Ling 566
Nov 2, 2023

Grammar and Processing

Overview

- Psycholinguistics and grammar design
  - What grammar has to say
  - What psychological evidence has to say
    - Acquisition
    - Production
    - Comprehension
- Universals

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.

Our model: Key characteristics

- Surface-oriented
- Constraint-based
- Lexicalist

Chomsky's 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 task-specific.
- Serious study of language use (production and comprehension) depends on having a well-developed theory of competence.

Brief remarks on language acquisition

- Chomsky's nativism is very 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 he assumes.
  - There may be more powerful learning methods than he assumes.

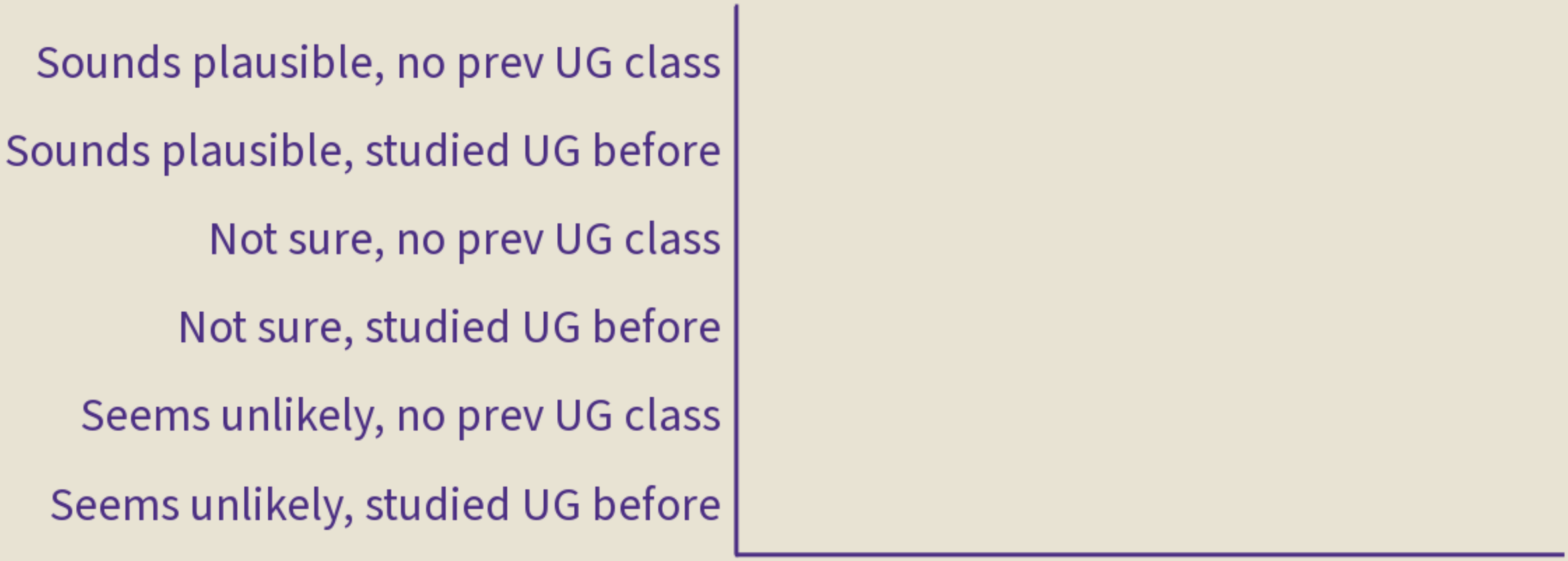
Brief remarks on language acquisition

- Chomsky's nativism is very 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 he assumes.
  - There may be more powerful learning methods than he 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're neutral with respect to this controversy.

When poll is active, respond at **pollev.com/emb**

Text **EMB** to **22333** once to join

Where do you currently stand wrt the hypothesis of innate UG?



