Ling 566
Nov 3, 2016

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

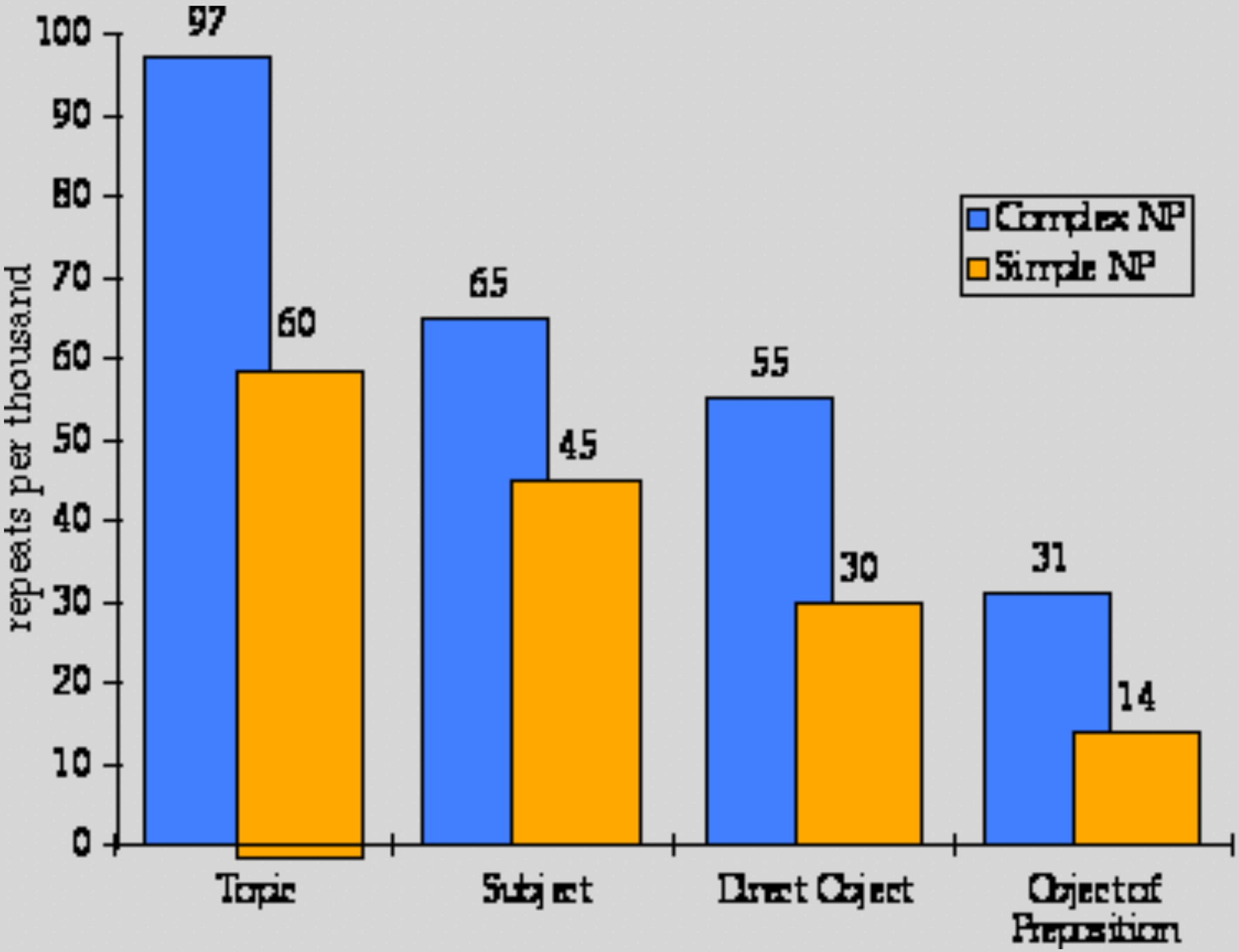
- Chomsky's nativism is very controversial
  - It is based on the "poverty of the stimulus" argument, and a model of learning as hypothesis testing.
  - The environment may be more informative than he assumes.
  - There may be more powerful learning methods than he assumes.
- There has not been much 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.

