The building blocks of HPSG grammars

In HPSG, sentences, words, phrases, and multisentence discourses are all represented as 

- = complexes of phonological, syntactic/semantic, and discourse information.

We can (and will) view HPSG grammars in two different ways:

1. From a linguistic perspective
2. From a formal perspective

Historical note: HPSG is based on Generalized Phrase Structure Grammar (GPSG) (Gazdar et al. 1985)

HPSG grammars from a linguistic perspective

From a linguistic perspective, an HPSG grammar consists of

a) a lexicon licensing basic words (which are themselves complex objects)
b) lexical rules licensing derived words
c) immediate dominance (id) schemata licensing constituent structure
d) linear precedence (lp) statements constraining word order
e) a set of grammatical principles expressing generalizations about linguistic objects

HPSG (typed) feature structures

HPSG is nonderivational, but in some sense, HPSG has several different levels (layers of features)

- A feature structure is a directed acyclic graph (DAG), with arcs representing features going between values

Each of these feature values is itself a complex object:

- The type sign has the features phon and synsem appropriate for it
- The feature synsem has a value of type synsem
- This type itself has relevant features (local and non-local)

Skeleton of a typed feature structure

In attribute-value matrix (AVM) form, here is the skeleton of an object:

Abbreviated skeleton

Things are often abbreviated when written down (although, the object itself still contains the same things):
An example tree

Let's walk through an example to illustrate how feature structures can be used, starting with this rather impoverished tree:

```
  she
/   \
drinks  wine
```

Some things to note about the tree

- Phonology (phon) is kept separate from syntax and semantics (synsem), allowing different processes to operate on them.
- We say that drinks is a finite verb by specifying its type (verb) and that the value of its vform feature is fin.
- We have some way to say that parts of the tree share identical information, e.g., that a VP and its head daughter V have many of the same properties.
- We use lists to encode subcategorization information, and these items are identified with elements in the tree—note, too, how selection is kept local.

Phrase structure grammar?

Even though it is called Head-driven Phrase Structure Grammar, the name is a misnomer:

- Nothing about the formalism forces you to use PS trees.
- In fact, technically, there are no trees as such, only features which encode objects akin to trees:
  - Types license particular schemata (e.g., head-comps-struc), and a DTRS list keeps track of the constituent daughters.
  - For ease of representation, we often display things in trees.
  - But the example two slides back is more accurately represented as on the next slide.

Lexicalized grammar

How do we start deriving such complex representations?

- One tenet of HPSG (akin to LFG) is that the lexicon contains complex representations of words.
- So, when words are built into phrases, we have all this information at our hands.

We can see this on the lexical entry on the next page, taken from Levine and Meurers (2005):

- For example, we can see that each word relates its syntactic argument structure (valence) with its semantics (content).
Lexical entry for *put*

<table>
<thead>
<tr>
<th>word</th>
<th>&lt;put&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>###</td>
<td></td>
</tr>
<tr>
<td><strong>SYNSEM</strong></td>
<td>LOCAL</td>
</tr>
<tr>
<td>CONTENT</td>
<td>put-relation</td>
</tr>
<tr>
<td>PUM</td>
<td>PPUT</td>
</tr>
</tbody>
</table>

Capturing dependencies

A grammatical framework needs to be able to capture the different grammatical dependencies of natural languages (cf. Levine and Meurers 2005, p. 5)

- **Local dependencies**: limited syntactic domain and largely lexical in nature
- **Non-local dependencies**: arbitrarily large syntactic domain and independent of lexicon

HPSG seems well-suited for this

Local dependencies

As with the other frameworks we've looked at, HPSG deals with local dependencies via the selectional properties of lexical heads (*head-driven*)

For example:

- Raising verbs select for an argument with which they share a subject
- Control (or equi) verbs select for an argument which has a co-indexed subject

Raising verb example

<table>
<thead>
<tr>
<th>word</th>
<th>&lt;seem&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>###</td>
<td></td>
</tr>
<tr>
<td><strong>SYNSEM</strong></td>
<td>LOCAL</td>
</tr>
<tr>
<td>CONTENT</td>
<td>seem'</td>
</tr>
</tbody>
</table>

Non-local dependencies

Instead of using transformations, HPSG analyzes unbounded dependency constructions (UDCs) by linking a filler with a gap

- Analysis relies on the feature **SLASH**
- The general idea is:
  - Trace lexical entry puts its local contents into a non-local **SLASH** set
  - This information is shared among the nodes in a tree
  - When the filler is realized, the information is removed from the **SLASH** set
HPSG grammars from a formal perspective

As with other frameworks we’ve examined, HPSG sets out to model the domain:

- Models of empirically observable objects need to be established, and
- Theories need to constrain which models actually exist.