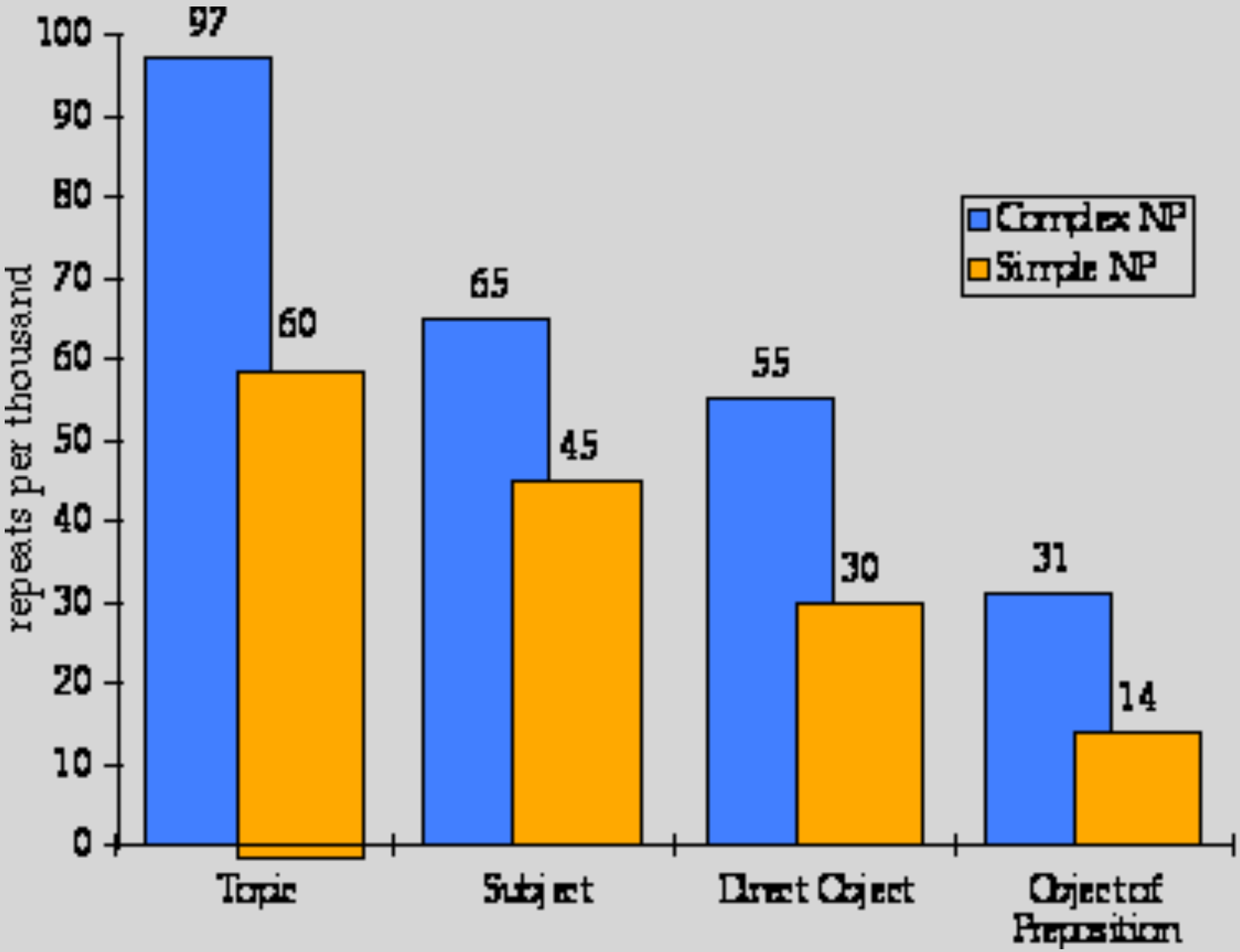
Total Results: 0

Production and Grammar

- Evidence for left-to-right effects
- Evidence for grammar in processing
- Evidence for top-down planning

Disfluencies are sensitive to structure:

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



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 claim that the wolves had raised the babies were rejected*.

vs.

(2) **The claim about the newborn babies were rejected*.

So why?

- Speculation: Clauses are their own agreement domains, so people don't mistake an NP in a lower clause as a trigger for agreement
- Original work: Kay Bock (1980s).

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 in at least the 1980s & 1990s years was whether language is processed in a modular fashion.

Tanenhaus's Eye-Tracking Experiments

- Participants wear a device on their heads that makes a videotape showing exactly what they're looking at.
- They listen to spoken instructions and carry out various tasks.
- They 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.

