Semantic underspecification: concepts and applications

Markus Egg
Humboldt-Universität Berlin

Semaine Nancéienne de sémantique formelle

Nancy, 24th March, 2010
Overview

• semantic underspecification: concepts
  – definition and background
  – fields of application and central issues

• semantic underspecification in Computational Semantics
  – pervasiveness of ambiguity
  – functionality of the syntax-semantics interface

• semantic underspecification: NLP applications
  – benefits of semantic underspecification for NLP applications
  – further processing of semantically underspecified expressions
  – semantic underspecification for large-coverage grammars and applications
Semantic underspecification: concepts

Definition and background
Introduction 1: definition and background

• underspecification is the deliberate omission of information from linguistic descriptions

• the goal is to capture several alternative realisations of a linguistic phenomenon in one single representation

• this avoids a disjunctive enumeration of the alternatives

• simple example 1: \{1, 2, 3, 4, 5\} can also be rendered as \{n|n \in \mathbb{N} \land n < 6\}

• simple example 2: agreement in English
  – simplistic phrase structure grammars must duplicate the relevant rules
    * \( S \rightarrow \text{NP\_SG} \ \text{VP\_SG} \)
    * \( S \rightarrow \text{NP\_PL} \ \text{VP\_PL} \)
  – constraint-based grammar makes do with one
    * \( S \rightarrow \text{NP} \quad \text{AGR: 1} \quad \text{VP} \quad \text{AGR: 1} \)
Introduction 2: definition and background

• typically there are two levels of description
  – an object-level of linguistic representation
  – a meta-level of describing these representations (often, only partially)
• underspecification emerged in phonology and was adopted in semantics in the early 1980’s
• in semantics, representations are about meaning
• underspecified semantic representations capture whole sets of different meanings
• these sets of meanings are alternative interpretations of one single linguistic form
• i.e., semantic underspecification is a means to deal with ambiguity
Introduction 3: definition and background

• typically, the ambiguities covered by semantic underspecification are structural
  – structural ambiguity arises if one (morpho-)syntactic structure corresponds to several meanings
  – these ambiguities are not due to ambiguous words like bank

(1) Every attorney hired a secretary
(2) Almost every attorney of every law firm in a big city hired a secretary
(3) The attorneys hired secretaries
(4) 500 companies own 6,000 computers
(5) beautiful dancer
Introduction 4: a simple example

- consider the ambiguous example (6) [= (1)]

(6) *Every attorney hired a secretary*

- in its readings, the semantic contributions of the NPs are *arranged differently*:

(7) $\forall x.\text{attorney}'(x) \to \exists y.\text{secretary}'(y) \land \text{hire}'(x, y)$

(8) $\exists y.\text{secretary}'(y) \land \forall x.\text{attorney}'(x) \to \text{hire}'(x, y)$

- its underspecified semantic representation abstracts away from this difference:

\[
\forall x. \text{attorney}'(x) \to \Box \quad \exists y. \text{secretary}'(y) \land \Box
\]

\[
\text{hire}'(x, y)
\]
Introduction 5: a simple example

• the ingredients of underspecified semantic representations
  – fragments of object-level semantic representations
  – glue points in these fragments
  – a relation between the glue points and the fragments

• if a glue point is related to a fragment, the fragment must be an (im-)proper part of the material that eventually attaches at the glue point

\[ \forall x. \text{attorney}'(x) \rightarrow \square \quad \exists y. \text{secretary}'(y) \land \square \]

\[ \text{hire}'(x, y) \]

Markus Egg, JSM 2010
Introduction 6: a simple example

- from underspecified semantic representations to the readings
  - correlate each glue point with one specific fragment (formally, a surjective mapping from fragments to glue points, the ‘plugging’ of Bos 2004)
  - this must respect the relation between glue points and fragments
  - one option correlates the top glue point with the left hand fragment

\[ \forall x. \text{attorney}'(x) \rightarrow \square \]

\[ \exists y. \text{secretary}'(y) \land \square \]

\[ \text{hire}'(x, y) \]

- replacing glue points by fragments yields the reading with wide scope for the universal quantifier

