On deriving semantic representations from dependencies

Michael Hahn  Detmar Meurers
Universität Tübingen
SFB 833, Project A4

(Research first presented at Depling 2011, Barcelona)

Second Tübingen-Berlin Meeting on Analyzing Learner Language
Tübingen December 6, 2011

Introduction

▶ CL in recent years has seen a rise in interest in tasks that involve the evaluation and comparison of meaning.
  ▶ e.g., Recognizing Textual Entailment Challenges
▶ We are exploring methods for comparing the meaning of answers to reading comprehension questions.
  ▶ basic research on evaluating meaning in context
  ▶ of practical relevance for intelligent tutoring system
▶ Meaning comparison can involve different representations
  ▶ comparisons of surface forms
    ▶ e.g., BLEU (Lin & Och 2004), ROUGE (Lin 2004) metrics
    ▶ deep semantic analysis and logical inference
      ▶ e.g., Bos & Markert (2006)
      ▶ idea: explore the space in-between the two extremes
▶ In this talk: derivation of underspecified semantic representations on the basis of dependency analysis.

Outline

Introduction
  Motivation
  Our Approach
Syntax-Semantics Interface Representations
  Example
  Deriving interface representation
Building LRS Representations
  Lexical Resource Semantics
  From interface terms to LRSs
Experiments
  Setup
  Corpus used
  Results
Summary

Our Approach

▶ Obtain semantic representations in two steps:
  ▶ Step 1: Transform dependency structure into a syntax-semantics interface representation for each word.
  ▶ Step 2: Compute LRS semantic representation.
  ▶ provides explicit semantic representation while also exposing all minimal building blocks of the semantics
  ▶ rich space of representations for comparing meanings even when no complete semantics can be obtained
▶ Our project deals with the automatic analysis of answers written by language learners: robustness needed
Syntax-Semantics Interface Representations

- Motivation of interface representations:
  - abstract away from variation in form and grammaticality
  - support simpler, more robust semantic construction
- Interface representations bear similarity to aspects of:
  - LFG f- and a-structures (Kaplan & Bresnan 1995)
  - Deep Syntactic Structure of Meaning Text Theory (Mel’cuk 1988)
- Our goal: achieve good coverage of language phenomena and deal with well-known argument-structure challenges:
  - local, middle-distance and long-distance relations
  - includes challenges for identifying argument structure: extraction, raising, control, passive, and their interaction (e.g., long-distance passive)
- In interface representations, properties needed to identify these relations are represented by a set of features:
  - defined for every word
  - information about categories, valency, modification, . . .

Deriving Interface Representations

- some of the features are straightforward to specify locally based on dependency and part-of-speech information
- recursive processing of dependencies needed for phenomena involving dependents which are not realized locally (extraction and non-finite constructions)
  - Procedural approach provides robustness in the presence of dependency parsing errors or ungrammaticality.
- Basic algorithm building interface representations:
  - First build interface representations for locally realized dependencies of each word.
  - Then recursively reconstruct other relations by moving, copying and adding dependents.
    - reconstruction starts from the least embedded verb, increasing depth of arguments
    - decisions are made locally, based only on one verb and its dependent at a time, using information from a lexicon
Lexical Resource Semantics (LRS)

- embeds a standard model-theoretic semantic language (Ty2) into typed feature structures as used in HPSG
- LRS is an **underspecified** semantic formalism:
  - semantic representations define a set of possible resolved formulae (not a single fully explicit logical form)
  - LRS includes a model theoretic interpretation, as opposed to evaluating the formulae outside the representation formalism as in MRS (Copestake et al. 2005)

An example with a scope ambiguity

LRS representation

- **INCONT**: core semantic contribution of head
- **EXCONT**: overall semantic representation
- **PARTS**: list containing all semantic subterms

\[
\begin{align*}
\text{INCONT} & : \text{love}(e) \\
\forall x(\text{man}(x) \rightarrow \alpha), \\
\exists y(\text{woman}(y) \land \beta), \\
\exists e(\text{love}(e) \land \text{subj}(e, x) \land \text{obj}(e, y)), \\
\text{love}(e), \text{man}, x, \ldots
\end{align*}
\]

The resolved overall semantics includes \( \forall \exists \) reading:

- **EXCONT** \( \forall x(\text{man}(x) \rightarrow \alpha) \)
- \( \alpha = \exists y(\text{woman}(y) \land \beta) \)
- \( \beta = \exists e(\text{love}(e) \land \text{subj}(e, x) \land \text{obj}(e, y)) \)