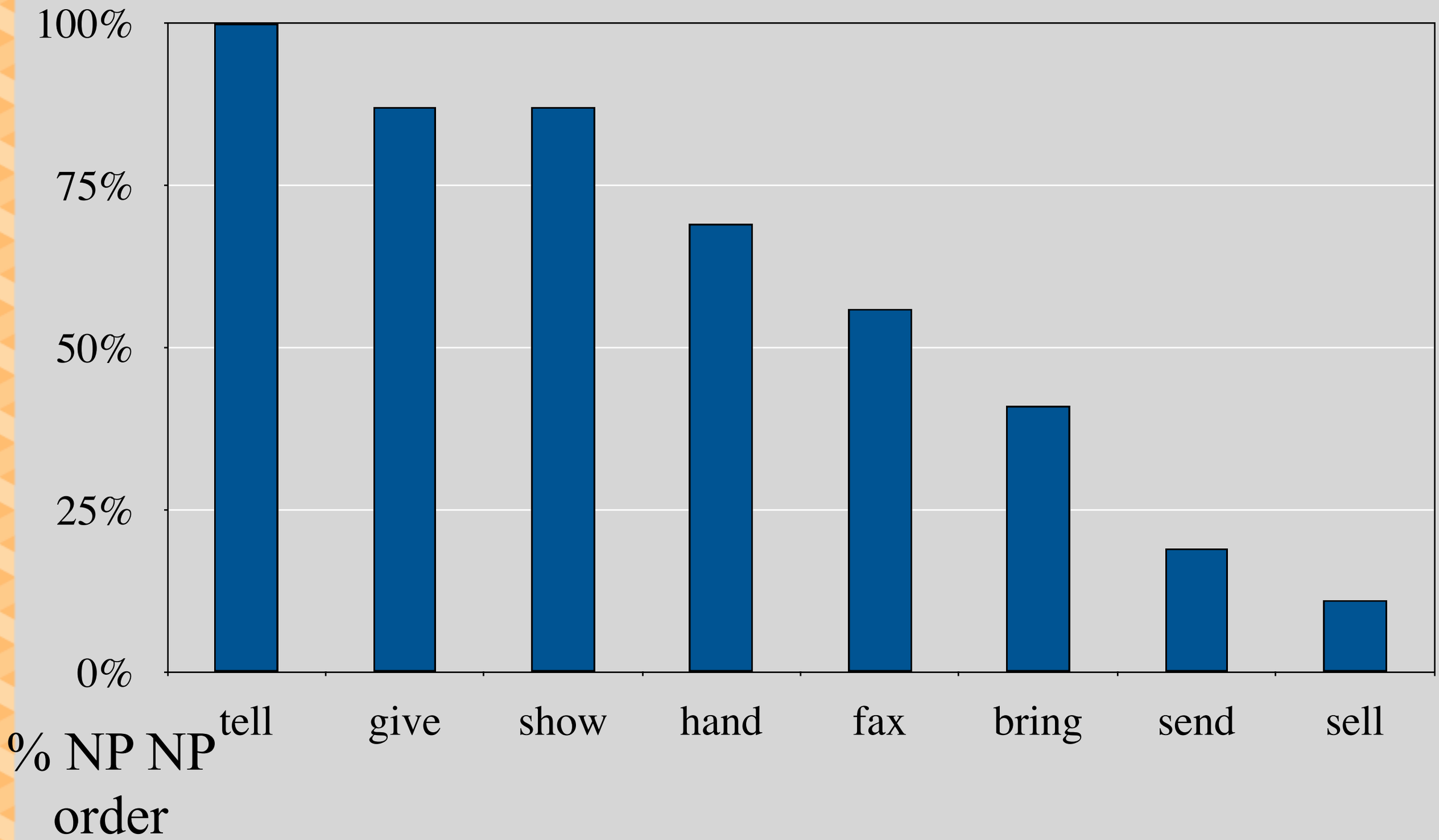
Speakers know a great deal about individual words

- 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*



Lexical biases influence processing

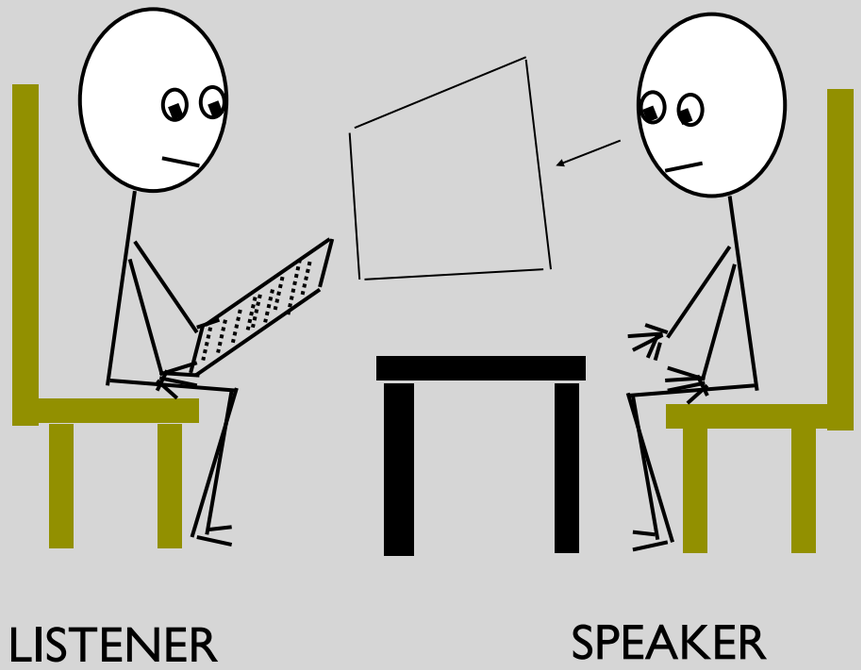
- Wasow et al ran a production experiment to test whether ambiguity avoidance would influence speakers' choice between (1) and (2):

  (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).