Markus Egg, JSM 2010
Semantic underspecification: concepts

Fields of application and central issues
Fields of application and central issues 1: scope

(9) Every attorney hired a secretary [= (6)]

• articles like every and a introduce quantifiers in their semantics
• quantifiers indicate (sets of) individuals about which a claim is to be made
  – this is called the restriction of the quantifier
  – e.g., every attorney indicates that a claim about every element of the set of attorneys is at stake
  – a secretary indicates that there will be a claim about one element from the set of secretaries
• scope of a quantifier in a semantic representation is that part of the representation that expresses the claim about the individuals indicated by the quantifier
• one quantifier may be in the scope of another as in (9)
Fields of application and central issues 2: scope

- scope ambiguities are the prototypical structural ambiguities
- scope can be modelled by the relation between fragments and glue points
- to handle scope ambiguity omit the scope relations that make up the differences between the readings
- the set of readings must be constrained appropriately, e.g., for nested NPs

(10) Every attorney of a law firm saw most clients

\[ \exists y. \text{law\_firm}'(y) \land \square \quad \forall x. (\text{attorney}'(x) \land \square) \rightarrow \square \quad \text{most}'(\text{client}', \lambda z. \square) \]

of'(x, y) \quad \text{see}'(x, z)
(11) *Every attorney of a law firm saw most clients*

- option 1: give the ‘a-law-firm’ fragment narrow scope
- then the ‘every-attorney’ or the ‘most-clients’ fragment may get widest scope
- i.e., this option describes **two readings**

$$\forall x. (\text{attorney}'(x) \land \Box) \rightarrow \Box$$

$$\exists y. \text{law\_firm}'(y) \land \Box$$

$$\text{see}'(x, z)$$

$$\text{of}'(x, y)$$

$$\text{most}'(\text{clients}', \lambda z. \Box)$$
Fields of application and central issues 4: scope

(12) *Every attorney of a law firm saw most clients*

- option 2: give the ‘a-law-firm’ fragment scope over ‘every-attorney’
- then the ‘most-clients’ fragment may scope above, between, or below them
- i.e., this option yields *three readings*

\[
\exists y. \text{law\_firm}'(y) \land \Box \\
\forall x. (\text{attorney}'(x) \land \Box) \rightarrow \Box \\
\text{of}'(x, y) \\
\text{see}'(x, z) \\
\text{most}'(\text{clients}', \lambda z. \Box)
\]
Fields of application and central issues 5: scope

- scope ambiguities do not affect DPs only
  - negation and modal expressions
    
    (13) *Everyone didn’t come*
    
    (14) *A unicorn seems to be in the garden*
  
  - conjunction
    
    (15) *I want to marry Peggy or Sue*
    
    (16) *Every attorney and his secretary met*
  
  - scope below the word level
    
    (17) *beautiful dancer*
    
    (18) *John’s former car*
    
    (19) *John almost died*
Fields of application and central issues 6: scope

- a related problem: modifier attachment ambiguities

(20) *We will discuss the meeting on Monday*

- *on Monday* modifies either *meeting* or the VP *discuss the meeting*
- i.e., there are two different syntactic structures with different interpretations

- can we unify the different interpretations in terms of an underspecified representation on the semantic level?

  - the two interpretations are closely related
  - all the readings of such ambiguities are composed of the same material, only the arrangement of the material is different
Fields of application and central issues 7: scope

- this suggests a representation in analogy to the one of structural ambiguity

\[ \text{discuss}'(\text{we}', \mathcal{L} \mathcal{E} \mathcal{M} \mathcal{I} \mathcal{N} \mathcal{G} \mathcal{Y} \mathcal{I} \mathcal{N} \mathcal{G} \mathcal{Y} (x)) \]

\[ \text{on}_{-} \text{monday}'(x) \]

\[ \text{meeting}'(x) \]

- NB: \( \mathcal{L} \mathcal{E} \mathcal{M} \mathcal{I} \mathcal{N} \mathcal{G} \mathcal{Y} \mathcal{I} \mathcal{N} \mathcal{G} \mathcal{Y} . P(x) \) refers to the only individual in the extension of \( P \)

