Ling 566
Nov 3, 2022

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.

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?

Sounds plausible, no prev UG class

Sounds plausible, studied UG before

Not sure, no prev UG class

Not sure, studied UG before

Seems unlikely, no prev UG class

Seems unlikely, studied UG before

Brief remarks on language acquisition

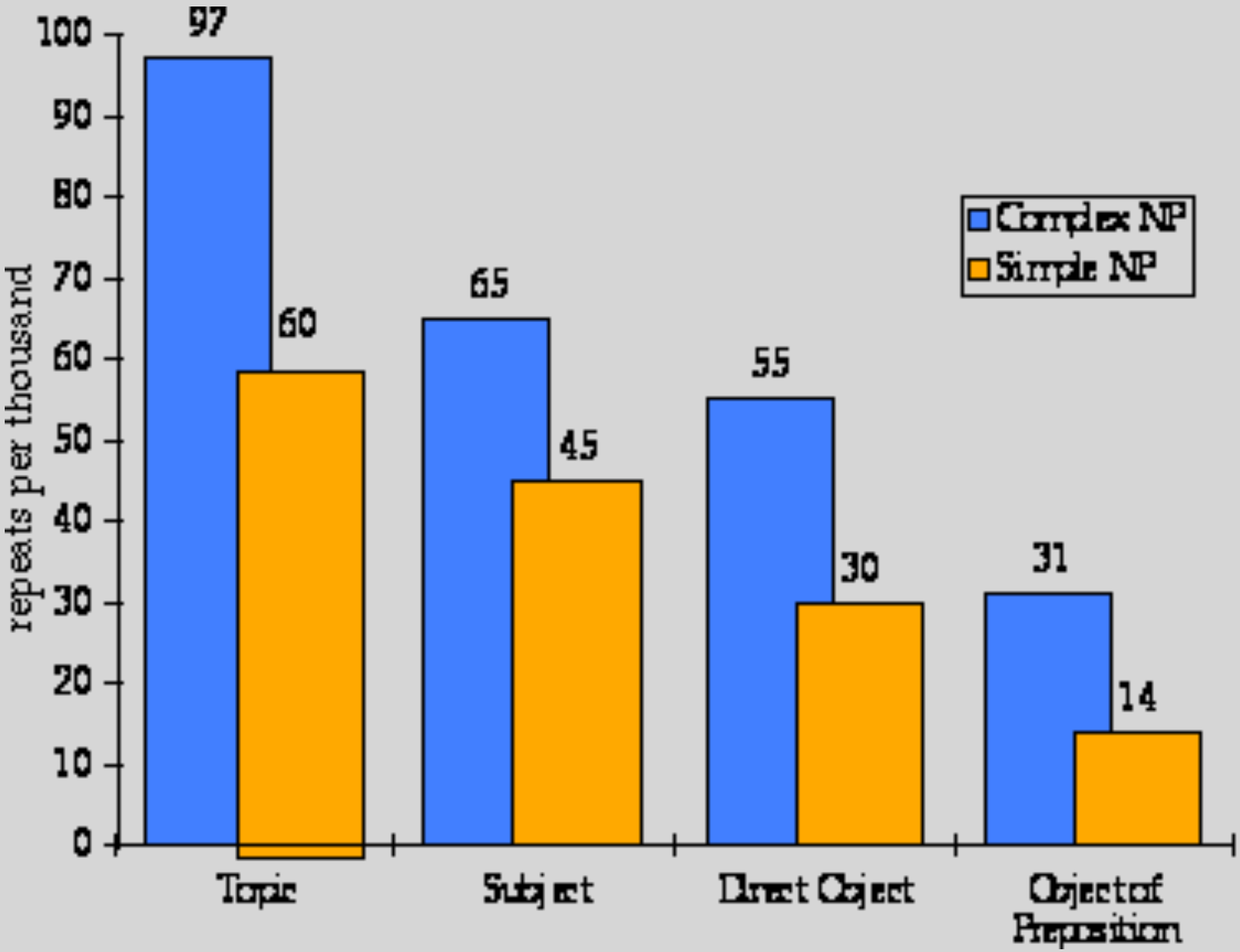
- Chomsky's nativism is very controversial
  - It is based on the "poverty of the stimulus" argument, and a model of learning as hypothesis testing.
  - The environment may be more informative than he assumes.
  - There may be more powerful learning methods than he assumes.
- There has not been much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