Thus, from a formal perspective, an HPSG grammar consists of

- the signature as declaration of the domain, and
- the theory constraining the domain.

The signature

- defines the ontology (‘declaration of what exists’):
  - which kind of objects are distinguished, and
  - which properties of which objects are modeled.
- consists of
  - the type (or sort) hierarchy and
  - the appropriateness conditions, defining which type has which appropriate attributes (or features) with which appropriate values.
  - Some atomic types have no feature appropriate for them

Example excerpt of a signature

Here, we leave out the appropriateness conditions and just show a hierarchy of types

```
object
  +
  boolean
  -
  head

  subst(antive)
   -
   func(tional)

  noun
  verb
  ...
  marker
  determiner
```

Models of linguistic objects

- As mentioned, the objects are modelled by feature structures, which are depicted as directed graphs.
- Since these models represent objects in the world (and not knowledge about the world), they are total with respect to the ontology declared in the signature. Technically, one says that these feature structures are
  - totally well-typed: Every node has all the attributes appropriate for its type and each attribute has an appropriate value.
  - sort-resolved: Every node is of a maximally specific type.

Structure sharing

The main explanatory mechanism in HPSG is that of structure-sharing, equating two features as having the exact same value (token-identical)

```
word
PRON
<walk> <walk>
SYNSEM|LOC
CAT|SUBCAT
NP|nom
3rd|sing
CONTENT
walk'
WALKER
```

The index of the NP on the SUBCAT list is said to unify with the value of LAUGHER.
Descriptions

A description language and its abbreviating attribute-value matrix (AVM) notation is used to talk about sets of objects. Descriptions consist of three building blocks:

- **Type** descriptions single out all objects of a particular type, e.g., word
- **Attribute-value pairs** describe objects that have a particular property. The attribute must be appropriate for the particular type of object, and the value can be any kind of description, e.g., \[ \text{spous} \ \text{name mary} \]
- **Tags** (structure sharing) to specify token identity, e.g., \[ \text{I} \]

Description example

A verb, for example, can specify that its subject be masculine singular (as Russian past tense verbs do):

(2) a. Ya spal
   \[ \text{masc, sg} \ \text{slept} \text{masc, sg} \]
   b. On spal
   \[ \text{he masc, sg} \ \text{slept} \text{masc, sg} \]

(3) On the verb's SUBJ list:

\[
\begin{array}{c|c|c|c|c|c|c}
\text{word} & \text{SYNSEM LOC} & \text{cat[head noun}} & \text{content index} & \text{num sing} & \text{gen masc} \\
\end{array}
\]

This doesn’t specify the entire (totally well-typed) feature structure, just what needs to be true in the feature structure.

Subsumption example

The description in (3) is said to **subsume** both of the following more specific (partial) feature structures:

(4) a.word
   \[ \text{SYNSEM LOC} & \text{cat[head noun}} & \text{content index} & \text{frst num sing} & \text{gen masc} \\
   \]
   b. word
   \[ \text{SYNSEM LOC} & \text{cat[head noun}} & \text{content index} & \text{frst num sing} & \text{gen masc} \\
   \]

Subsumption

Feature structure descriptions have **subsumption** relations between them.

- A more general description subsumes a more specific one.
- A more general description usually means that less features are specified.

HPSG from a linguistic perspective (again)

Now that we have these feature structures, how do we use them for linguistic purposes?

1. Specify a signature/ontology which allows us to make linguistically-relevant distinctions and puts appropriate features in the appropriate places
2. Specify a theory which constrains that signature for a particular language
   - Lexicon specifies each word and the different properties that it has
     - There can also be relations (so-called lexical rules) between words in the lexicon
   - Phrasal rules, or principles, allow words to combine into phrases
A tour of Pollard and Sag (1994)

We’ll start with the signature and theory from Pollard and Sag (1994).

In the next series of slides, you should:

- begin to understand what everything means
- begin to understand the connection between linguistic theory and its formalization in HPSG
- begin to gain an appreciation for a completely worked-out theory

Why the complicated structure?

