Ling 566
Oct 31, 2019

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:
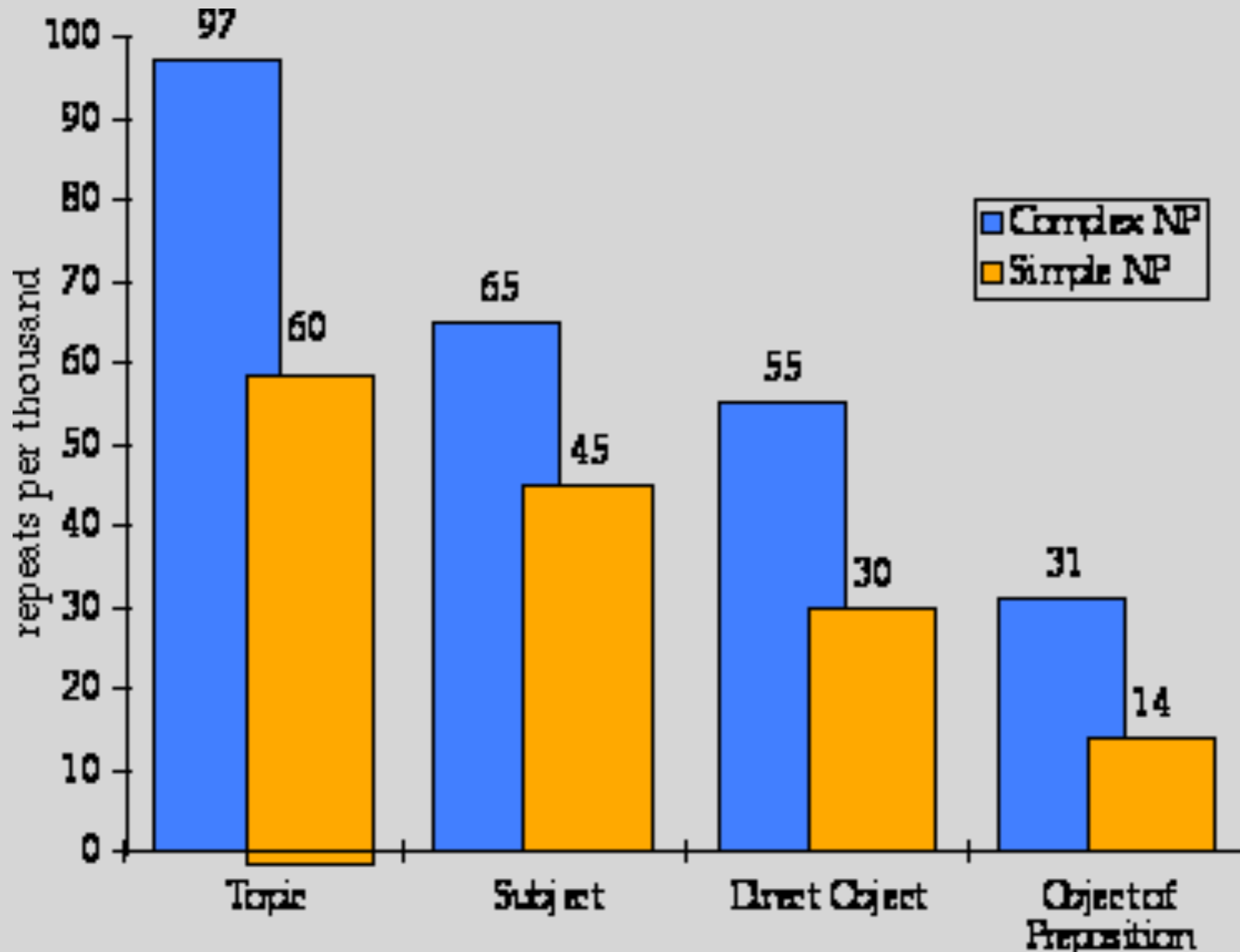
Production errors are sensitive to syntactic structure