- the readings of (20)

(21) \[ \text{discuss}'(\text{we}', \mathcal{L} \mathcal{E} \mathcal{M} \mathcal{I} \mathcal{N} \mathcal{G} \mathcal{Y} \mathcal{I} \mathcal{N} \mathcal{G} \mathcal{Y} (x))) \]

(22) \[ \text{on}_{-} \text{monday}'(\text{discuss}'(\text{we}', \mathcal{L} \mathcal{E} \mathcal{M} \mathcal{I} \mathcal{N} \mathcal{G} \mathcal{Y} \mathcal{I} \mathcal{N} \mathcal{G} \mathcal{Y} (x))) \]
Fields of application and central issues 8: ellipsis

- elliptical sentences are underspecified because material has been left out
- this material can be reconstructed from a (usually preceding) source sentence

(23) John slept. Mary did, too.

- elliptical (target) sentences have the same structure as their source sentence
- the sentences differ with respect to parallel elements (in (23), John and Mary)
- the meaning of the target sentence is the one of the source sentence, with the target parallel element replacing the source parallel element
- further syntactic restrictions apply (Sag 1976; Fiengo and May 1994)

(24) John greeted everyone that Bill did

- for a wide range of ellipses see Webber (1978) and Gawron and Peters (1990)

(25) John revised his paper before the teacher did and Bill did, too
Fields of application and central issues 9: ellipsis

- challenges of ellipsis
  - identify the source sentence
    - ok for VP ellipses in parsed English corpora (Hardt 1997; Nielsen 2004)
    - identifying source sentences during processing is restricted to intrasentential contexts (Gregory and Lappin 1998; Egg and Erk 2002)
  - distinguish the parallel element in the source sentence
    - the rest of the source sentence is then combined with the target parallel element to obtain the fully specified target sentence
    - e.g., by Higher-Order Unification (Dalrymple et al. 1991)
      - source and target sentence mean the same modulo parallel elements
      - the meaning of the source sentence is equated with a relation $P$ applied to the semantics of the source parallel element(s)
      - the target sentence means $P$ applied to the semantics of the target parallel element(s)
Fields of application and central issues 10: ellipsis

- example 1: *John slept. Mary did, too.*
  - we look for a \( P \) such that \( \text{sleep}'(j) = P(j) \)
  - i.e., \( P = \text{sleep}' \), and the meaning of the target sentence is \( \text{sleep}'(m) \)
  - such equations can be solved with algorithms like the one of Huet (1975)

- example 2: *Max loves his mother and Bill does too.*
  - two readings: whom does Bill love? (‘strict’ vs. ‘sloppy’)
  - the relevant equation: \( P(m) = \text{love}'(m, \text{mother-of}'(m)) \)
  - first (relevant) solution: \( P = \lambda x. \text{love}'(x, \text{mother-of}'(x)) \) (‘loving one’s own mother’)
  - second (relevant) solution: \( P = \lambda x. \text{love}'(x, \text{mother-of}'(m)) \) (‘loving Max’s mother’)

Markus Egg, JSM 2010
Fields of application and central issues 11: others

- further fields of application include
  - lexical ambiguity, e.g., for chemical terms like *butene* (Reyle 2006)
    * but- indicates 4 carbon atoms and -ene, a double bond
    * the position of the double bond is open, it could be either CC=CC or C=CCC
    * these terms further instantiate underspecification: remaining valencies are saturated with hydrogen atoms, unless specified
  - metonymy (Hobbs et al. 1993; Markert and Hahn 2002; Markert and Nissim 2009) and aspectual coercion (Pulman 1997; Egg 2005)

(26) *The Boston office called*

(27) *Amelie played the Moonlight Sonata for five minutes/for five days*
Fields of application and central issues 12: others

- further fields of application include
  - underspecification of argument labelling (Copestake 2009)
    * labels can be meaningless, e.g., ‘ARGₙ’ in (Robust) Minimal Recursion Semantics
    * instead of thematic roles (agent, experiencer etc.)
    * these roles could later be inferred from lexical databases like Wordnet (Fellbaum 1998) or Propbank (Palmer et al. 2005)
  - incompletely transmitted messages due to problems in production, transmission, or reception
Semantic underspecification in Computational Semantics

