HPSG

Ivan Sag, Tom Wasow, & Emily Bender
The Course Web Site

http://hpsg.stanford.edu/05inst/
## Schedule of Classes

<table>
<thead>
<tr>
<th>Date</th>
<th>Topic</th>
<th>Readings</th>
</tr>
</thead>
<tbody>
<tr>
<td>28 June</td>
<td>Psycholinguistic Motivation</td>
<td>SWB Ch.9, pp. 294–309</td>
</tr>
<tr>
<td></td>
<td>Feature Structure Models</td>
<td>SWB Ch.3, pp. 49–74</td>
</tr>
<tr>
<td>30 June</td>
<td>Valence, Part of Speech and Phrasal Projection</td>
<td>SWB Ch.4, pp. 93–122.</td>
</tr>
<tr>
<td></td>
<td>Signs and Compositionality</td>
<td>SWB Ch. 5.</td>
</tr>
<tr>
<td>5 July</td>
<td>Overview</td>
<td>SWB Ch. 6, pp. 165–191.</td>
</tr>
<tr>
<td></td>
<td>Binding Theory</td>
<td>SWB Ch. 7, pp. 203–215.</td>
</tr>
<tr>
<td></td>
<td>Implementing HPSG</td>
<td></td>
</tr>
<tr>
<td>7 July</td>
<td>The Lexicon</td>
<td>SWB Ch. 8.</td>
</tr>
<tr>
<td>12 July</td>
<td>Passive, Raising, Control</td>
<td>SWB Ch. 10; Ch. 11, pp. 333–345; Ch. 12,</td>
</tr>
<tr>
<td>14 July</td>
<td>Filler-Gap Constructions</td>
<td>SWB Ch. 14.</td>
</tr>
<tr>
<td></td>
<td>Wrap-Up</td>
<td></td>
</tr>
</tbody>
</table>
What does grammar have to do with psychology?

Three ways it could be relevant:

* It provides insight into how children acquire language.
* It provides insight into how speakers produce utterances.
* It provides insight into how listeners understand utterances.
Mainstream position:

- Grammar represents knowledge of language ("competence").
- This is distinct from use of language ("performance").
- We can draw a strong conclusion about language acquisition, namely, most grammatical knowledge is innate and domain-specific.
- Serious study of language use (production and comprehension) depends on having a well-developed theory of competence.
Brief remarks on language acquisition

- Idea of an innate language organ is controversial
  - It is based on the “poverty of the stimulus” argument, and a model of learning as hypothesis testing.
  - The environment may be more informative than this assumes.
  - There may be more powerful learning methods than this assumes.
- There has not been much work on language acquisition using constraint-based lexicalist theories like ours; but
  - Explicit formulation is a prerequisite for testing learning models
  - Our feature structures could model richer context information.
- We are open-minded with respect to this controversy.
Production

- People plan their utterances to some extent, but typically don’t have sentences fully formulated before they start uttering.

- This is evident from the existence of disfluencies, such as *uh, um*, repetitions, false starts, etc.
A disfluent utterance

because you see I, uh, some of our people, [pause and clears throat] who are doing LEs, uumm, have to consider which paper [pause] to do,
Disfluencies are sensitive to structure:

Repeat rate of *the* varies with position and complexity of the NP it introduces:

![Bar chart showing repeat rate of *the* in different positions and complexities of the NP.](chart.png)
Production errors are sensitive to syntactic structure

Agreement errors are more common with PP complements than sentential complements: errors like (2) are significantly more common than errors like (1).

(1) *The only generalization I would dare to make about our customers are that they’re pierced.

Bock & Cutting (1992)

vs.