- LOCAL & NONLOCAL: Most linguistic constructions can be handled locally, but non-local constructions (e.g., extraction) require different mechanisms
- CATEGORY, CONTENT, and CONTEXT: roughly, these correspond to syntactic, semantic, and pragmatic notions, all of which are locally determined
- HEAD and SUBCAT: a word’s syntactic information comes in two parts: its own lexical information (part of speech, etc.) and information about its arguments

Properties of particular part-of-speech

What SUBCAT does

The SUBCAT list can be thought of as akin to a word’s valency requirements

- Items on the SUBCAT list are ordered by obliqueness—akin to LFG—not necessarily by linear order
- The SUBCAT Principle, described below, will describe a way for a word to combine with its arguments– That is, we will still need a way to go from the SUBCAT specification to some sort of tree structure

NB: Here, we will use a single SUBCAT list, but later we will switch to a VALENCE feature, which contains both a SUBJ and COMPS list
Locality of SUBCAT

SUBCAT selects a list of SYNSEM values, not SIGN values.

- If you work through the ontology, this means that a word does not have access to the DTRS list of items on its own SUBCAT list.
- Intuitively, this means that a word cannot dictate properties of the daughters of its daughters.

⇒ Constructions are thus restricted to local relations.

CONTENT information

The CONTENT feature specifies different semantic information.

- A feature appropriate for nominal-object objects (a subtype of content objects) is INDEX (as shown on the next slide).
- Agreement features can be stated through the INDEX feature.
- Note that CASE was put somewhere else (within HEAD), so CASE agreement is treated differently than person, number, and gender agreement (at least in English).

Semantic representations

Indices

Abbreviations for describing lists

Attention: 〈 〉 and 〈 〉 describe all lists of length one!
Abbreviations of common AVMs

Pollard and Sag (1994) use some abbreviations to describe synsem objects:

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Abbreviated AVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP</td>
<td>synsem</td>
</tr>
<tr>
<td></td>
<td>LOCAL</td>
</tr>
<tr>
<td></td>
<td>CATEGORY</td>
</tr>
<tr>
<td></td>
<td>HEAD verb</td>
</tr>
<tr>
<td></td>
<td>CONTENT index</td>
</tr>
<tr>
<td></td>
<td>SUBCAT ()</td>
</tr>
<tr>
<td>S</td>
<td>synsem</td>
</tr>
<tr>
<td></td>
<td>LOCAL</td>
</tr>
<tr>
<td></td>
<td>CATEGORY</td>
</tr>
<tr>
<td></td>
<td>HEAD verb</td>
</tr>
<tr>
<td></td>
<td>CONTENT</td>
</tr>
<tr>
<td></td>
<td>SUBCAT ()</td>
</tr>
<tr>
<td>VP</td>
<td>synsem</td>
</tr>
<tr>
<td></td>
<td>LOCAL</td>
</tr>
<tr>
<td></td>
<td>CATEGORY</td>
</tr>
<tr>
<td></td>
<td>HEAD verb</td>
</tr>
<tr>
<td></td>
<td>CONTENT</td>
</tr>
<tr>
<td></td>
<td>SUBCAT (synsem)</td>
</tr>
</tbody>
</table>

The Lexicon

The basic lexicon is defined by the Word Principle as part of the theory. It defines which of the ontologically possible words are grammatical:

\[ \text{word} \rightarrow \text{lexical-entry}_1 \lor \text{lexical-entry}_2 \lor \ldots \]

with each of the lexical entries being descriptions, such as e.g.:

\[
\begin{align*}
\text{word} & \rightarrow \text{phon} \quad \langle \text{laughs} \rangle \\
\text{SYNSEM LOC} & \rightarrow \text{CAT} \quad \text{NP} (\text{nom} \langle 3 \text{rd}, \text{sing} \rangle, \text{NP} [\text{acc} \langle] \\
\text{cont} & \rightarrow \text{content} \quad \langle \text{laughs} \rangle \\
\end{align*}
\]

An example lexicon

\[
\begin{align*}
\text{word} & \rightarrow \text{phon} \quad \langle \text{laughs} \rangle \\
\text{SYNSEM LOC} & \rightarrow \text{CAT} \quad \text{NP} (\text{nom} \langle 3 \text{rd}, \text{sing} \rangle, \text{NP} [\text{acc} \langle] \\
\text{cont} & \rightarrow \text{content} \quad \langle \text{laughs} \rangle \\
\end{align*}
\]

\[
\begin{align*}
\text{word} & \rightarrow \text{phon} \quad \langle \text{laughs} \rangle \\
\text{SYNSEM LOC} & \rightarrow \text{CAT} \quad \text{NP} (\text{nom} \langle 3 \text{rd}, \text{sing} \rangle, \text{NP} [\text{acc} \langle] \\
\text{cont} & \rightarrow \text{content} \quad \langle \text{laughs} \rangle \\
\end{align*}
\]
**Types of phrases**

In order to put words from our lexicon into a sentence, we have to define what makes an acceptable sentence structure:

- Each phrase has a `dtrs` attribute (words do not have this attribute), which has a `constituent-structure` value.
- This `dtrs` value loosely corresponds to what we normally view in a tree as daughters.
  - Additionally, “tree branches” contain grammatical role information (adjunct, complement, etc.).
- By distinguishing different kinds of `constituent-structures`, we define what kinds of phrases exist in a language.