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 over the past 20 years has been 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.

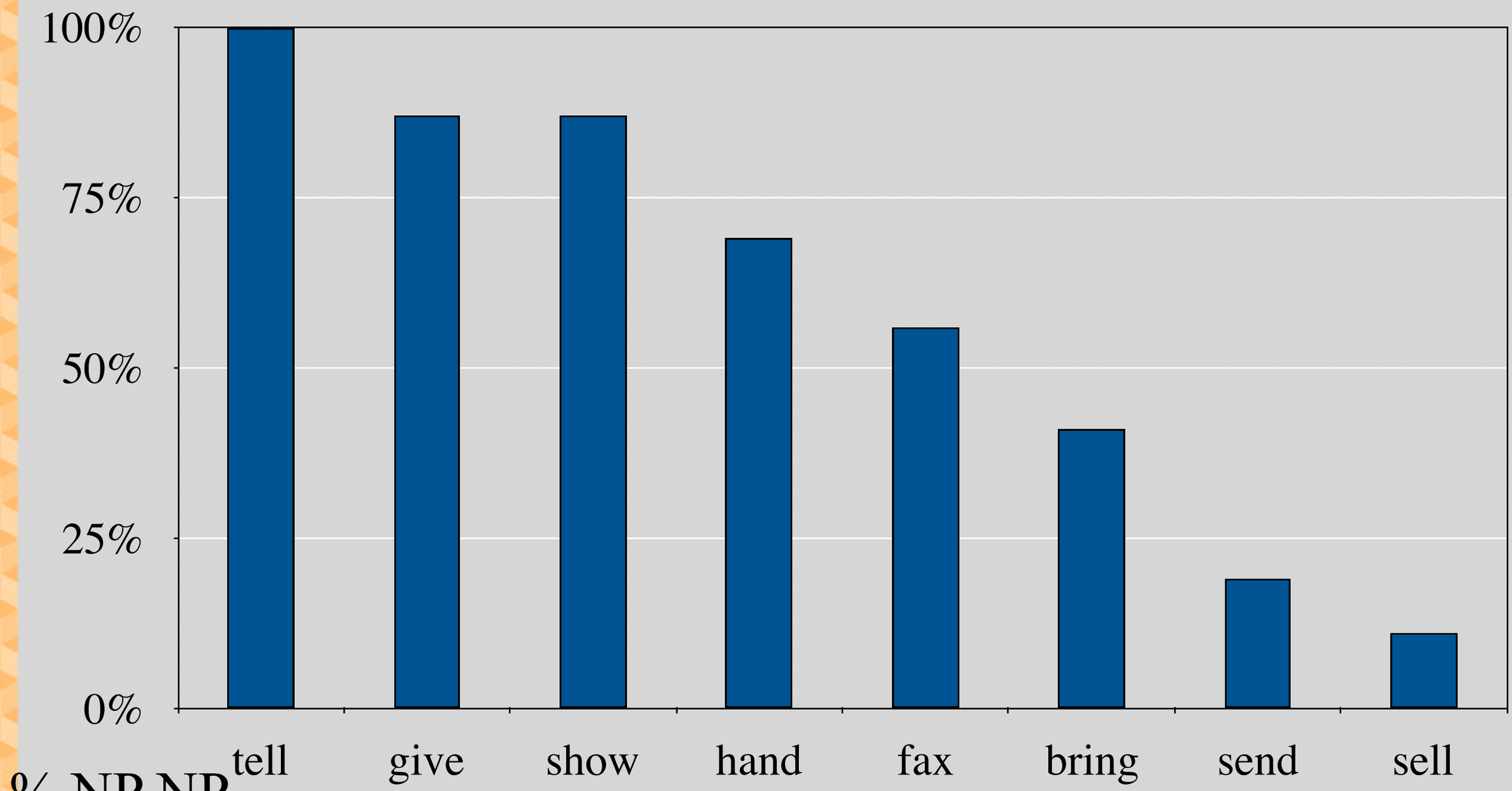
Speakers know a great deal about individual words