Pervasiveness of ambiguity
Pervasiveness of ambiguity 1

- ambiguities from different sources (lexical, anaphoric binding, structural) interact and multiply

(28) Früher stellten die Frauen der Inseln am Wochenende Kopftücher mit Blumenmotiven her, die ihre Männer an den folgenden Montagen auf dem Markt im Zentrum der Hauptinsel verkauften (Hans Uszkoreit)

‘The women from the islands used to produce headscarves with flower motifs over the weekend, which were sold by their husbands on the following Mondays at the market in the centre of the main island.’

- this sentence has 258,048 possible interpretations
Pervasiveness of ambiguity 2

- in particular, structural ambiguity must be processed in an efficient fashion
  - for complex expressions, generating and enumerating the readings would be too long-winded
  - example: the number of readings of a sentence with four nested NPs like in (29) [= (2)] is 14 (the Catalan number for 4)

(29) *Almost every attorney of every law firm in a big city hired a secretary*

- one of the formulae for the Catalan number of $n$: $\frac{(2n)!}{n!(n+1)!}$
- the first ten Catalan numbers:
  1, 2, 5, 14, 42, 132, 429, 1430, 4862, 16796

Markus Egg, JSM 2010
Pervasiveness of ambiguity 3

- pilot study on scope conducted at the Universität des Saarlandes
  - leading question: is scope really an issue out there in real life?
  - basis: German newspaper corpus (Frankfurter Rundschau) with syntactic annotations from the TIGER treebank (Brants and Hansen 2002)
  - around 10% of the sentences provided interesting cases of scope ambiguity
  - the complexity of these cases may well exceed the examples from linguistic textbooks

(30) *Die Mitglieder der Gruppe verteilen ein Handbuch an alle Anwesenden*

‘The members of the group distribute a handbook to all participants’

scope relations: there is a kind of handbook such that participants obtain a concrete copy of the handbook
Pervasiveness of ambiguity 4

• in large-scale NLP applications, the ambiguity problem is greatly enhanced by spurious ambiguities
  – e.g., 5% of the representations in the Rondane Treebank (Copestake and Flickinger 2000) have more than 650,000 solutions (Koller et al. 2008)
  – record holder is (31) with about $4.5 \times 10^{12}$ scope readings:

    (31) Myrdal is the mountain terminus of the Flåm rail line (or Flåmsbana) which makes its way down the lovely Flåm Valley (Flåmsdalen) to its sea-level terminus at Flåm.

  – the median number of scope readings per sentence is 56 (Koller et al. 2008)
Semantic underspecification in Computational Semantics

Functionality of the syntax-semantics interface

Markus Egg, JSM 2010
Functionality of the syntax-semantics interface

- ambiguity means that there is one form with several meanings
- no immediate functional mapping from syntactic to semantic structure
- one can restore a 1-1 relation between syntactic and syntactic structure by multiplying syntactic structures: Montague Grammar, Generative Grammar
  - assume one syntactic structure per reading
  - these structures are situated on a not directly observable syntactic level (e.g., Logical Form in Generative Grammar, Hornstein 1995)
  - e.g., LF structures for (1) (simplified):
    (32) \([S [every \text{attorney}]_i [S [a \text{secretary}]_j [S t_i \text{ hired } t_j]]]\)
    (33) \([S [a \text{secretary}]_j [S [\text{every attorney}]_i [S t_i \text{ hired } t_j]]]\)
  - but in NLP, surface-oriented syntactic analyses prevail

Markus Egg, JSM 2010
Functionality of the syntax-semantics interface 2

• one could also relinquish the 1-1 relation and assume a nondeterministic relation between syntax and semantics
  – scope-bearing items enter the semantic representation unscoped and can receive scope non-deterministically
  – e.g., in Cooper Storage approaches (Cooper 1983) or the Categorial Grammar of Steedman (2007)