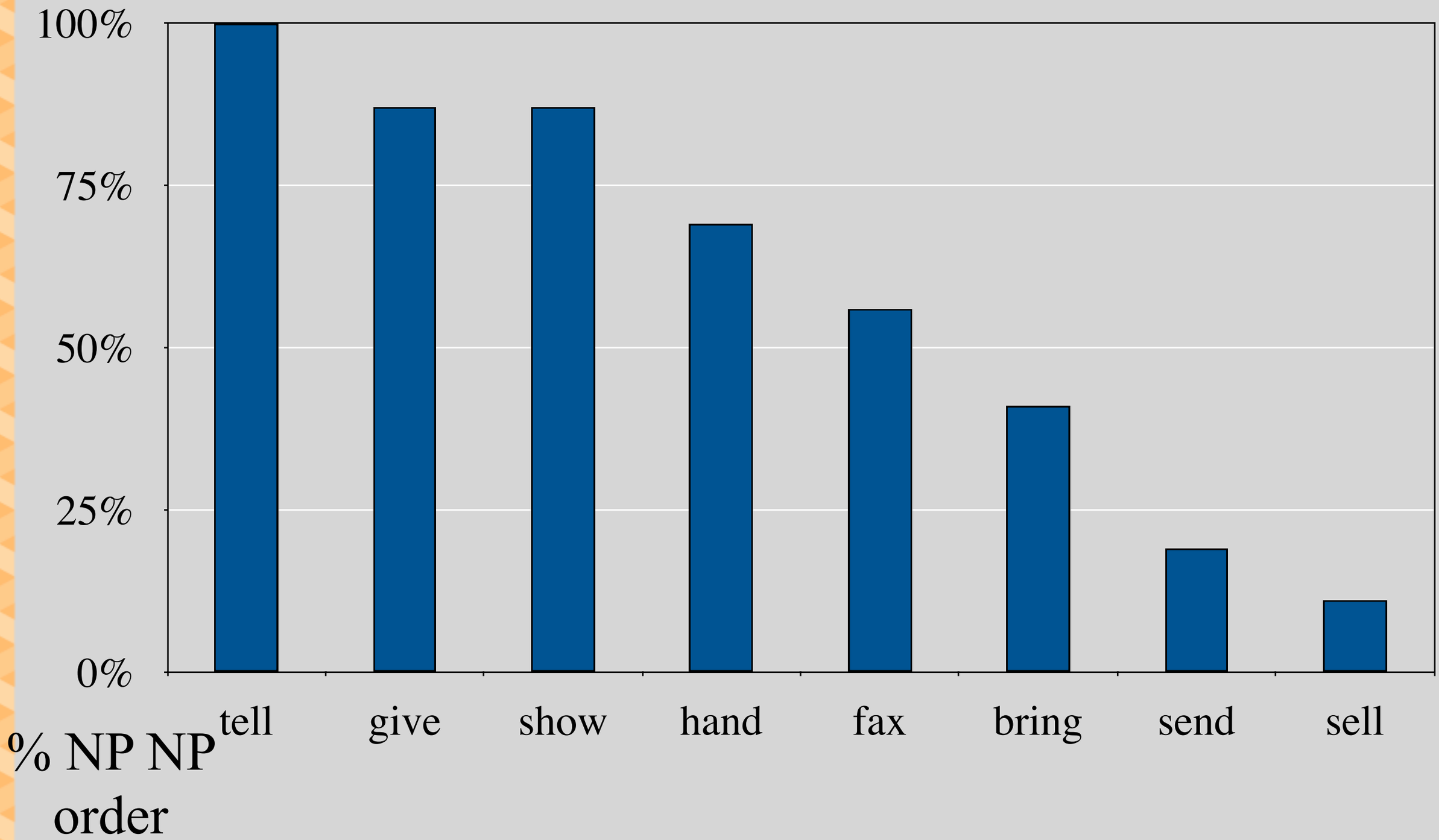
Speakers know a great deal about individual words