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*

% NP NP order
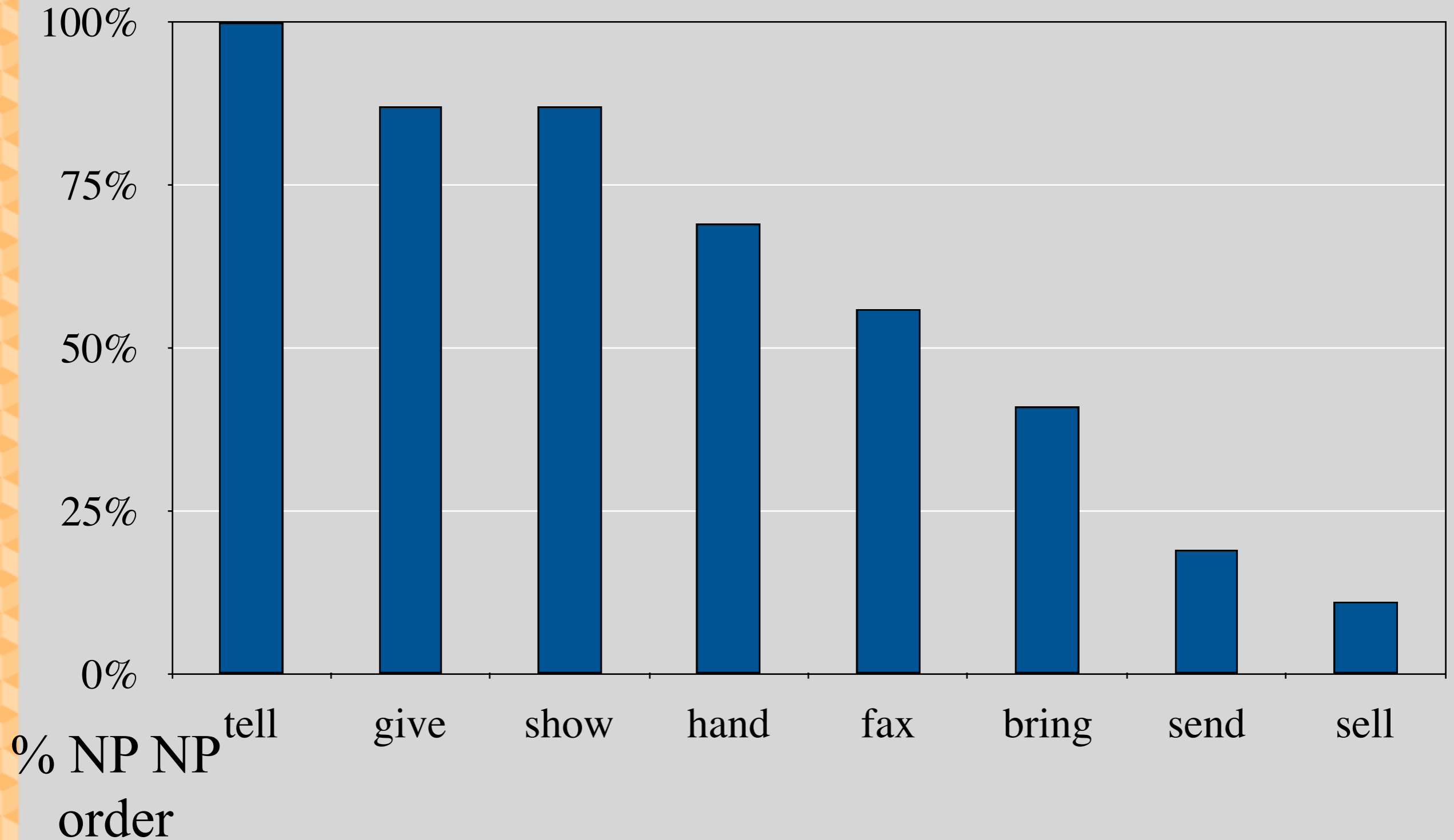
Lexical biases influence processing