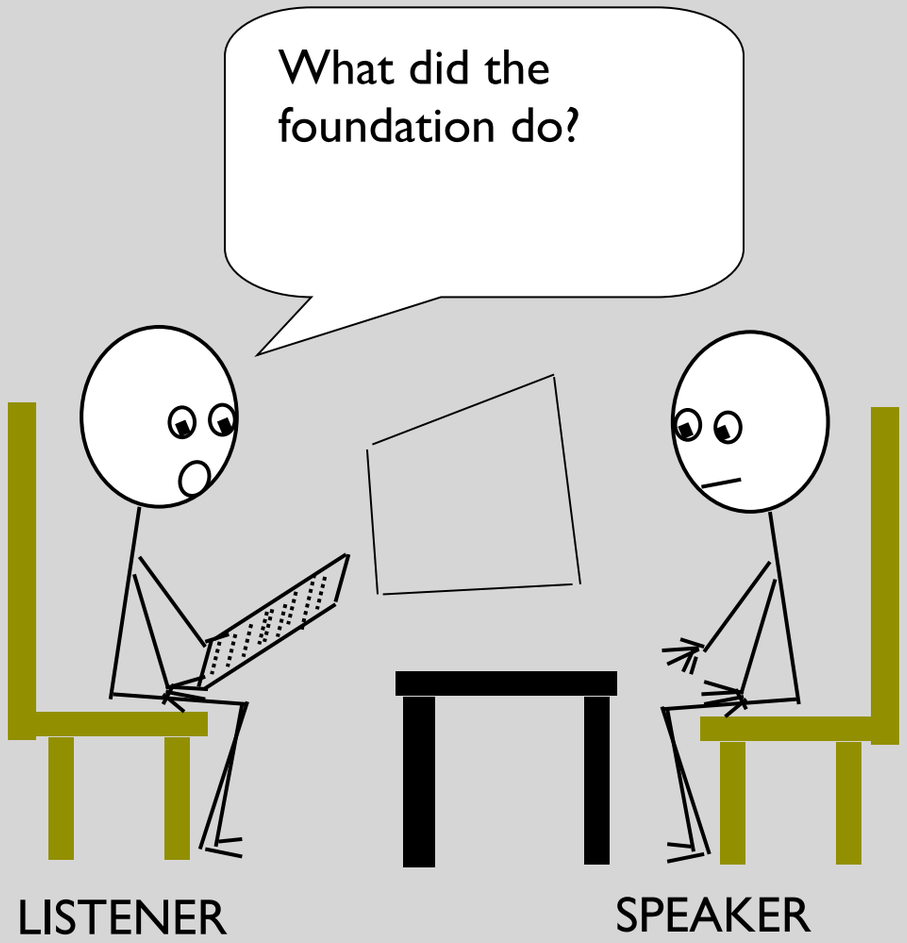
Experimental Method



1. Speaker silently reads a sentence:

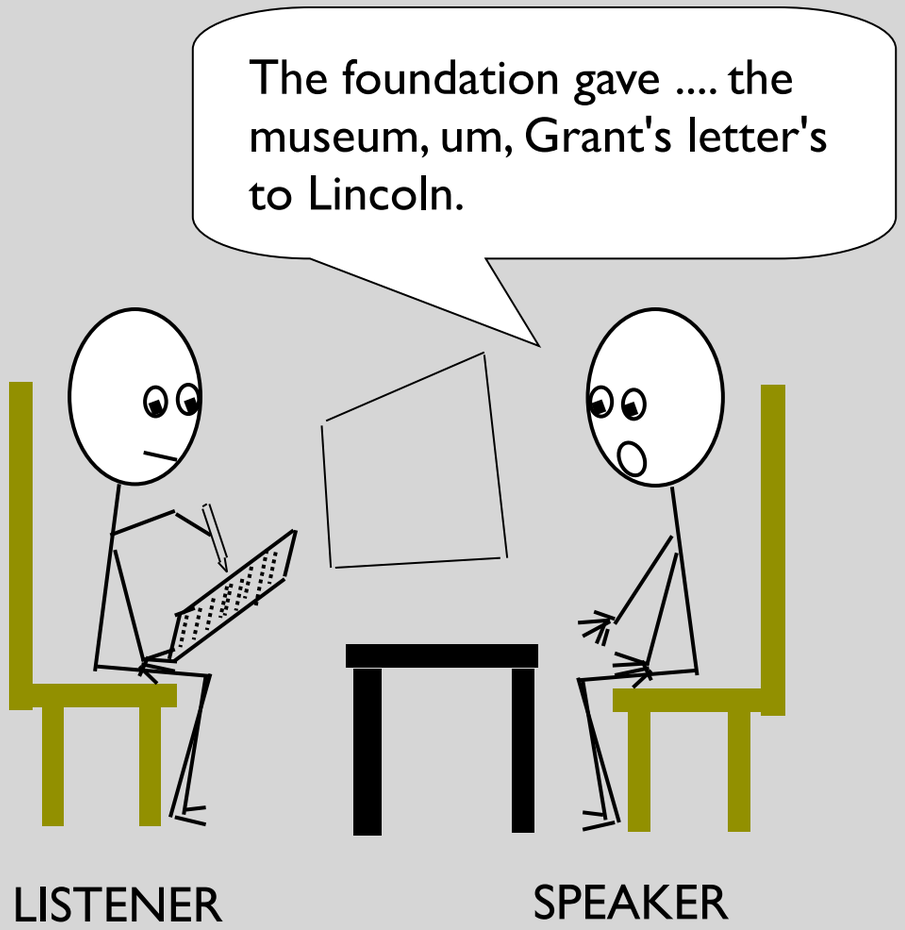
A museum in Philadelphia received Grant's letters to Lincoln from the foundation.