• underspecification preserves the 1-1 relation by adapting semantics to syntax
  – assume a single semantic representation per syntactic structure
  – such representations can stand for whole sets of readings
  – they are thus meta-level expressions
  – the relation of meta-level expressions and readings is well-defined
  – this fits in with the surface orientation of syntactic analyses in NLP
Functionality of the syntax-semantics interface 3

- assuming a one-many relation between syntactic and semantic structures would be inadequate for purely semantic reasons
- it enforces interpreting an ambiguous expression as disjunction of its readings

(34) [\(= (1)\)] *Every attorney hired a secretary*
- e.g., (34) has the readings ‘\(\forall \exists\)’ (‘one-each’) and ‘\(\exists \forall\)’ (‘one-for-all’):
- thus, its interpretation would be \(\forall \exists \lor \exists \forall\) (‘either one-each or one-for-all’)

- but this predicts a wrong interpretation for negated ambiguous expressions:

(35) *It is not the case that every attorney hired a secretary*
- prediction: \(\neg (\forall \exists \lor \exists \forall) = \neg \forall \exists \land \neg \exists \forall\) (‘neither one-each nor one-for-all’)
- in words: all of the readings are negated

- formally, this is a negated disjunction of readings
Functionality of the syntax-semantics interface 4

- this conflicts with intuitions about the negation of (34), which interpret it as
  \( \neg \forall \exists \lor \neg \exists \forall \) (‘either not one-each or not one-for-all’)
  - in words: one of the readings is negated
  - formally, this is a disjunction of negated readings

- alternative approach
  - model the meaning of ambiguous expressions as sets of readings
    * the meaning of (34): \( \{ \forall \exists, \exists \forall \} \)
    * the meaning of the negation of (34): \( \{ \neg \forall \exists, \neg \exists \forall \} \)
  - model their assertion as a disjunction of the elements of this set
    * the meaning of the assertion of (34): \( \forall \exists \lor \exists \forall \)
    * the meaning of the assertion of the negation of (34): \( \neg \forall \exists \lor \neg \exists \forall \)
Semantic underspecification: in NLP applications

Benefits of semantic underspecification for NLP applications

Markus Egg, JSM 2010
Benefits of underspecification 1: delayed disambiguation

- further processing of semantic representations of structurally ambiguous expressions
  - modules for the further processing of semantic representations call for efficiently processable representations
  - many structural ambiguities are irrelevant for further processing, e.g., in Machine Translation

(36) *The attorneys hired secretaries*

(37) *Die Anwälte stellten Sekretärinnen ein*

- consequence: postpone the enumeration of individual readings and let it be guided by demand
- this was the strategy e.g. of the Verbmobil MT system (spontaneous spoken dialogue in the domain of scheduling meetings, Wahlster 2000)

Markus Egg, JSM 2010
Benefits of underspecification 2: portability

- underspecification formalisms share basic intuitions on structural ambiguities
  - meta-level expressions describe object-level semantic representations
  - deliberate omission of differences between readings
  - focus on (underspecification of) scope relations

- these formalisms include
  - Quasi-Logical Form (QLF; Alshawi 1992)
  - Underspecified Discourse Representation Theory (UDRT; Reyle 1993)
  - Hole Semantics (HS; Bos 2004)
  - Glue Language Semantics (‘Glue’: Dalrymple et al. 1997)
  - Constraint Language for Lambda Structures (CLLS; Egg et al. 2001)
  - Minimal Recursion Semantics (MRS; Copestake et al. 2005)
  - Regular Tree Grammars (Koller et al. 2008)
Benefits of underspecification 3: portability

• semantic modules with underspecification can be based on surface-oriented syntactic modules
  – strategies of and intuitions on syntax-semantics interfaces can be reused for different syntactic modules
  – underspecification makes possible very flexible syntax-semantics interfaces
  – there is no need to rearrange syntactic structures for semantic processing

• surface-oriented syntactic frameworks that are used as the basis of semantic modules with underspecification include
  – Head-driven Phrase Structure Grammar (HPSG; Pollard and Sag 1994)
  – Lexical-Functional Grammar (LFG; Dalrymple 2001)
  – Lexicalised Tree-Adjoining Grammar (LTAG; XTAG-Group 2001)
  – Combinatory Categorial Grammar (CCG; Bos 2008)
Benefits of underspecification 4: portability