Semantic specification in lexical entries

- **man**: \( \text{INCONT} \{ \text{man}(x) \} \)
  - **PARTS**: \( \{ \text{man}, x \} \)
- **every**: \( \text{INCONT} \{ \forall x(\text{man} \rightarrow \text{man}) \} \)
  - **PARTS**: \( \{ \text{man}, \text{man} \} \)
- **loves**: \( \text{INCONT} \{ \text{love}(e) \} \)
  - **PARTS**: \( \{ \text{love}(e) \land \text{subj}(e, x) \land \text{obj}(e, y) \}, \ldots \)
### Adapting LRS to dependency framework

- LRS was developed for constituency-based HPSG.
- The syntactic composition constraints of LRS can be translated into lexicalized, dependency-based formalisms such as our interface representations.
- **Why?**
  - Strictly lexicalized: all subexpressions originate in lexicon
  - Phrasal notions can be restated in terms of dependencies
  - Maximal projection = set of direct and indirect dependents
  - \( \text{excont of the sentence} = \text{excont of the root word} \)

### From interface terms to LRSs

- Semantics for each word built separately, based only on the interface representation of the word
- The LRS representation is successively built up by applying rewrite rules which add constraints and elements to \( \text{PARTS} \)

### Building LRS Representations

**Example Rule**

\( \text{cat} = \text{verb} : \)
- \( \text{INCONT} \) unified with \( \text{PREd} (\text{INDEX}) \)
- add \( \exists \text{INDEX} \alpha \land \beta, \text{INCONT} \) and their subexpressions

For example, for (4) we obtain:

**Interface Representation:**

\[
\begin{align*}
\text{PRED} & \quad \text{love} \\
\text{INDEX} & \quad e \\
\text{PARTS} & \quad \langle \ldots, \text{love}(e), \exists e \alpha \land \beta, \ldots \rangle
\end{align*}
\]

**Semantics:**

\[
\begin{align*}
\text{INCONT} & \quad \text{love}(e), \exists e \alpha \land \beta \\
\text{PARTS} & \quad \langle \alpha \land \beta, \text{love}, e \rangle
\end{align*}
\]
On deriving semantic representations from dependencies

Michael Hahn, Detmar Meurers

Introduction
Motivation
Our Approach
Syntax-Semantics Interface Representations
Example Deriving interface represent.
Building LRS Representations Lexical Resource Semantics
From interface terms to LRSs
Experiments
Setup Corpus used Results
Summary

Setup

- Task: Evaluate approach on German learner corpus
  - quality and robustness of semantic representation
- Two ways to obtain dependency analysis
  - manual dependency annotation
  - automatic dependency parses using MaltParser (Nivre et al. 2007) trained on Tüba-D/Z (Telljohann et al. 2004) converted to dependencies (Versley 2005)
- Step 1: Automatically derive interface representations based on dependency analysis and lexical information
- Step 2: Automatically derive LRS structures based on interface representations
- Evaluate correctness of resulting LRS structure

Corpus used

- CREG-109 corpus (Ott & Ziai 2010), sub-corpus of the Corpus of Reading Comprehension Exercises in German (CREG, Meurers et al. 2010)
- answers to reading comprehension exercises written by US college students at beginner/intermediate levels
- 109 sentences (sentence length: avg. 8.26, max. 17)
- 17 ungrammatical sentences
  - errors in word order, agreement, and case government
- manual annotation of LRS semantics
  - For ungrammatical sentences, the semantics of a grammatical target hypothesis was annotated.

Results

- Percentage of answers with completely correct analyses:

<table>
<thead>
<tr>
<th></th>
<th>Grammatical</th>
<th>Ungrammatical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manual</td>
<td>82.6 %</td>
<td>70.5 %</td>
</tr>
<tr>
<td>Automatic</td>
<td>65.5 %</td>
<td>47.1 %</td>
</tr>
</tbody>
</table>

- Generally only annotated semantics counted as correct.
  - Exception: adverbial modifiers verbs (where Montagovian and Neo-Davidsonian representations were accepted)
We presented a system that automatically derives underspecified, model-theoretic semantic representations from dependency parses of German learner sentences.

Two step process:

- transforming dependency structures into syntax-semantics interface representations
  - generalizes over a range of syntactic and morphological options for realizing the same meaning
- derive LRS semantic representations using simple rules based on interface representations

We successfully tested the approach using the CREG-109 learner corpus.

Lexical Resource Semantics can fruitfully be used with dependency-based syntactic representations.

Dependency analysis supports robust generation of semantic representations for leaner language.

References