(2) *The only generalization about our customers are that they’re pierced.
Some high-level sentence planning is necessary, too

• *Ich habe dem Mann, den ich gesehen habe geholfen.*
  I have the-dat man who-acc I seen have helped
  “I helped the man I saw”

• *Ich habe den Mann, dem ich geholfen habe gesehen.*
  I have the-acc man who-dat I helped have seen.
  “I saw the man I helped”

• The choice between *dem* and *den* depends on the choice of verbs several words later.
A production model should allow interaction of top-down and left-to-right information

- Grammar plays a role in production.
- Partial grammatical information should be accessible by the production mechanism as needed.
- This argues against grammatical theories that involve sequential derivations with fixed ordering.
- Our theory of grammar has the requisite flexibility.
Comprehension

- Early work tried to use transformational grammar in modeling comprehension

- The Derivational Theory of Complexity: The psychological complexity of a sentence increases with the number of transformations involved in its derivation.

- Initial results seemed promising, but later work falsified the DTC.
Some relevant quotes

- “The results show a remarkable correlation of amount of memory and number of transformations”
  – Chomsky, 1968

- “[I]nvestigations of DTC…have generally proved equivocal. This argues against the occurrence of grammatical derivations in the computations involved in sentence recognition”
  – Fodor, Bever, & Garrett, 1974
Another quote

• “Experimental investigations of the psychological reality of linguistic structural descriptions have…proved quite successful.”  
  – Fodor, Bever, & Garrett, 1974

• In particular, they concluded that “deep structures” and “surface structures” were psychologically real, but the transformations relating them weren’t.
Early Evidence for the Psychological Reality of Deep Structures

- The proposed DS for (2) had three occurrences of *the detective*, while the proposed DS for (1) had only two:

  1. *The governor asked the detective to prevent drinking.*
  2. *The governor asked the detective to cease drinking.*

- In a recall experiment, *detective* was significantly more effective in prompting people to remember (2) than (1)
Typical Problem Cases for the DTC

(1) *Pat swam faster than Chris swam.*
(2) *Pat swam faster than Chris did.*
(3) *Pat swam faster than Chris.*

• The DTC predicts that (1) should be less complex than (2) or (3), because (2) and (3) involve an extra deletion transformation.

• In fact, subjects responded more slowly to (1) than to either (2) or (3).
What should a psychologically real theory of grammar be like?

- The “deep structure” distinctions that are not evident on the surface should be represented.
- The transformational operations relating deep and surface structures should not be part of the theory.
- Our information-rich trees include all of the essential information in the traditional deep structures, but without the transformations.
Jerry Fodor claims the human mind is “modular”

“A module is…an informationally encapsulated computational system -- an inference-making mechanism whose access to background information is constrained by general features of cognitive architecture.”
-- Fodor, 1985

A central issue in psycholinguistics over the past 20 years has been whether language is processed in a modular fashion.
Semantic Incrementality

• Listeners construct partial interpretations of utterances based on what they have heard so far.

• This is based on partial syntactic analysis, plus knowledge of lexical semantics (and common sense):

  I found the book on the \{ coffee table, water table \}

• This observation suggests that processing is incremental, with syntactic and semantic processing tightly linked -- i.e., non-modular
Tanenhaus’s Eye-Tracking Experiments

- Participants wear a device on their heads that makes a video showing exactly what they’re looking at.
- They listen to spoken instructions and carry out various tasks.
- The eye-tracking provides evidence of the cognitive activity of participants that can be correlated with the linguistic input.
Non-linguistic visual information affects lexical access

- Participants’ gaze settled on a referent before the word was completed, unless the initial syllable of the word was consistent with more than one object.

- For example, participants’ gaze rested on the pencil after hearing 
  *Pick up the pencil* 
  more slowly when both a pencil and a penny were present.
Non-linguistic visual information affects syntactic processing

• Eye movements showed that people hearing (1) often temporarily misinterpreted *on the towel* as the destination.

  (1) *Put the apple on the towel in the box.*

• When *on the towel* helped them choose between two apples, such misparses were significantly less frequent than when there was only one apple.
General Conclusion of Eye-Tracking Studies

- People use whatever information is available as soon as it is useful in interpreting utterances.

- This argues against Fodorian modularity.

- It argues for a model of language in which information is represented in a uniform, order-independent fashion.
Speakers know a great deal about individual words

- Individual lexical items have many idiosyncrasies in where they can occur, and in where they tend to occur.