- 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

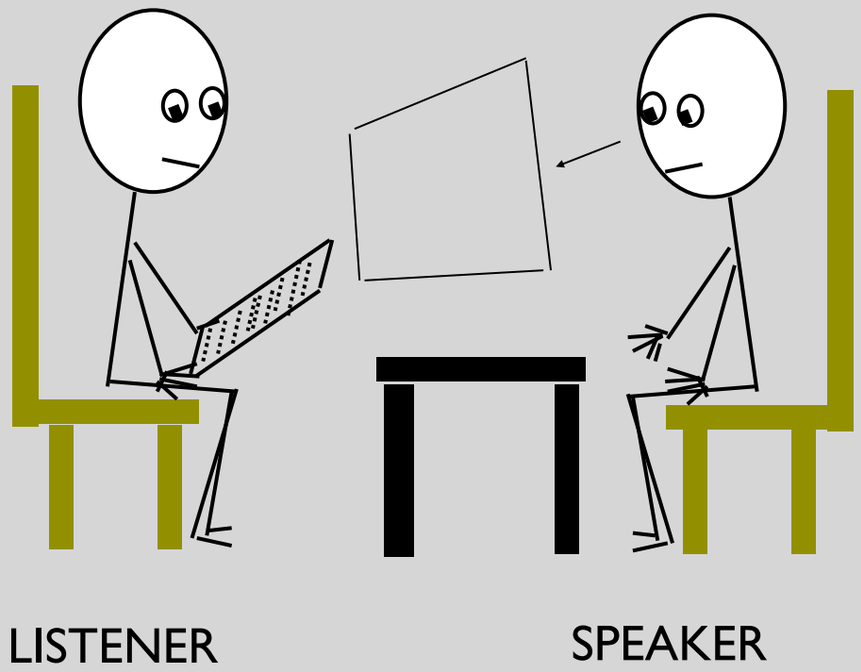
- Wasow et al ran a production experiment to test whether ambiguity avoidance would influence speakers' choice between (1) and (2):

  (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).