- 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).
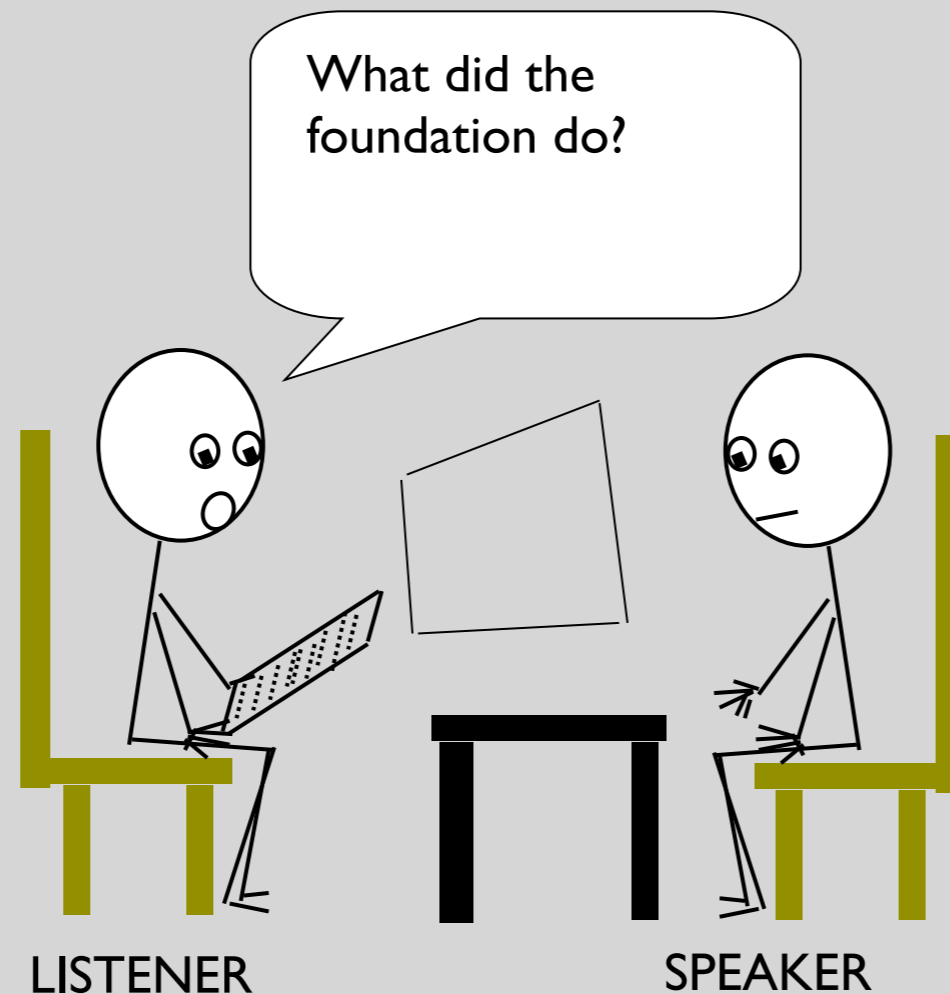
Experimental Method