- realised combinations of syntactic and semantic formalisms

<table>
<thead>
<tr>
<th></th>
<th>HPSG</th>
<th>LFG</th>
<th>LTAG</th>
<th>CCG</th>
</tr>
</thead>
<tbody>
<tr>
<td>MRS</td>
<td>(Copestake et al. 2005)</td>
<td>(Oepen et al. 2004)</td>
<td>(Kallmeyer and Joshi 1999)</td>
<td></td>
</tr>
<tr>
<td>Glue</td>
<td>(Asudeh and Crouch 2001)</td>
<td>(Dalrymple 1999)</td>
<td>(Frank and van Genabith 2001)</td>
<td></td>
</tr>
<tr>
<td>HS</td>
<td>(Chaves 2002)</td>
<td></td>
<td>(Kallmeyer and Joshi 2003)</td>
<td></td>
</tr>
</tbody>
</table>
Benefits of underspecification 5: portability

- translations between underspecification formalisms underlying CS modules are possible
- there is a core common to all these formalisms, but no general equivalence, since formalisms are of different expressivity (Koller et al. 2008)
- realised translation steps
  - from HS to CLLS (Koller et al. 2003)
  - from MRS to CLLS (Fuchss et al. 2004)
  - from QLF to HS (Lev 2005)
- with such translations, one could combine the advantages of several CS modules, e.g.
  - wide coverage
  - efficient methods of further processing
Benefits of underspecification 6: hybrid processing

• semantic underspecification can bridge the gap between deep and shallow processing
  – instead of combining deep and shallow syntactic analyses, semantics is the level where the two are combined
  – advantage 1: combine the best of both worlds
    * shallow processing is robust, fast, and has wider coverage
    * deep processing is more accurate and provides information that is crucial for further NLP modules (e.g., argument structure)
  – advantage 2: semantics is the level at which information from different threads of processing must be integrated anyway

• this calls for appropriately formulated semantic representation formalisms that can express various levels of underspecification
Benefits of underspecification 7: hybrid processing

- Robust Minimal Recursion Semantics (RMRS; Copestake 2006) was designed for this task
- shallow syntactic analyses provide a part of the information to be gained from deep analysis, e.g.
  - part-of-speech taggers: no connection between semantic contributions of different words
  - NP chunkers: no argument linking
- i.e., the semantic information derivable from the results of a shallow parse can only be a part of that to be derived from the results of a deep parse
- semantic representations of different depths must be compatible
  - to combine results from parallel deep and shallow processing
  - to make shallow semantic analyses deep by adding more information
Benefits of underspecification 8: hybrid processing

- example: argument linking
  - NP chunkers do not relate verbs and their syntactic arguments
    \[(38) \text{Max met Mary}\]
  - then semantic analysis cannot identify individuals in NP and verb semantics
    \[(39) \text{named}(x_1, \text{Max}), \text{meet}(x_2, x_3), \text{named}(x_4, \text{Mary})\]
  - a deep semantic analysis would relate them (Copestake 2006)
    \[(40) \text{named}(x_1, \text{Max}), \text{meet}(x_2, x_3), \text{named}(x_4, \text{Mary}), x_1 = x_2, x_3 = x_4\]
  - i.e., the formalism must be able to separate the information that a shallow analysis can deliver
    \[(41) \text{named}(x_1, \text{Max}), \text{meet}(x_1, x_4), \text{named}(x_4, \text{Mary})\]
Semantic underspecification: in NLP applications

Further processing of semantically underspecified expressions
Further processing of underspecified expressions 1

- resolution of underspecified expressions
  - enumeration of the readings by resolving the constraints
  - for some formalisms, so-called solvers are available for this task
    * a MRS solver is part of the Linguistic Knowledge System (Copestake and Flickinger 2000)
    * there is a solver for HS, too (Blackburn and Bos 2005)
    * for the language of dominance constraints, a number of solvers is available (Koller, Regneri, and Thater 2008)
  - however, considering the huge number of readings, a full resolution is often neither efficient nor useful
  - still, resolution is a kind of benchmark test for efficient encoding of underspecification
Further processing of underspecified expressions 2