- For example, the verb *behoove* occurs only with the subject *it* (and only in certain verb forms), and the verb *beware* has only the base form.

- We also know that the transitive use of *walk* is much rarer than the intransitive.
V-NP-NP vs. V-NP-PP Frequency in the NYT

<table>
<thead>
<tr>
<th>Verb</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>tell</td>
<td>100%</td>
</tr>
<tr>
<td>give</td>
<td>75%</td>
</tr>
<tr>
<td>show</td>
<td>50%</td>
</tr>
<tr>
<td>hand</td>
<td>25%</td>
</tr>
<tr>
<td>fax</td>
<td>10%</td>
</tr>
<tr>
<td>bring</td>
<td>10%</td>
</tr>
<tr>
<td>send</td>
<td>0%</td>
</tr>
<tr>
<td>sell</td>
<td>0%</td>
</tr>
</tbody>
</table>
Lexical biases influence processing

- Arnold, et al (2004) ran a production experiment to test whether ambiguity avoidance would influence speakers’ choice between (1) and (2):

  (1) *They gave Grant’s letters to Lincoln to a museum.*
  (2) *They gave a museum Grant’s letters to Lincoln.*

- Lexical bias of the verbs turned out to be a significant predictor of which form speakers used (and ambiguity avoidance turned out not to be).
A psychologically real grammar should be lexicalist

- Early generative grammars downplayed the lexicon.
- Now, however, the importance of the lexicon is widely recognized.
- This aspect of grammar has been developed in greater detail in our theory than in any other.
- It would be easy to add frequency information to our lexicon, though there is debate over the wisdom of doing so.
Conclusion

- Grammatical theory should inform and be informed by psycholinguistic experimentation and corpus studies.

- This has happened less than it should have.

- Existing psycholinguistic evidence favors a constraint-based, lexicalist approach (like ours).
and now to our theory.....
Problems with Context-free Grammar

- Potentially arbitrary rules
- Gets clunky quickly with cross-cutting properties
- Not quite powerful enough for natural languages

Solution: Replace atomic node labels with feature structures.
## Cross-cutting Grammatical Properties

<table>
<thead>
<tr>
<th></th>
<th>3rd singular subject</th>
<th>plural subject</th>
</tr>
</thead>
<tbody>
<tr>
<td>direct object NP</td>
<td>denies</td>
<td>deny</td>
</tr>
<tr>
<td>no direct object NP</td>
<td>disappears</td>
<td>disappear</td>
</tr>
</tbody>
</table>
Two Kinds of Language Models

• Speakers’ internalized knowledge (their grammar)

• Set of sentences in the language
Things Involved in Modeling Language

- Real world entities
- Models
- Descriptions of the models
Feature Structure Descriptions

\[
\begin{bmatrix}
\text{FEATURE}_1 & \text{VALUE}_1 \\
\text{FEATURE}_2 & \text{VALUE}_2 \\
\vdots \\
\text{FEATURE}_n & \text{VALUE}_n
\end{bmatrix}
\]
A Lexical Entry

\[\langle \text{bird} , \begin{bmatrix} \text{POS} & \text{noun} \\ \text{NUM} & \text{sg} \end{bmatrix} \rangle\]
Some Entities

entity
NAME   Stanford University
TEL     650-723-2300

entity
NAME   John Hennessy
TEL     650-723-2481

entity
NAME   Stanford Linguistics
TEL     650-723-4284
Type Hierarchies

A type hierarchy....