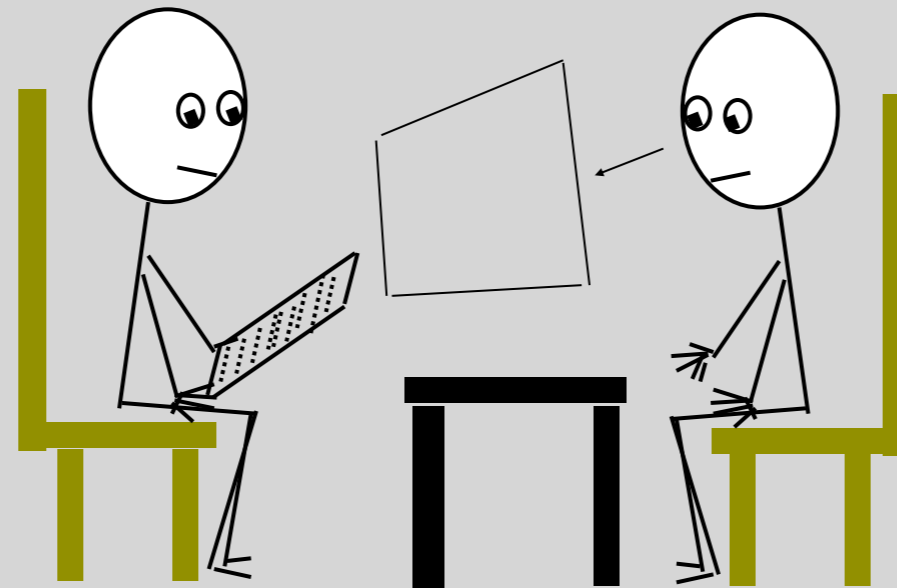
LISTENER   SPEAKER
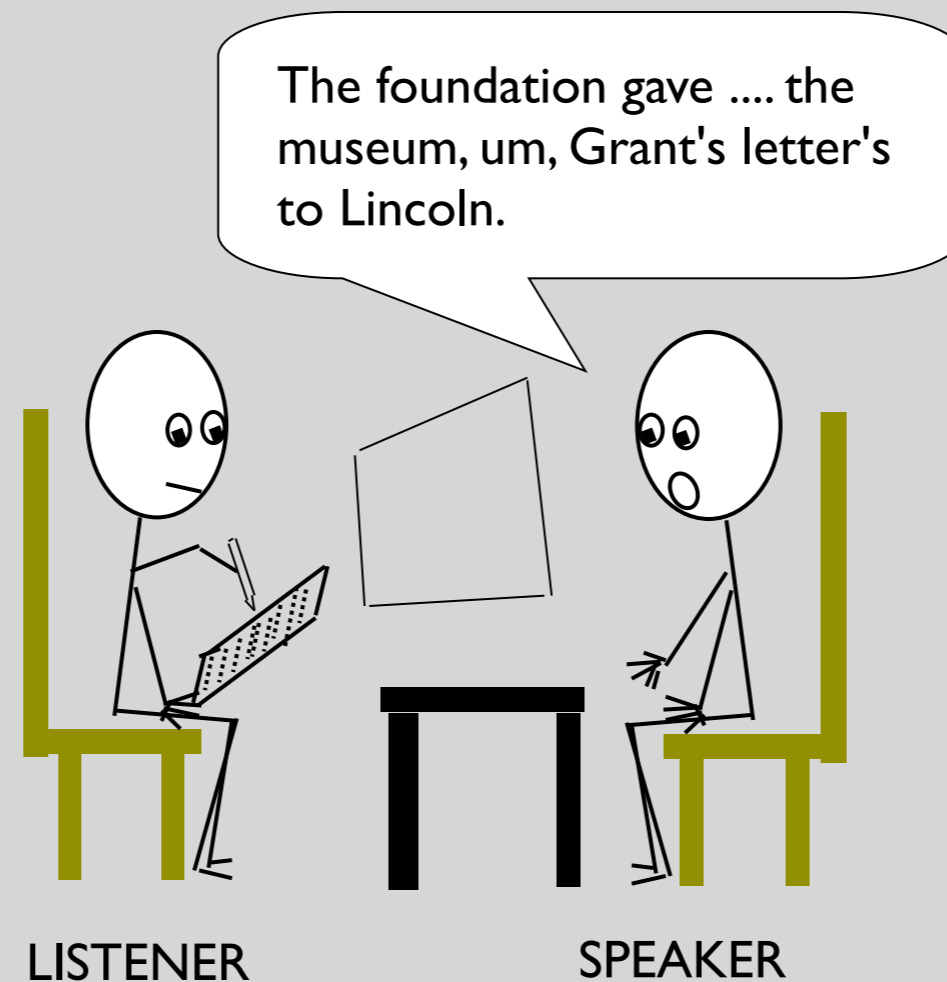
1. Speaker silently reads a sentence:

