Deep Linguistic Analysis, Interfaces and World Knowledge

Stefan Müller
Deutsche Grammatik
Institut für Deutsche und Niederländische Philologie
Fachbereich Philosophie und Geisteswissenschaften
FU Berlin
Stefan.Mueller@fu-berlin.de

Introduction

History of CL

Different Approaches (Uszkoreit, 2001)

Constraint-Based Grammar

• Constraint-based grammars were developed as an alternative to generative grammars.
• Frameworks: GPSG, LFG, HPSG, and CxG (with the respective basic assumptions)
• Some of these frameworks started out as generative grammars, but the underlying assumptions changed in the mid of the 90s.
• Example for a constraint on English sentences:

(1) a. Kim loves Sandy.
    b. * I loves Sandy.

The subject has to agree with the verb in person and number.

• Constraint-based view: If nothing is said, everything goes.

See Pullum and Scholz, 2001 for consequences of this view.

The Framework

• the framework: Head-Driven Phrase Structure Grammar (HPSG) (Pollard and Sag, 1987, 1994; Müller, 2007).
• Since Sag, 1997 this is compatible with basic insights from Construction Grammar (CxG).
• Since it is a constraint-based grammar nowadays, it is not a phrase structure grammar.
• The framework allows for and uses headless constructions (Müller, 1999, Chapter 10).
• Recent development Sign-based CxG (Sag, 2007, 2010).
• Embodied CxG (Bergen and Chang, 2005) can be reformulated as a HPSG (see Müller, 2010a, Chapter 9.6.2).
• Minimalist Grammars correspond to the main schemata in HPSG (Müller, To appear)
Basic Ingredients: Features and Values

To describe language we need only three things:

- features
- values
- identities between values

Values can be simple or complex:

```
FIRST NAME  max
SURNAME     meier
BIRTHDAY    10.10.1985

FIRST NAME  peter
SURNAME     meier
BIRTHDAY    10.05.1960
FATHER ...  
MOTHER ...  
```

Types

In addition we have types:

```
FIRST NAME  max
SURNAME     meier
BIRTHDAY    10.10.1985
FATHER ...  
MOTHER ...  
```

Types define which features belong to a description of certain objects.
Types are organized in hierarchies →
generalizations over all kinds of linguistic objects

Heads and Arguments

Basic aspects of the sentence structure are determined by the verb.

(2) a. Peter sleeps.
b. Peter likes Maria.
c. Peter talks about the financial crises.

The verb is called head.
The non-verbal elements in (2) are arguments of the verb.

Valence Representations

Descriptions of lexemes contain a list with descriptions of the syntactic and semantic properties of their arguments.

(3) Verb  ARG-ST              SUBJ        COMPS
sleeps  ⟨NP[nom]⟩            ⟨NP[nom]⟩  ( )
likes  ⟨NP[nom], NP[acc]⟩    ⟨NP[nom]⟩  ⟨NP[acc]⟩
talks  ⟨NP[nom], PP[about]⟩  ⟨NP[nom]⟩  ⟨PP[about]⟩

For English, the ARG-ST list is mapped onto two valence features: SUBJ and COMPS.
An Intransitive Sentence

\[ V[\text{SUBJ } \langle \rangle, \text{COMPS } \langle \rangle] \]

\[ \text{NP[nom]} \]

\[ \text{Peter} \]

\[ \text{sleeps} \]

A Transitive Sentence

\[ V[\text{SUBJ } \langle \rangle, \text{COMPS } \langle \rangle] \]

\[ \text{NP[nom]} \]

\[ V[\text{SUBJ } \langle 1 \rangle, \text{COMPS } \langle 2 \rangle] \]

\[ \text{NP[acc]} \]

\[ \text{Kim} \]

\[ \text{likes} \]

\[ \text{Sandy} \]

Constituent Structure

- Remember that I claimed that we need features and values only?
  So what are these trees?

\[ \text{NP} \]

\[ \text{Det} \]

\[ \text{N} \]

\[ \text{the} \]

\[ \text{man} \]

\[ \text{PHON } \langle \text{the man} \rangle \]

\[ \text{HEAD-DTR } \]

\[ \text{PHON } \langle \text{man} \rangle \]

\[ \text{NON-HEAD-DTR} \]

\[ \text{PHON } \langle \text{the} \rangle \]

Languages with free(er) constituent order

- languages like German and Persian:
  all arguments (of finite verbs) represented under \text{COMPS}.
- An arbitrary element from the list can be combined with the head.
Demo: Valence, Head Features, and Order

Grammar 4:

(7) der Mann die Frau kennt
the man.NOM the woman.ACC knows

Grammar 9:

(8) a. der Mann der Frau das Buch gibt
the man.NOM the woman.DAT the book.ACC gives
   b. der Mann das Buch der Frau gibt
the man.NOM the book.ACC the woman.DAT gives

Semantics: Nominal Objects

• The meaning of *dog* is usually represented as:

(9) \( \lambda x \text{dog}'(x) \)

