Overview

- What we’re trying to do
- The pieces of our grammar
- Two extended examples
- Reflection on what we’ve done, what we still have to do
- Reading questions
What We’re Trying To Do

• Objectives
  • Develop a theory of knowledge of language
  • Represent linguistic information explicitly enough to distinguish well-formed from ill-formed expressions
  • Be parsimonious, capturing linguistically significant generalizations.

• Why Formalize?
  • To formulate testable predictions
  • To check for consistency
  • To make it possible to get a computer to do it for us
How We Construct Sentences

• The Components of Our Grammar
  • Grammar rules
  • Lexical entries
  • Principles
  • Type hierarchy (very preliminary, so far)
  • Initial symbol (S, for now)

• We combine constraints from these components.

• Q: What says we have to combine them?
An Example

A cat slept.

• Can we build this with our tools?
• Given the constraints our grammar puts on well-formed sentences, is this one?
Lexical Entry for \textit{a}

- Is this a fully specified description?
- What features are unspecified?
- How many word structures can this entry license?
Lexical Entry for *cat*

- Which feature paths are abbreviated?
- Is this a fully specified description?
- What features are unspecified?
- How many word structures can this entry license?
Effect of Principles: the SHAC

```
\left[ \begin{array}{c}
\text{word} \\
\text{HEAD} \\
\text{SYN} \\
\text{VAL} \\
\text{SEM} \\
\text{noun} \\
\text{AGR} \\
\text{3sing} \\
\text{GEND} \\
\text{neut} \\
\text{D} \\
\text{SPR} \\
\text{COMPS} \\
\text{MOD} \\
\text{MODE} \\
\text{INDEX} \\
\text{RESTR} \\
\text{INDEX} \\
\text{ref} \\
\text{k} \\
\text{RELN} \\
\text{cat} \\
\text{INSTANCE} \\
\text{k} \\
\end{array} \right]
```
Description of Word Structures for cat
Description of Word Structures for \( a \)
Building a Phrase

[ ]

[ ] [ ]
Constraints Contributed by Daughter Subtrees
Constraints Contributed by the Grammar Rule

[phrase
SYN [ VAL [ SPR ⟨⟩]]]

[7]

[6]

word

[det
HEAD
AGR [ 3sing
GEND neut]
COUNT +
COMPS ⟨⟩]

[7]

SEM

[mode none
INDEX k
RESTR ⟨[RELN a BV k]⟩]

[7]

word

[noun
HEAD
AGR [ 3sing
GEND neut]
COUNT +]

[7]

SEM

[mode ref
INDEX k
RESTR ⟨[RELN INSTANCE k]⟩]
A Constraint Involving the SHAC
Effects of the Valence Principle
Effects of the Head Feature Principle

phrase
  SYN
  VAL
    COMPS
    MOD

word
  SYN
  VAL
    COMPS
    MOD
  MODE none
  INDEX k
  RESTR
    RELN {BV}
      INSTANCE k

word
  SYN
  VAL
    COMPS
    MOD
  MODE ref
  INDEX k
  RESTR
    RELN {cat}
      INSTANCE k
Effects of the Semantic Inheritance Principle
Effects of the Semantic Compositionality Principle
Is the Mother Node Now Completely Specified?

\[
\begin{array}{c}
\text{phrase} \\
\text{SYN} \\
\text{VAL} \\
\text{SEM} \\
\end{array}
\begin{aligned}
&\text{HEAD} \ 6 \\
&\text{SPR} \ (\ ) \\
&\text{COMPS} \ 3 \\
&\text{MOD} \ 4 \\
&\text{MODE} \ 8 \\
&\text{INDEX} \ k \\
&\text{RESTR} \ A \oplus B
\end{aligned}
\]

\[
\begin{array}{c}
\text{word} \\
\text{SYN} \\
\text{VAL} \\
\text{SEM} \\
\end{array}
\begin{aligned}
&\text{HEAD} \ [\text{det} \ 2] \\
&\text{AGR} \ 2 \\
&\text{COUNT} \ + \\
&\text{COMPS} \ (\ ) \\
&\text{SPR} \ (\ ) \\
&\text{MOD} \ (\ ) \\
&\text{MODE} \ \text{none} \\
&\text{INDEX} \ k \\
&\text{RESTR} \ A \langle \text{RELN} \ a \rangle k
\end{aligned}
\]

\[
\begin{array}{c}
\text{word} \\
\text{SYN} \\
\text{VAL} \\
\text{SEM} \\
\end{array}
\begin{aligned}
&\text{HEAD} \ [\text{noun} \ 6] \\
&\text{AGR} \ 2 \\
&\text{3sing \ GEND \ neut} \\
&\text{COMPS} \ (\ ) \\
&\text{SPR} \ (7) \\
&\text{MOD} \ (4) \\
&\text{MODE} \ \text{ref} \\
&\text{INDEX} \ k \\
&\text{RESTR} \ B \langle \text{RELN} \ \text{cat} \ Instance \ k \rangle
\end{aligned}
\]
Lexical Entry for *slept*
Another Head-Specifier Phrase

Key

- HSR
- SHAC
- Val Prin
- HFP
- SIP
- SCP

© 2003 CSLI Publications
Is this description fully specified?