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

The foundation gave .... the museum, um, Grant's letter's to Lincoln.

LISTENER   SPEAKER

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
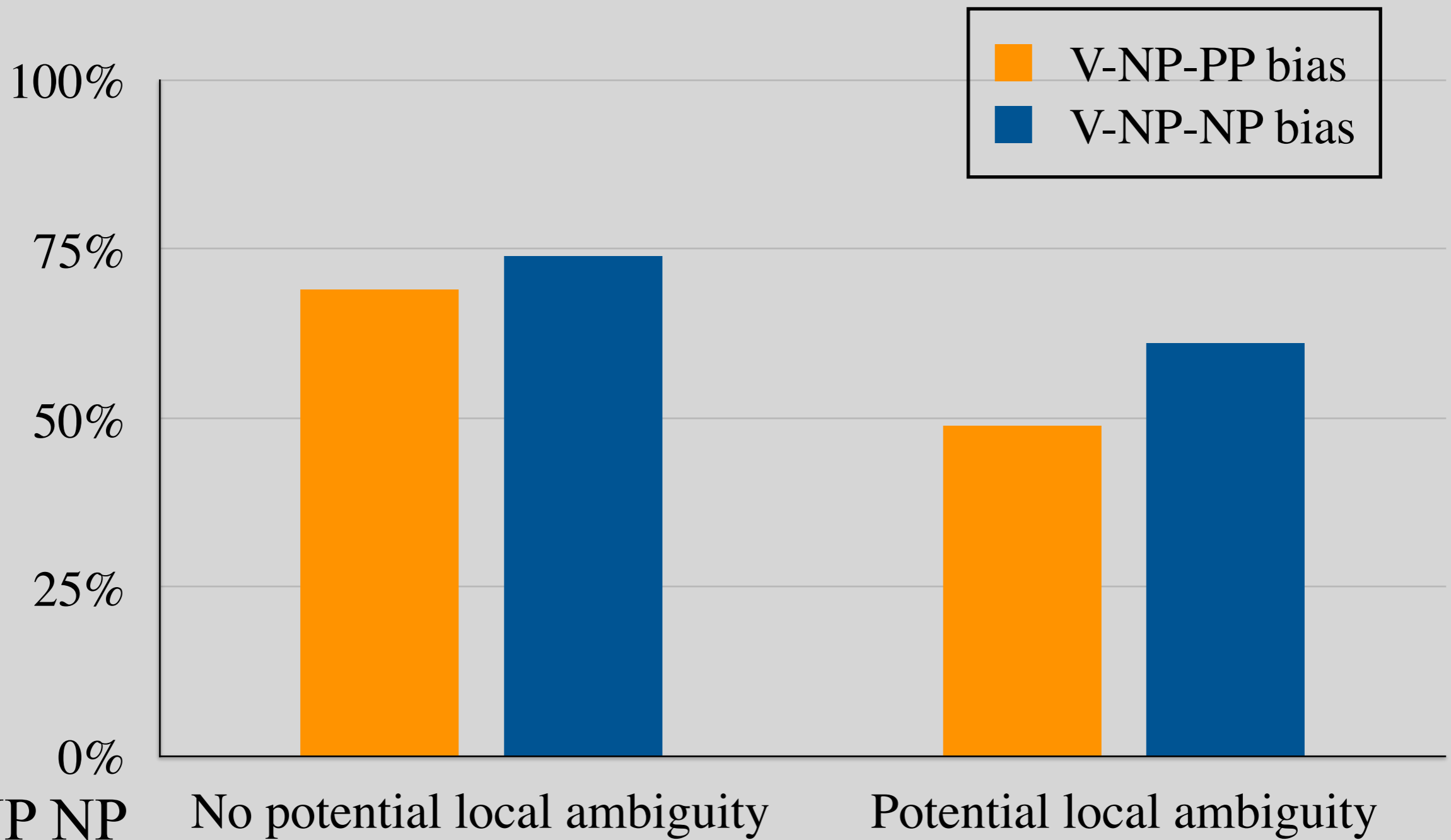
Reverse ambiguity effect

- 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

- How can a process-neutral grammar more accurately model language production? What exactly does it mean to be a process-neutral grammar, and what does this look like in practice?

- I can see how this is true with the grammar in this book with partial information about surrounding structure and semantic information on each part of the sentence. However, if when we are talking, we process sequentially from begging to end, wouldn't our grammar fall short? I feel like everything starts with the head verb of the sentence in our grammar. It needs to be there to have a grammatical sentence, and it's what selects for an NP specifier.