• Features and values (Copestake, Flickinger, Pollard and Sag, 2005):

(10) \[
\begin{bmatrix}
\text{IND} & \text{PER} & 3 \\
\text{NUM} & \text{sg} & \text{index} \\
\text{RELS} & \text{INST} & 1 \\
\text{mr} & \text{dog} & 1
\end{bmatrix}
\]

\( \text{mr} \) is a referential index (person, number(, gender) for pronoun binding). 
\( \text{RELS} \) is a list of restrictions.
Semantics: Verbal Objects

- Meaning of *chase* with an event variable:
  \[ \lambda y \lambda x \lambda e \text{chase}'(e,x,y) \]

- Features and values:
  \[
  \begin{bmatrix}
  \text{IND} & \text{event} \\
  \text{EVENT} & \text{index} \\
  \text{AGENT} & \text{index} \\
  \text{PATIENT} & \text{index} \\
  \text{mrs} \\
  \end{bmatrix}
  \]

Linking

- Linking of valency information and semantic contribution
  \[
  \begin{bmatrix}
  \text{HEAD} & \text{VFORM fin verb} \\
  \text{ARG-ST} & \text{NP[nom] } \text{NP[acc]} \\
  \text{IND} & \text{event} \\
  \text{AGENT} & \text{index} \\
  \text{PATIENT} & \text{index} \\
  \text{mrs} \\
  \end{bmatrix}
  \]

- Referential indices of NPs are identified with the semantic roles.

Composition (Simplified)

DemO: Scope and Underspecification

Dowty, 1979, Section 5.6 on *again*, Egg (1999) on (14):

(14) a. dass Max wieder alle Fenster öffnete
    that Max again all windows openend
    'that Max opened all windows again'

b. *again*(\(\forall\text{(CAUSE(open)})\); repetitiv

c. *again*(\(\forall\text{(open)})\); repetitiv

d. \text{CAUSE}(\text{again}'(\forall\text{(open)})); restitutiv

Minimal Recursion Semantics: Copestake, Flickinger, Pollard and Sag, 2005
Selectional Restrictions

- Ambiguous or not?
  (15) Das Buch liest das Kind.
  the book reads the child
- Read requires an animate (human?) subject.
- Wellformed or not?
  (16) Peter isst einen Stein.
  Peter eats a stone
- We can exclude it, but:
  (17) Peter isst keine Steine, denn Steine kann man nicht essen.
- Ontological constraints may be violated under negation.

World Knowledge and Grammar

- Boas (2003) discusses resultative constructions like (18):
  (18) But most of the dogs were unhappy and would bark themselves hoarse. (BNC)
- Suggests an analysis in which lexical items contain world knowledge to rule out (19):
  (19) * The dog barked the postman hoarse.
- But what about (20)?
  (20) # The dog barked and as a result of this the postman became hoarse.

For (20) we need the full machinery with inferences and so on anyway (Müller, 2005).

Worldknowledge and Grammar

Event-based frame semantic representation of the prototypical sense of *melt*

<table>
<thead>
<tr>
<th>GOAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Ag)</td>
</tr>
</tbody>
</table>

Ag: Entity exerting energy that can lead objects to liquefy
Pt: Object that warms up and/or liquefies as the result of energy being applied to it
p3: SYN: PP
SEM: denoting the end result state (or end location of a path) of a mass that has been liquefied by heat

According to Boas, p3 is provided by world knowledge.

More on World Knowledge

(21) a. Dann erzählt Juliane Lumumba von den Tonbändern im Archiv, die wegen fehlender Klimaanlage in der tropischen Hitze zu einer schwarzen Masse schmolzen.\(^1\)
   b. Dann ging mal das Schreibpapier aus oder die bestellte Ladung Kerzen war zu Wachs geschmolzen, ehe sie den Hafen erreicht hatte.\(^2\)
   c. Erz wird „direkt“ zu Eisenschwamm reduziert, der dann zu flüssigem Roheisen geschmolzen wird.\(^3\)

\(^1\)Frankfurter Rundschau, 05.08.1997, p. 3.
\(^3\)Die Presse, 11.01.1992; Voest will vom Staat acht Milliarden Schilling.
World Knowledge and Melting Points

Why is a sentence like (22) strange?

(22) # Dann erzählte Juliane Lumumba von den Stahlträgern im Archiv, die wegen fehlender Klimaanlage in der tropischen Hitze zu einer schwarzen Masse schmolzen.

Summary and Outlook

- Sketch of a constraint-based theory
- Syntax and semantics are treated in the same structure and are linked
- Syntactic and semantic (and other constraints) are unordered
- World knowledge should be separate but linked. Complex inference mechanisms are needed.
- Processing procedures should be separate from linguistic knowledge.
- Currently in Berlin: German, Persian, Chinese, Maltese, Danish, ...
- Grammars use the same “core” → generalizations over properties that are shared by several languages. (Müller, 2013)
- Approach different from general MGG approach:
  We start with the individual languages and search for generalizations.