- redundancy elimination
  - identifying and elimination of spurious ambiguities

(42) The boy met the girl

(43) The president of every country is a crook

- considering the severity of the redundancy problem, this is very desirable
- ideally, this should be done before the readings are enumerated
- early approaches filter out equivalent readings after resolution (Moran 1988; Hurum 1988)
- in the meantime, spurious ambiguities can be reduced before resolution (Koller, Regneri, and Thater 2008)
Further processing of underspecified expressions 3

- reasoning with underspecified expressions (van Eijck and Jaspars 1996; Reyle 1996)
  - e.g., (44) entails (45), no matter which reading of (44) is at stake
    (44) Every woman loves a man
    (45) Amélie loves a man
  - different kinds of consequence relations can be used (van Deemter 1996)

- integration of preferences

(46) A secretary was hired by every attorney
  - from syntax (S vs. O, c-command, θ-hierarchies), semantics (strength of interpretations), lexicon (kind of determiner), conceptual knowledge
  - much more work could be done in this area . . .
Underspecification for large grammars

- MRS is used in the English Resource Grammar (ERG),
  - a broad-coverage HPSG-based grammar of English (Flickinger 2002)
  - online demo: http://erg.delph-in.net/logon

- based on experience with the ERG, the LinGO Grammar Matrix was
developed (Flickinger, Bender, and Oepen 2003)
  - a framework for the development of broad-coverage grammars
  - used for grammars of languages from different language families

- Glue is used in the ParGram (Parallel Grammar) Project (Butt et al. 2002)
  - wide coverage grammars for typologically diverse languages in LFG
  - based on XLE parser/grammar development platform (Kaplan et al. 2002)

- HS is used within XTAG (wide-coverage grammar for English encoded in
  LTAG; Doran et al. 2000), (Joshi et al. 2007)
Underspecification in NLP applications 1

- **Machine Translation**: LOGON (Norwegian → English; Oepen et al. 2004)
  - MRS is used for semantic transfer
  - the semantic component of the LFG-based parser for Norwegian
  - for the HPSG-based generator for English
- **Information Extraction**: SciBorg (IE from chemistry papers; Rupp et al. 2008)
- **Question Answering**: Quetal (Frank et al. 2004)
  - crosslingual Question Answering using RMRS structures from shallow parsers
  - also uses parts of the SProUT (Shallow Processing with Unification and Typed Feature Structures) processing platform (Becker et al. 2002)
- **ontology extraction** (Nichols et al. 2005; Herbelot and Copestake 2008)
Underspecification in NLP applications 2

- further applications were pursued in the project DeepThought (Callmeier et al. 2004)
  - information extraction for business intelligence
  - e-mail response management for customer relationship management
  - creativity support for document production and collective brainstorming
  - deep and shallow modules of processing were combined in the Heart of Gold architecture (Schäfer 2007)
    * this architecture integrates modules for deep and shallow syntactic analysis (from chunker to deep parser)
    * it is designed language-independently
    * it was used in other other projects, too, e.g., Quetal
Underspecification in NLP applications 3

• the project DeepThought (ctd)
  – innovative parsers are included in the system, too, e.g., the RASP system (Robust accurate statistical parsing; Briscoe and Carroll 2002)
    ∗ this parser is intermediate in depth between tagging and deep processing
    ∗ it outputs simple dependency structures and does not use lexical subcategorization information
    ∗ this degree of informativity can be expressed precisely in RMRS
Summary

• semantic underspecification is an important concept in present-day NLP
• it has been researched extensively and is formally very well understood
• it is one of the techniques that keep semantic representations tractable in spite of the ambiguity problem
• it makes the definition of semantic interfaces easier
• still, much more work needs to be done for the eventual derivation of single, fully specified interpretations

Markus Egg, JSM 2010
References


Dalrymple, M., S. Shieber, and F. Pereira (1991). Ellipsis and higher-order...


Faculty of Mathematics and Computer Science, Universität des Saarlandes, Saarbrücken.