- 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

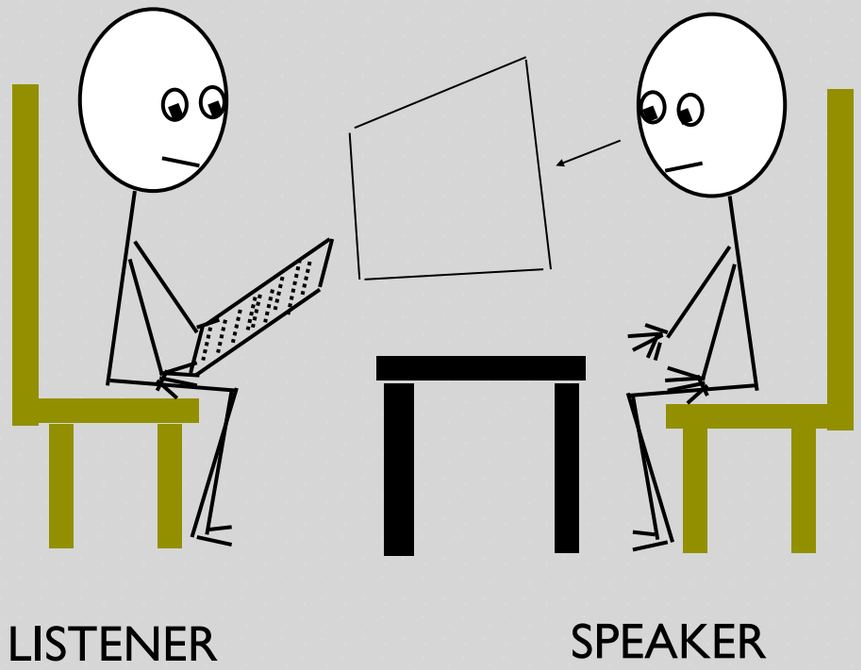
- Wasow et al ran a production experiment to test whether ambiguity avoidance would influence speakers' choice between (1) and (2):

  (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).