Experimental Method, continued



2. The sentence disappears from the screen.
   The listener reads the next question from a list.

Experimental Method, continued

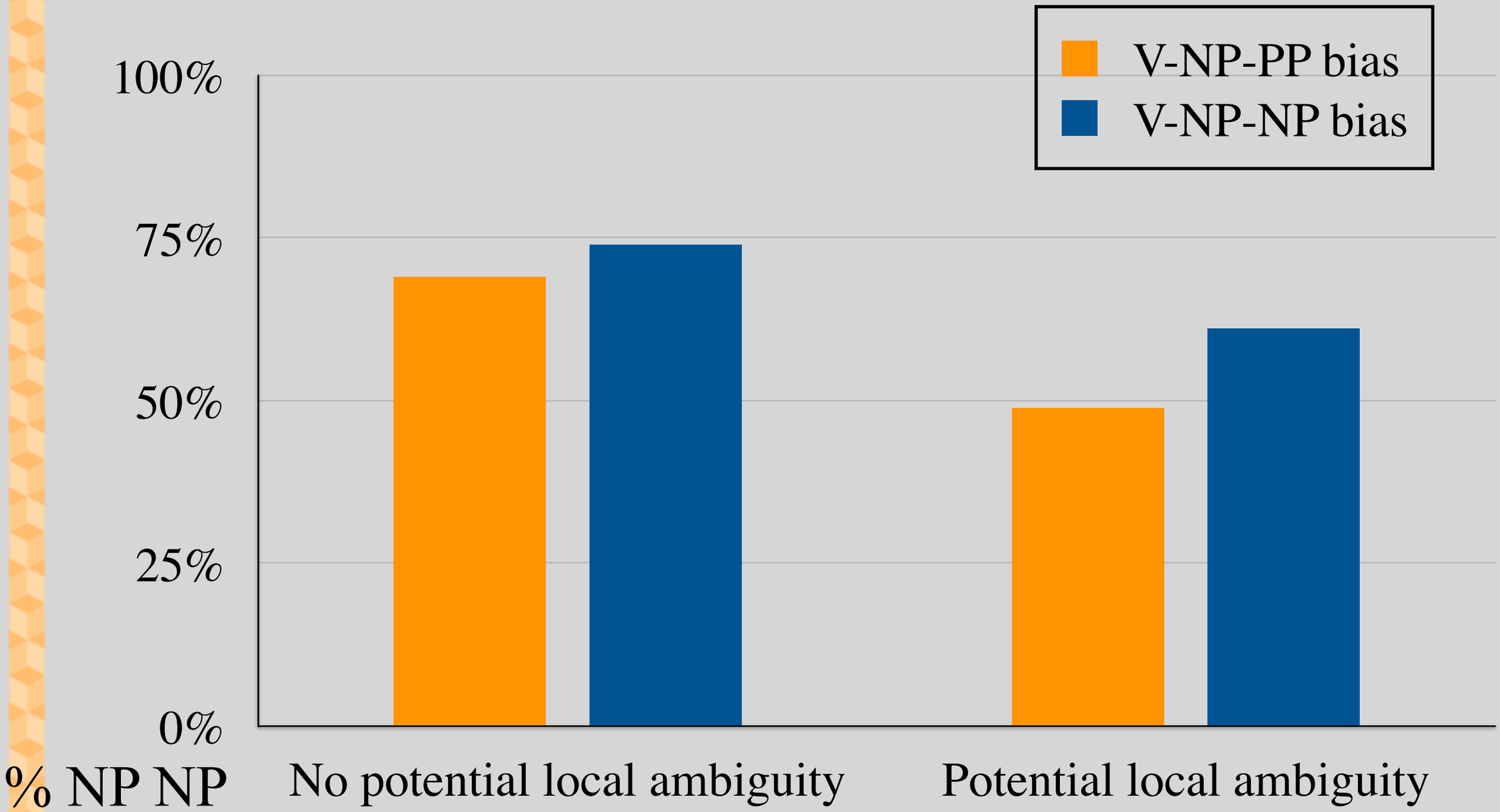


Poll!

3. The speaker answers the listener's question.

   The listener chooses the correct response on a list (from two choices).

Experimental Results on Local Ambiguity



Reverse ambiguity effect

- Arnold, Wasow, Asudeh & Alrenga 2004 *Journal of Memory & Language*
- Re-ran the experiment with slightly better methodology and found a *stronger* reverse ambiguity effect.

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.
- This has happened less than it should have.
- Existing psycholinguistic evidence favors a constraint-based, lexicalist approach (like ours).

Universals?