Reading Questions

- We tend to be focusing on written/theoretically "perfect" structures when building our grammar and discussing the potential universality of our grammar. From this I'm wondering how grammars of actual performed speech might have any differences. A grammar of performances would probably be pretty different than a "competent" grammar, I wonder if speech disfluencies and other performance-specific features of language would help or hinder a grammar that is attempting to touch on universality. I'm also curious at a high level of what differences a performance-grammar would have from the one we're building.

Reading Questions

- Surface-oriented grammars basically ignore any innate knowledge of the world(such as fences being related to bars) in order to have a deterministic view of the world right? If so isn't a lot of the information we encode into our lexical entries a simplification of some of this world view? Aka lexical entry for loves has a lover and a loved one. Is that not external information?

- Separately the textbook is saying that our grammar can work within the human(with external knowledge) world because of it declarative and does not have any destructive properties. Which part of our tree are we positing would be used by an entity? The whole tree? The root node?

Reading Questions

- What would a non-lexicalist grammar look like? Would a non-lexicalist grammar be even remotely efficient?

- There was a small note about frequency of certain words affecting how we process and interpret garden path sentences (e.g. in the horse raced past the barn fell, raced is more frequently a finite verb form instead of a passive participle). I'm curious to know--is there a version of all these lexical entries having an entry for frequency? I know this would depend on what corpus you're taking frequency (counts/total) over, but is this a thing--like probabilistic HPSG?

Reading Questions

- I can't help but wonder what lexical rules would look like for a universal grammar. What would one need to begin forming those beyond a catalog of all possible human phonetics?

- The idea of phrase structure rules as 'natural candidates for universality' (Pg. 307) is attractive. I wonder about how much the general constraint they enforce (order) falls apart in highly agglutinative languages/languages without strong word order. Would this concern be eschewed by considering morphosyntax to be formed by structure rules?

Reading Questions

- On pg. 307 the book states: "Some features (e.g. PER, GEND, COUNT) clearly reflect properties of the world or of human thought, whereas others (e.g. CASE, FORM) seem specifically linguistic)." However, to me the ability to map features like GEND and COUNT to real world properties seems language specific. I think the section was saying that our ability to map some features onto real world concepts gives some merit to the idea of the universality of grammar, but because even the "real world" features aren't universally represented I am confused about this suggestion.

- Why do we want universal grammar anyway?

Reading Questions

- One of my main questions after reading the explanations of how our grammar fits with the psycholinguistic study results, such as in the specific example about our treatment of valence, is whether designing the grammar in a way that would be compatible with such psycholinguistic evidence was a goal from the outset of the grammar development process? Or, rather, if the grammar was first mostly built without direct consideration of psycholinguistic data (perhaps there wasn't such wide data available when it was first being developed?), and the fact that ends up jiving with the psycholinguistic evidence is more taken as confirmation for the suitability of our grammar's design choices?

Reading Questions

- I can understand why a linguistic theory would want to model human language cognition as closely as possible - but is this important when we start modeling language with a computer? Should we be trying to create a model that suits how computers process language? Or do the same traits apply? I guess I'm curious how this grammar gets modified when we have to apply it to a computational task (maybe I just have to wait until 567…)

Reading Questions

- The book states "So if a grammar is to be realistic it needs to characterize linguistic knowledge in a way that allows for efficient computation of partial analyses" (Pg 304). From an implementation perspective how do we analyze the well-formedness of sentences when we have such an arrangement, since some parts of a sentence are interdependent?

Reading Questions

- If I understand correctly, the chapter suggests that people process a sentence left-to-right from the linear surface form, and build meaning gradually word by word as they read in the input stream, without constructing any potential underlying structure in their mind. If this is the case, how would people understand a head-final language sentence, where more crucial information are closer to the end of the sentence? It seemed to me that surface information is not enough in this case. You need to have a structure skeleton ready first, and wait for the head word, then you can retrieve any relevant information once the structure is filled?

Reading Questions

- So how to psycholinguists criticize HPSG?

Reading Questions

- In 9.5, the discussion of frequency, as it pertains to the aforementioned garden path sentences, raised questions about computational heuristics for resolving ambiguities. What sorts of tools are available for parsing phrases that our grammar (as it exists) can't resolve? I'm reminded of using Markov chains for the purpose of part-of-speech tagging, another area where the resolution of ambiguity is a high priority.