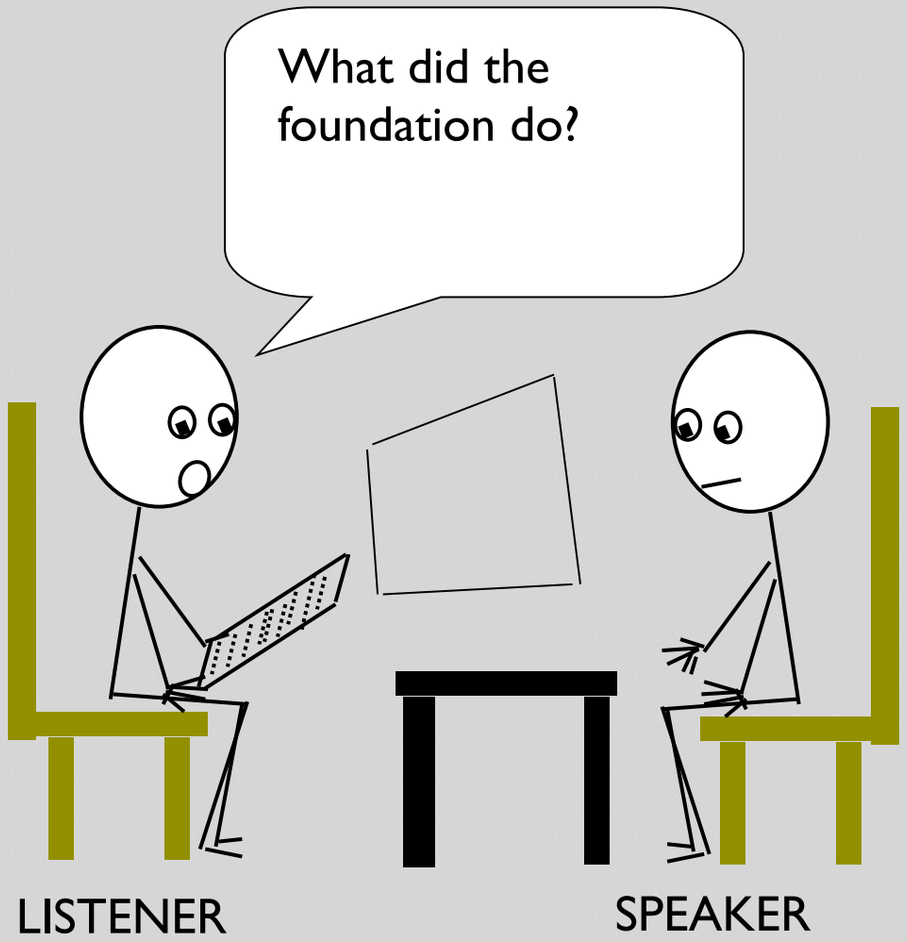
Experimental Method



1. Speaker silently reads a sentence:

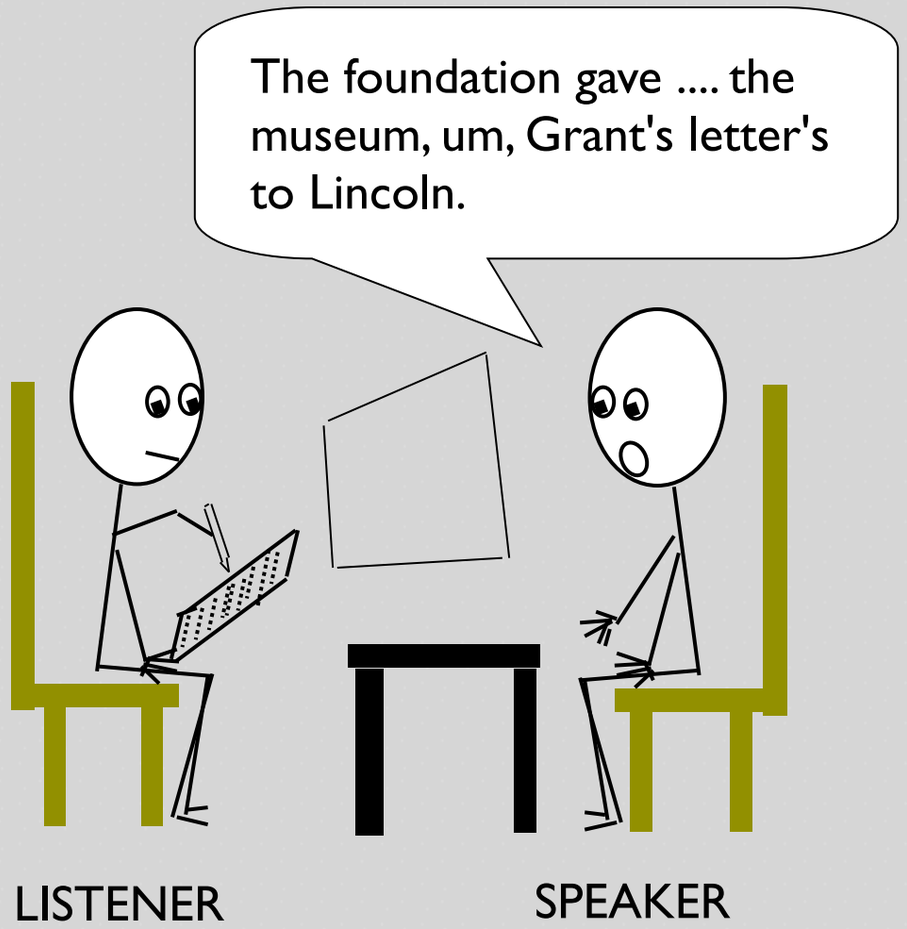
A museum in Philadelphia received Grant's letters to Lincoln from the foundation.