- P&P (top-down): attempts to relate multiple typological properties to single parameters.
- Grammar Matrix (bottom-up(-ish)): attempts to describe many languages in a consistent framework and then takes stock of common constraints.

What aspects of our grammar fragment so far seem English-specific?

Total Results: 0

What aspects of our grammar fragment so far seem like they might be crosslinguistically useful or even universal?

Total Results: 0

Universals?

- Case constraint
- SHAC
- Binding theory
- Head-complement/-specifier/-modifier
- Head Feature Principle
- Valence Principle
- Semantic Compositionality Principle
- ...

Overview

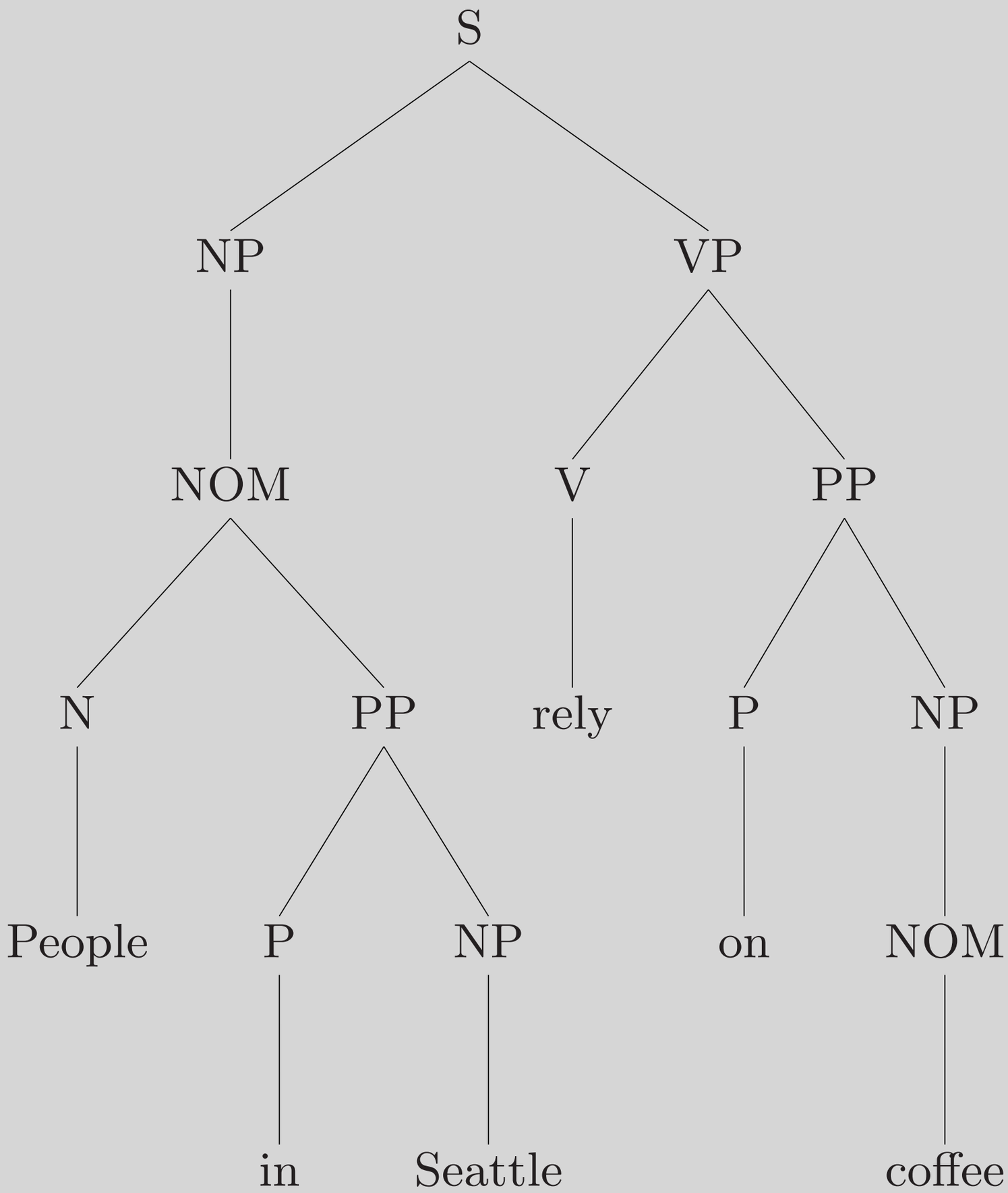
- Psycholinguistics and grammar design
  - What grammar has to say
  - What psychological evidence has to say
    - Acquisition
    - Production
    - Comprehension
- Universals