• ... states what kinds of objects we claim exist (the types)

• ... organizes the objects hierarchically into classes with shared properties (the type hierarchy)

• ... states what general properties each kind of object has (the feature and feature value declarations).
A Type Hierarchy

entity

[NAME, TEL]

organization

[FOUNDERS]

university

[FOUNDERS->PRESIDENT]

department

[individual]

[BIRTHDAY]

[FOUNDERS->CHAIR]
Our Entities Again

\[
\begin{align*}
\text{university} & \\
\text{NAME} & \text{Stanford University} \\
\text{FOUNDERS} & \langle \text{Leland Stanford, Jane Stanford} \rangle \\
\text{PRESIDENT} & \text{John Hennessy} \\
\text{TEL} & 650-723-2300 \\
\end{align*}
\]

\[
\begin{align*}
\text{individual} & \\
\text{NAME} & \text{John Hennessy} \\
\text{BIRTHDAY} & 9-22-52 \\
\text{TEL} & 650-723-2481 \\
\end{align*}
\]

\[
\begin{align*}
\text{department} & \\
\text{NAME} & \text{Stanford Linguistics} \\
\text{FOUNDERS} & \langle \text{Joseph Greenberg, Charles Ferguson} \rangle \\
\text{CHAIR} & \text{Beth Levin} \\
\text{TEL} & 650-723-4284 \\
\end{align*}
\]
university
NAME Stanford University

FOUNDERS
\(\langle\text{individual} \begin{bmatrix} \text{NAME} & \text{Leland Stanford} \end{bmatrix}, \text{individual} \begin{bmatrix} \text{NAME} & \text{Jane Stanford} \end{bmatrix}\rangle\)

PRESIDENT
\(\begin{bmatrix} \text{individual} \text{NAME} & \text{John Hennessy} \\ \text{BIRTHDAY} & 9-22-52 \\ \text{TEL} & 650-723-2481 \end{bmatrix}\)

TEL 650-723-2300
department
NAME Stanford Linguistics

FOUNDERS

(individual
NAME Joseph Greenberg
BIRTHDAY 5-28-15
)

(individual
NAME Charles Ferguson
BIRTHDAY 7-6-21
)

CHAIR

(individual
NAME Beth Levin
BIRTHDAY 8-26-55
TEL 650-723-4284
)

TEL 650-723-4284
<table>
<thead>
<tr>
<th>TYPE</th>
<th>FEATURES/VALUES</th>
<th>IST</th>
</tr>
</thead>
<tbody>
<tr>
<td>entity</td>
<td>[NAME  string            ]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[TEL  number             ]</td>
<td></td>
</tr>
<tr>
<td>organization</td>
<td>[FOUNDERS  list(individual)]</td>
<td>entity</td>
</tr>
<tr>
<td>university</td>
<td>[PRESIDENT  individual   ]</td>
<td>organization</td>
</tr>
<tr>
<td>department</td>
<td>[CHAIR  individual       ]</td>
<td>organization</td>
</tr>
<tr>
<td>individual</td>
<td>[BIRTHDAY  date          ]</td>
<td>entity</td>
</tr>
</tbody>
</table>
Tags

department : \[ \begin{bmatrix} TEL & 1 \\ CHAIR & \begin{bmatrix} TEL & 1 \end{bmatrix} \end{bmatrix} \]
Combining Constraints

\[
\begin{bmatrix}
\text{TEL} & 650-723-4284 \\
\end{bmatrix}
\&
\begin{bmatrix}
\text{NAME} & \text{Stanford Linguistics} \\
\end{bmatrix}
\]

= 

\[
\begin{bmatrix}
\text{NAME} & \text{Stanford Linguistics} \\
\text{TEL} & 650-723-4284 \\
\end{bmatrix}
\]
Incompatible Constraints

\[
\begin{bmatrix}
\text{university} \\
\text{NAME} & \text{Stanford University}
\end{bmatrix}
\quad \text{and} \quad
\begin{bmatrix}
\text{university} \\
\text{NAME} & \text{Harvard University}
\end{bmatrix}
\]

cannot be combined; neither can the following:

\[
\begin{bmatrix}
\text{department} \\
\text{TEL} & 650-555-4284
\end{bmatrix}
\]

\[
\begin{bmatrix}
\text{individual} \\
\text{TEL} & 650-555-4284
\end{bmatrix}
\]
But is compatible with any of the following:

\[
\text{university} \]

\[
\text{individual} \\
\text{NAME} \quad \text{Sailor Moon}
\]