Experimental Method, continued



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

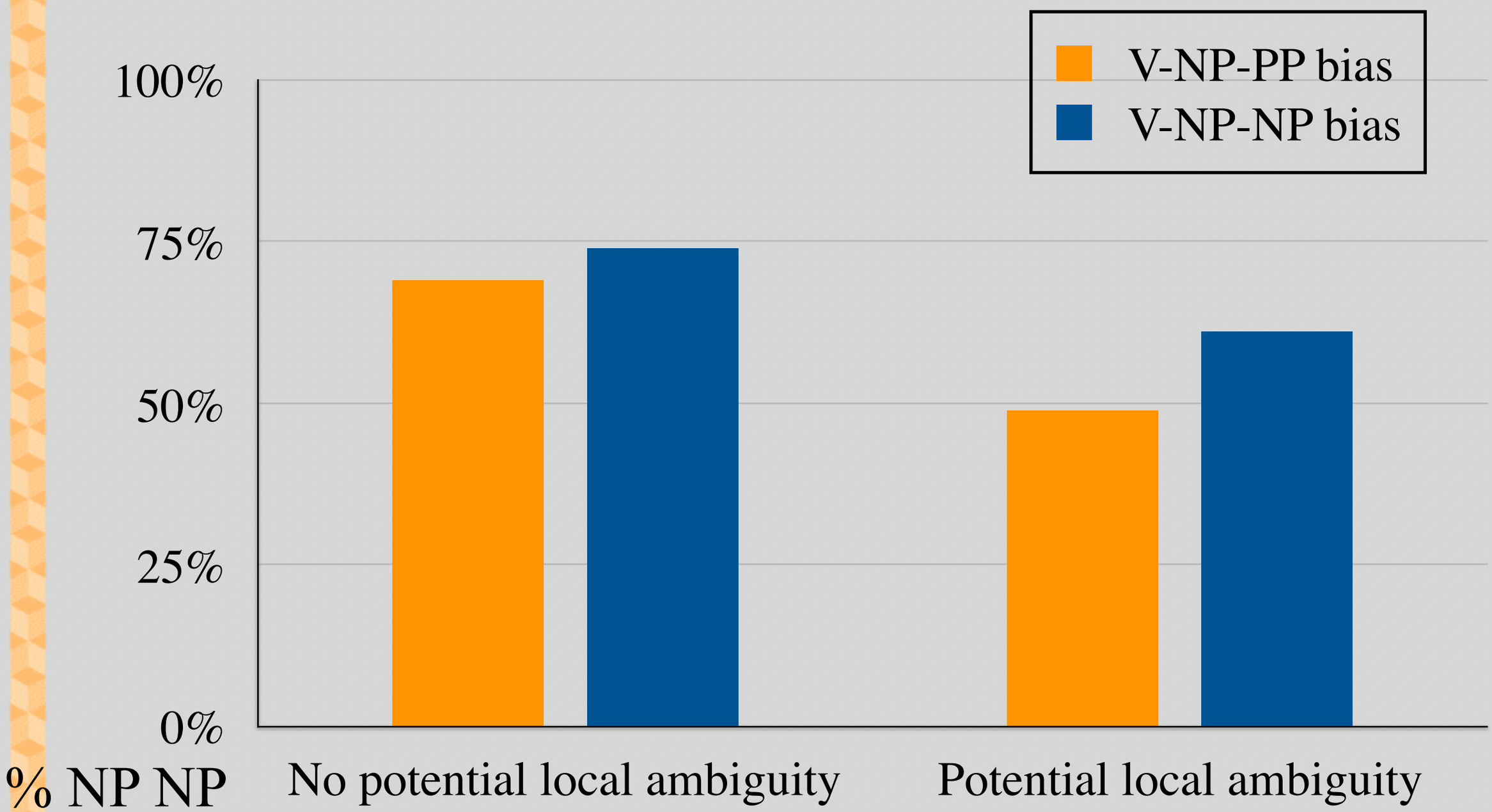
Experimental Method, continued



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.