Midterm

- Posted Monday (11/6)
- Due next Monday (11/13)
- No collaboration
- Send questions to me by email
- Check Canvas read-only midterm Q&A discussion for Q&A :)

Practice sentence

- People in Seattle rely on coffee



RQs: Squiggles

- What are atom, index, and list(tau) in the Type Hierarchy? (I see that they are discussed in 9.7.2 but the squiggly bits are a bit hard to understand)

GENERAL TYPES		
TYPE	FEATURES/CONSTRAINTS	IST
feat-struc		
atom		*feat-struc*
index		*feat-struc*
list		*feat-struc*
$list(\tau)$	$\begin{bmatrix} \text{FIRST} & \tau \\ \text{REST} & list(\tau) \end{bmatrix}$	*list*
l-sequence	$\begin{bmatrix} \text{FIRST} & atom \\ \text{REST} & \langle word \rangle \mid \langle lexeme \rangle \end{bmatrix}$	*list*

Preliminaries

According to our approach, a grammar G is defined by the following components:

- a finite set of features: $\mathcal{F} = \{\text{SYN, SEM, HEAD, AGR, } \ldots\}$,
- a finite set of primitive items:
  $\mathcal{A}_{atom} = \mathcal{A}_{pol} \cup \mathcal{A}_{gr.atom} \cup \mathcal{A}_{mode} \cup \mathcal{A}_{reln}$, where:
  1. $\mathcal{A}_{pol} = \{+, -\}$,
  2. (a set of ground atoms) $\mathcal{A}_{gr.atom} = \{1\text{st}, 2\text{nd}, 3\text{rd}, \text{sg}, \text{pl}, \ldots, \text{run}, \text{dog}, \ldots\}$,
  3. $\mathcal{A}_{mode} = \{\text{prop}, \text{ques}, \text{dir}, \text{ref}, \text{none}\}$, and
  4. $\mathcal{A}_{reln} = \{\textbf{walk}, \textbf{love}, \textbf{person}, \ldots\}$,
- a denumerably infinite set of primitive items: $\mathcal{A}_{index} = \mathcal{A}_{ind} \cup \mathcal{A}_{sit}$, where:
  1. $\mathcal{A}_{ind} = \{i, j, \ldots\}$ and
  2. $\mathcal{A}_{sit} = \{s_1, s_2, \ldots\}$,
- the distinguished element *elist* (*empty-list*), discussed below,
- a finite set of types: $\mathcal{T} = \{\textit{noun}, \textit{agr-pos}, \textit{plural}, \textit{expression}, \ldots\}$,

1. $\mathcal{A}_{ind} = \{i, j, \ldots\}$ and
2. $\mathcal{A}_{sit} = \{s_1, s_2, \ldots\}$,

- the distinguished element *elist* (*empty-list*), discussed below,
- a finite set of types: $\mathcal{T} = \{noun, agr\text{-}pos, plural, expression, \ldots\}$,
- a type hierarchy with a tree structure associated with constraint inheritance (for instance, the type hierarchy represented by the tree and table in Section 9.2.1 and 9.2.2),
- a set $\mathcal{LT} \subset \mathcal{T}$ called the *leaf type* (a type τ is a *leaf type* if it is associated with a leaf in the type hierarchy tree, i.e. if τ is one of the most specific types),
- a set of list types (if τ is a type, then $list(\tau)$ is a list type),
- a set of grammar rules (see Section 9.2.4),
- a set of principles,
- a lexicon (which is a finite set of lexical entries like those in Section 9.2.6), and
- a set of lexical rules (like those in Section 9.2.5).