\[
\text{department} \\
\text{NAME} \quad \text{Metaphysics} \\
\text{CHAIR} \quad \text{Alexius Meinong, Jr.}
\]
For example,

\[
\begin{bmatrix}
\text{individual} \\
\text{NAME} & \text{Sailor Moon} \\
\text{TEL} & 888-234-5789
\end{bmatrix}
\]
Another Pair of Incompatible Constraints

\[
\begin{bmatrix}
\text{BIRTHDAY} & 10-10-1973 \\
\text{PRESIDENT} & \text{individual} \\
\text{NAME} & \text{Sailor Moon}
\end{bmatrix}
\]
Equivalent Feature Structures

\[
\begin{bmatrix}
\text{TEL} & 1650-723-4284 \\
\text{CHAIR} & \begin{bmatrix} \text{TEL} & 1 \end{bmatrix}
\end{bmatrix}
\]

\[
\begin{bmatrix}
\text{TEL} & 1 \\
\text{CHAIR} & \begin{bmatrix} \text{TEL} & 1650-723-4284 \end{bmatrix}
\end{bmatrix}
\]

\[
\begin{bmatrix}
\text{CHAIR} & \begin{bmatrix} \text{TEL} & 1 \end{bmatrix}
\end{bmatrix}
\]

\[
\begin{bmatrix}
\text{TEL} & 1650-723-4284 \\
\text{CHAIR} & \begin{bmatrix} \text{TEL} & 279650-723-4284 \end{bmatrix}
\end{bmatrix}
\]

\[
\begin{bmatrix}
\text{TEL} & 279 \\
\text{CHAIR} & \begin{bmatrix} \text{TEL} & 650-723-4284 \end{bmatrix}
\end{bmatrix}
\]

\[
\begin{bmatrix}
\text{TEL} & \mathbb{N} \\
\text{CHAIR} & \begin{bmatrix} \text{TEL} & \mathbb{N}650-723-4284 \end{bmatrix}
\end{bmatrix}
\]
The defendant disappeared.
The Beginnings of Our Type Hierarchy

$feature - structure$

$expression \ldots$

$word$ $phrase$
A Feature for Part of Speech

\[
NP = \begin{bmatrix}
\text{phrase} \\
\text{HEAD} & \text{noun}
\end{bmatrix}
\]

\[
\langle \text{bird}, \begin{bmatrix}
\text{word} \\
\text{HEAD} & \text{noun}
\end{bmatrix} \rangle
\]
Type Hierarchy for Parts of Speech I

expression

feat − struc

pos

word phrase noun verb det prep adj conj
Type Hierarchy for Parts of Speech II

```
feat - struc

expression

[HEAD]

word phrase

agr-pos

[AGR]

noun verb

[DET]

AUX

head

prep adj conj
```

© 2003 CSLI Publications
Underspecification

\[ V = \begin{bmatrix} \text{word} \\ \text{HEAD} & \text{verb} \end{bmatrix} \]

\[ \text{VP} = \begin{bmatrix} \text{phrase} \\ \text{HEAD} & \text{verb} \end{bmatrix} \]

\[ \begin{bmatrix} \text{HEAD} & \text{verb} \end{bmatrix} \]
Formalizing the Notion of Head

- Expressions have a feature \text{HEAD}
- \text{HEAD}'s values are of type \text{pos}
- For \text{HEAD} values of type \text{agr-cat}, \text{HEAD}'s value also includes the feature \text{AGR}
- Well-formed trees are subject to the Head Feature Principle
The Head Feature Principle

• Intuitive idea: Key properties of phrases are shared with their heads
• The HFP: In any headed phrase, the HEAD value of the mother and the head daughter must be identical.
• Sometimes described in terms of properties “percolating up” or “filtering down”, but this is just metaphorical talk
A Tree is Well-Formed if …

• It and each subtree are licensed by a grammar rule or lexical entry
• All general principles (like the HFP) are satisfied.
• NB: Trees are part of our model of the language, so all their features have values (even though we will often be lazy and leave out the values irrelevant to our current point).