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

Reading Questions

- On page 274, under GENERAL TYPES, there are types *atom*, *list*, *l-sequence* that I don't quite understand. Also, the whole FIRST/REST thing is a bit unclear to me. Could you explain these in a bit more detail?

- What's the relationship between lexemes, phonological forms, and atoms? Are atoms subsets of phonological forms or are they one and the same thing?

- If a list happens to be empty, what is the FIRST of that *list*?

TYPE	FEATURES/CONSTRAINTS	IST
feat-struc		
atom		*feat-struc*
index		*feat-struc*
l-rule	$\left[\begin{array}{ll}\text{INPUT} & \textit{l-sequence}\left\langle \text{X} , \left[\text{SEM} \; / \; \boxed{2}\right]\right\rangle \\ \text{OUTPUT} & \textit{l-sequence}\left\langle \text{Y} , \left[\text{SEM} \; / \; \boxed{2}\right]\right\rangle \end{array}\right]$	*feat-struc*
list		*feat-struc*
list(τ)	$\left[\begin{array}{ll}\text{FIRST} & \tau \\ \text{REST} & \textit{list}(\tau) \end{array}\right]$	*list*
l-sequence	$\left[\begin{array}{ll}\text{FIRST} & \textit{atom} \\ \text{REST} & \langle \textit{word} \rangle \mid \langle \textit{lexeme} \rangle \end{array}\right]$	*list*

Reading Questions

- Is pyscholinguistics really in the remit of explanatory power of our theory of grammar here? Isn't it better to keep our grammar constrained to syntactic descriptions of language?
- Is biological plausibility the main reason we care about order independence in the grammar? Are there computational advantages, too?

Reading Questions

- If we agree that HPSG supports incremental processing, what would that processing look like? Is there evidence to suggest that the processing starts at a few leaves, constructs a node, intakes another leaf and adds that leaf to the previous node? Do we process in a bottom-up kind of way?

- Is there evidence that this implies HPSG is a "better" way of processing language than other methods which are less similar to the way humans naturally process it?

Reading Questions

- Language processing is incremental and language users start processing/predicting what is been said even before the utterance is completed. In your current experience, have language processing technologies come remotely close that? Or are we not there yet technologically speaking?

The sheep in the pen had been sleeping and were about to wake up.

- As humans, we can read this sentence and rapidly understand *pen* to mean an enclosure rather than a writing implement. Is there anything in HPSG that makes that kind of inference possible? Are there separate lexical entries for items (like *pen* and *pen*) that don't differ at all grammatically but are different in their semantics?

Reading Questions

- We've been really careful to distinguish the theory from the formalism all along, but I'm not sure which one we're arguing from in chapter 9. It seems like the sentence processing discussion is citing facts about trees and lexical rules that belong to the formalism. Of course, the theory and the formalism have to be related in some way, but it's not totally clear to me how to go from facts about the formalism to facts about the theory. Are we, in fact, arguing from the formalism here, and if so, how are we claiming the formalism is related to the theory?

Reading Questions

- It's emphasized that the system we are developing doesn't make use of "destructive operations." What might be an example of a destructive operation, and why is this important?
- What does surface oriented mean? What would a non-surface oriented theory look like?

Reading Questions

- I don't understand how the incremental processing of information supports constraint-based models of language. I understand how it's evidence against transformational models, but not evidence for constraint-based.

- Additionally I didn't quite understand what it means for a grammar to be process-neutral, and how our grammar is and transformational grammar is not.

Reading Questions

- Chapter 9 says that HPSG is more similar to the way humans process language than, say, transformational grammars. In what ways is HPSG still different from human language processing?
- We often focus more on the competence side of language than on the performance part. However, I assume that implemented grammars often have to deal with disfluencies. What are some of the ways that they try to make sense of disfluencies?