RQs: FIRST/REST

- Where in the book is FIRST/REST mentioned prior to this chapter? I noticed it in the type hierarchy and looked through previous chapters and did not see it. What is the purpose of FIRST and REST?
- The Constant Lexeme Lexical Rule on page 259 does not mention FIRST, but in the chapter 9 grammar summary on page 282, the output of the Constant Lexeme Lexical Rule changes the [FIRST tag1]. Why is this?

RQs: Lexical thingies

- In what situation do we have to use "lexical entry" rather than "lexical sequence"?

RQs: Other theories

- What is Transformational Grammar?
- What makes a generative grammar generative?

RQs: Other theories

- I haven't really studied a lot of theories of syntax, so I was wondering, what are some theories that aren't surface oriented? How are they different on a surface level (ha) from what we do in HPSG?
- Can you use an example to elaborate on why transformational grammars are not process-neutral?

RQs: Other theories

- I wonder if we can draw some parallel between HPSG and Optimality Theory, as the way HPSG differs from transformational grammar reminded me of how OT differs from rule-based phonology. It appears that both HPSG and OT incorporate the idea of "filtering out" bad outputs from an abstract pool of infinitely many possible forms, with each possibility being somewhat given equal consideration.

- Strong Lexicalism: Are there analyses of any phenomena that can falsify this hypothesis?

RQs: Other theories

- Different theories of syntax seem to raise different kinds of questions, and have different emphases that are dependent on their design features. For instance, transformational grammars like MP/GB love to argue for configurational analyses because this is the main thing that their theory provides. The lexicalist side might do the opposite (?). Is it 'correct' to allow a particular framework influence the way a linguist reason about linguistic patterns?

RQs: Language and cognition

- Is it important that a theory of syntax reflect human language processing?
- On page 308: "Further, our grammatical theory suggests a number of parallels between the kinds of information structures needed to account for linguistic competence and those employed in other cognitive domains." Some similar concepts to "Other cognitive domains" have been mentioned several times above, and I'd like to know what exactly they are.
- In this chapter, knowledge such as that needed to disambiguate word sense (e.g., 'pen' as a fenced enclosure vs. a writing instrument) is referred to as non-linguistic knowledge. I am wondering what other knowledge this term encompasses and if there are applications that leverage this type of knowledge to address ambiguity?

RQs: Garden path sentences

- Why are they called Garden Path?
- Are there tools for parsing/recognizing garden-path sentences? Does HPSG play any role in building tools like such?

RQs: Incremental processing

- Also for partial completions of sentences, like if someone starts asking or saying something and gets interrupted and answered, if you feed that data into this grammar would it still draw trees for the interrupted phrase and consider it well-formed because it's incremental? Since technically they were starting to say a fully fleshed out well-formed sentence of English.

RQs: Incremental processing

- When talking about echo questions, the consideration of grammaticality came to mind. Say in a situation where speaker B is echoing a question back to speaker A like on page 301, and speaker A answers speaker B before speaker B finished their echo question (ex: Speaker B: "Who did you say is --" Speaker B: "Señora Maria Consuelo Bustamante y Bacigalupo"), is Speaker B's sentence considered ill formed because it was interrupted and therefore unfinished? or do we assume the rest of the sentence would round it out grammatically?

RQs: Incremental processing

- On page 301, the book talks about "echo questions" and how realistic grammars need to allow for the "efficient incremental computation of partial analyses". In (53), the book mentions that the listener can likely extrapolate the remainder of the question around the asterisks. Is the goal for our grammar then to, also around the asterisks, also be able to interpret what the remainder of the question is, if its meant to incrementally progress words/sounds as it receives them?

RQs: Incremental processing

- Related to this incremental interpretation, say we had an interaction as follows: "Can you take out the trash? / I can't." - In this example, would one goal of our grammar to be able to contextually tell that "can't" is related to the action of taking out the trash (perhaps storing that information in the SEM RELN list)?

RQs: Frequentistic knowledge

- How is lexical probability captured in hpsg? From example (55): "The sheep in the pen..." How can our grammar select the 'fenced enclosure' sense of pen rather than the 'writing implement'? Both are grammatical, and a listener could do it by nonlinguistic knowledge. Is this resolved by items in the RESTR list, or does it have to be done by incorporating probabilities from a corpus?

RQs: Computational modeling

- As mentioned in the textbook, our grammar seems to model how humans actually speak and process language fairly well - is this still the case when we apply our grammar to build NLP systems?