[Diagram of syntactic and semantic structures with labeled parts such as phrase, synonym, valence, composition, mode, index, and restrictions.]
Does the top node satisfy the initial symbol?
RESTR of the S node

\[ \langle \text{RELN BV} \ k, \text{RELN INST} \ k, \text{RELN SLEEPER} \ k, \ldots \rangle \]
Another Example

S

NP

D  NOM

the  N

photos  P

of  NP

the

VP

V  ADV

disappeared  yesterday

D  N

the  suspect
The following diagram illustrates the head features extracted from the lexical entries of the sentence: "the photos of the suspect disappeared yesterday."
the photos of the suspect disappeared yesterday
Valence Features: Lexicon, Rules, and the Valence Principle
Required Identities: Grammar Rules

S

[1] NP

[2] D

the

[3] N

photos

[SPR 2]

[COMPS 3]

[4] PP

disappeared

[SPR 1]

[5] D

of

[SPR 5]

[COMPS 4]

[6] V

yesterday

[SPR 1]

[MOD 6]
Two Semantic Features: the Lexicon & SIP

The photos of the suspect disappeared yesterday.
RESTR Values and the SCP

A ⊕ B ⊕ C ⊕ D ⊕ E ⊕ F ⊕ G

A ⊕ B ⊕ C ⊕ D ⊕ E

A ⊕ B ⊕ C ⊕ D ⊕ E

B ⊕ C ⊕ D ⊕ E

C ⊕ D ⊕ E

D ⊕ E

F ⊕ G

RELN the BV j

RELN photo INST j CONTENT k

RELN disap. SIT s3 D-ER j

RELN yest. ARG s3

RELN the BV k

RELN suspect INST k

the

photos

of

disappeared

yesterday
An Ungrammatical Example

What’s wrong with this sentence?
An Ungrammatical Example

What's wrong with this sentence?

So what?
An Ungrammatical Example

The Valence Principle

*\(\text{S}\)

\(\text{NP}\) [\text{CASE\ acc}] \n
\(\text{them}\)

\(\text{V}\)

\(\text{sent}\)

\(\text{NP}\) [\text{SPR}\ \langle 1 \rangle ]

\(\text{NP}\)

\(\text{NP}\)

\(\text{D}\ \text{a}\)

\(\text{N}\ \text{letter}\)
An Ungrammatical Example

HeadSpecifier Rule

*\( S \)

\[
\text{NP} \quad \begin{array}{c}
\text{[CASE acc]}
\end{array}
\text{VP} \quad \begin{array}{c}
\text{[SPR} \quad \begin{array}{c}
\text{⟨1⟩}
\end{array}
\text{]} \end{array}
\text{NP} \quad \begin{array}{c}
\text{[SPR} \quad \begin{array}{c}
\text{⟨1⟩NP[nom]}\rangle
\end{array}
\text{]} \end{array}
\text{NP} 
\text{NP}
\]

contradiction

\begin{array}{c}
\text{them}
\end{array}
\begin{array}{c}
\text{sent}
\end{array}
\begin{array}{c}
\text{us}
\end{array}
\begin{array}{c}
\text{a letter}
\end{array}
Exercise in Critical Thinking

• Our grammar has come a long way since Ch 2, as we've added ways of representing different kinds of information:
  • generalizations across categories
  • semantics
  • particular linguistic phenomena: valence, agreement, modification

• What else might we add? What facts about language are as yet unrepresented in our model?
Overview

• What we’re trying to do
• The pieces of our grammar
• Two extended examples
• Reflection on what we’ve done, what we still have to do
• Reading questions
• Next time: Catch up & review
• In what way does the actual meaning of the two structures assigned to this sentence differ?

• *We sent two letters to Lee.*

• Are they really both grammatical?
We send two letters to Lee.
We send two letters to Lee.
Reading Questions

• How do we know what features to put into a *predication*?

• Would *letters* as in letters of the alphabet have the same lexical entry as *letters* like what’s usually sent in the mail?

• How do we represent the difference in meaning between *send* and *sent*?
Reading Questions

• How do we get enough different INDEX values for a whole dictionary?
• Why sometimes s and sometimes $s_n$, and not t, u, v?
• How can to be semantically empty and still have a meaningful INDEX value?
• How can the head of a phrase be semantically empty?
• Why does letter share its INDEX with it’s SPR?
Reading Questions

- Does set of well-formed structures correspond exactly to the set of well-formed English sentences?
- Do we have to understand the squiggly bits?
- Why bother formalizing?
- Don’t these feature structures get ridiculously large?
Reading Questions

• Does English have dative case?

• Is it redundant to have a feature CASE for English given that we mostly use prepositions to mark 'case'?

• English nouns (other than pronouns) are underspecified for CASE. How do we figure out their particular CASE values when they are used in a tree?
Reading Questions

• Is position alone enough to tell whether something is SPR or COMPS?

• Will this approach work for morphologically complex languages as well?

• What ever happened to NOM?

• Is it worth memorizing the rules now?

• Why aren’t we using NumP?
Reading Questions

• Is top-down or bottom-up more efficient in actual processing?

• How can we possibly do "simultaneous satisfaction" of all constraints?

• What are the best practices for writing trees going bottom-up (order of things to put in)?
Reading Questions

• Does not having to realize semantic roles mean we can license semantically weird sentences?

• Can we build a grammar that works with more than one sentence at a time? (I.e., paragraphs)