**Universal Principles**

But how exactly did that last example work?

- `drinks` has head information specifying that it is a verb and so forth, and it also has subcategorization information specifying that it needs a subject and an object.
  - The head information gets percolated up (The HEAD Principle).
  - The subcategorization information gets “checked off” as you move up in the tree (The SUBCAT Principle).

Such principles are treated as linguistic universals in HPSG.

**Head-Feature Principle:**

- In prose: The HEAD feature of any headed phrase is structure-shared with the HEAD value of the head daughter.
- Specified as a constraint:

\[
\text{phrase}[\text{dtrs head-structure}] \rightarrow \text{SYNSEM}[\text{LOC|CAT|HEAD } \text{dtrs|HEAD-DTR|SYNSEM|LOC|CAT|HEAD } \text{E}]
\]
Subcat Principle:

In a headed phrase, the SUBCAT value of the head daughter is the concatenation of the phrase’s SUBCAT list with the list (in order of increasing obliqueness of SYNSEM values of the complement daughters).

\[
\text{DTRS headed-structure} \rightarrow \begin{cases} 
\text{SYNSEM[LOC][CAT][SUBCAT]} & \text{DTRS} \\
\text{HEAD-DTR[SYNSEM][LOC][CAT][SUBCAT]} & \oplus \oplus \oplus \\
\text{COMP-DTRS[SYNSEM2sign]} & \oplus \oplus \oplus 
\end{cases}
\]

with \(\oplus\) standing for list concatenation, i.e., \(\text{append}\), defined as follows

\[
\text{e-list} \oplus \oplus := \begin{cases} 
\text{FIRST} & \oplus \oplus \\
\text{REST} & \oplus \oplus \oplus \oplus 
\end{cases}
\]

Immediate Dominance (ID) Schemata

- There is an inventory of valid ID schemata in a language
- Every headed phrase must satisfy exactly one of the ID schemata
  - Which ID schema is used depends on the type of the DTRS attribute
  - This goes back to the ontology of phrases we saw earlier

Formally, though, these constraints are phrased as the universal principles were

Immediate Dominance Principle (for English):

\[
\text{phrase [DTRS headed-struct]} \rightarrow \begin{cases} 
\text{SYNSEM[LOC][CAT]} & \text{HEAD} & \text{ verb} & \text{word} \\
\text{SUBCAT} & \oplus & \oplus & \oplus \\
\text{head-comps-struc} & \oplus & \oplus & \oplus \\
\text{HEAD} & \oplus & \oplus & \oplus \\
\text{word} & \oplus & \oplus & \oplus 
\end{cases}
\]

(Head-Subject)

\[
\vdash \oplus \oplus 
\]

Fallout from these Principles

- Note that agreement is handled neatly, simply by the fact that the SYNSEM values of a word’s daughters are token-identical to the word’s SUBCAT items.

One question remains before we can get the structure we have above:

- How exactly do we decide on a syntactic structure?
- i.e., Why is it that the object was checked off low and the subject was checked off at a higher point?

Answer: because of the ID schemata used

Towards Head Adjunct Structures

Lexical entry of an attributive adjective

So, in the example of She drinks wine, the DTRS value over drinks wine is a head-comps-struc, while the DTRS over the whole sentence is a head-subj-struc
Lexical entry of an attributive adjective
Version without redundant specifications

SPEC Principle:

Marking Principle:
Lexical entry of the marker *that*

```
PHON <that>
SYNSEM|LOC|CAT
HEAD SPEC LOC|CAT SUBCAT ⟨⟩ MARKING that
```

Sketch of an example for a head-marker structure

```
PHON <that,John,laughs>
SYNSEM|LOC|CAT
HEAD SPEC LOC|CAT SUBCAT ⟨⟩ MARKING unmarked
```

A few more points on HPSG

- We can view a grammar as a set of constraints: formulas which have to be true in order for a feature structure to be well-formed.
  - With such a view, parsing with HPSG falls into the realm of constraint-based processing.
- Two important points about relating descriptions are subsumption and unification, loosely defined as:
  - **subsumption**: the description $F$ subsumes the description $G$ iff $G$ entails $F$; i.e., $F$ is more general than $G$.
  - **unification**: the description of $F$ and $G$ unify iff their values are compatible.
- **Closed World Assumption**: there are no linguistic species beyond what is specified in the type hierarchy.