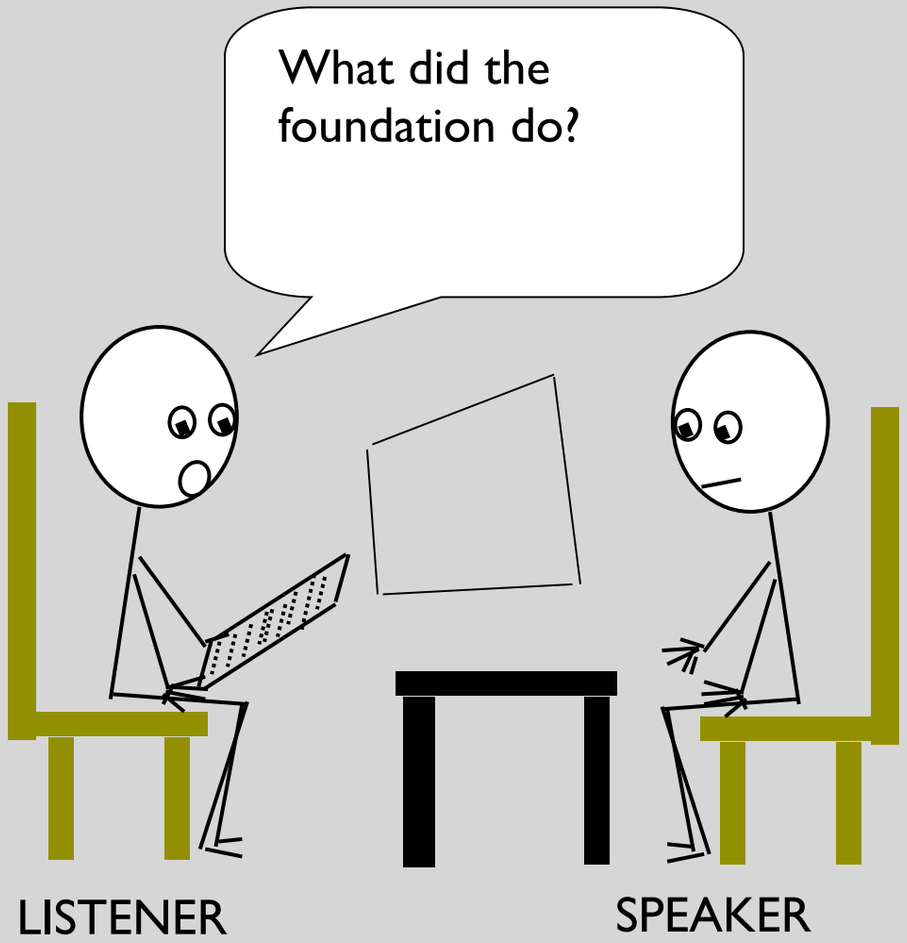
Experimental Method



1. Speaker silently reads a sentence:

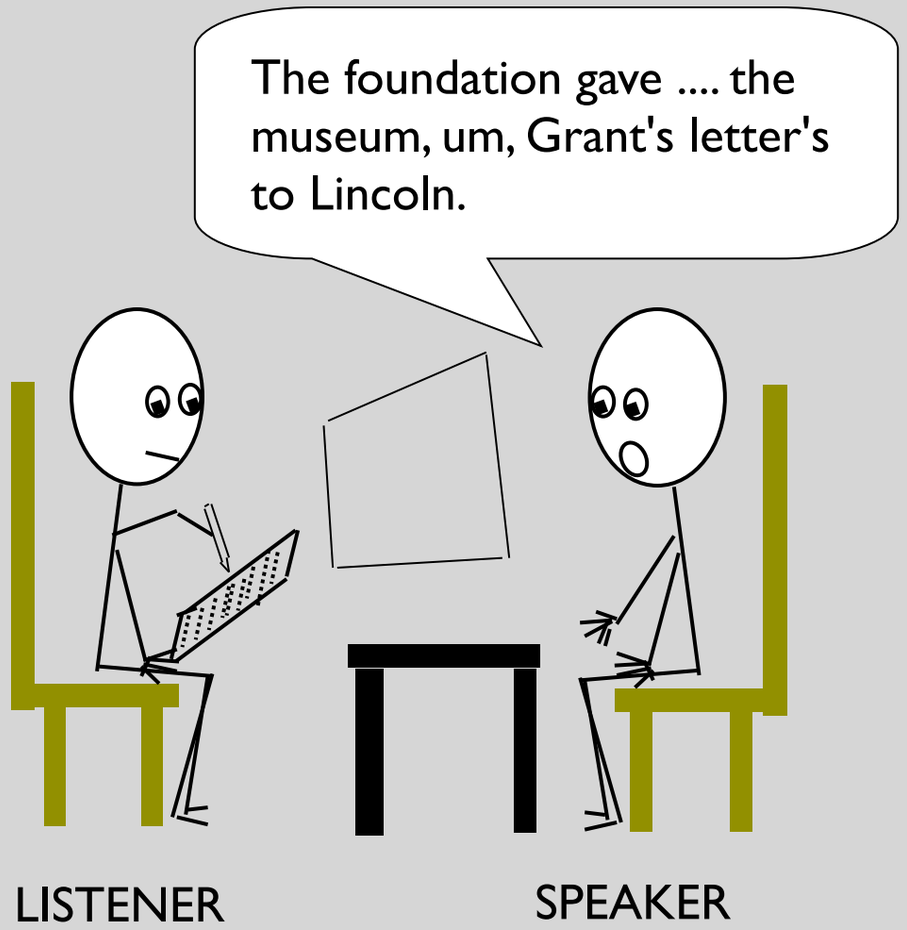
A museum in Philadelphia received Grant's letters to Lincoln from the foundation.

Experimental Method, continued



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

Experimental Method, continued

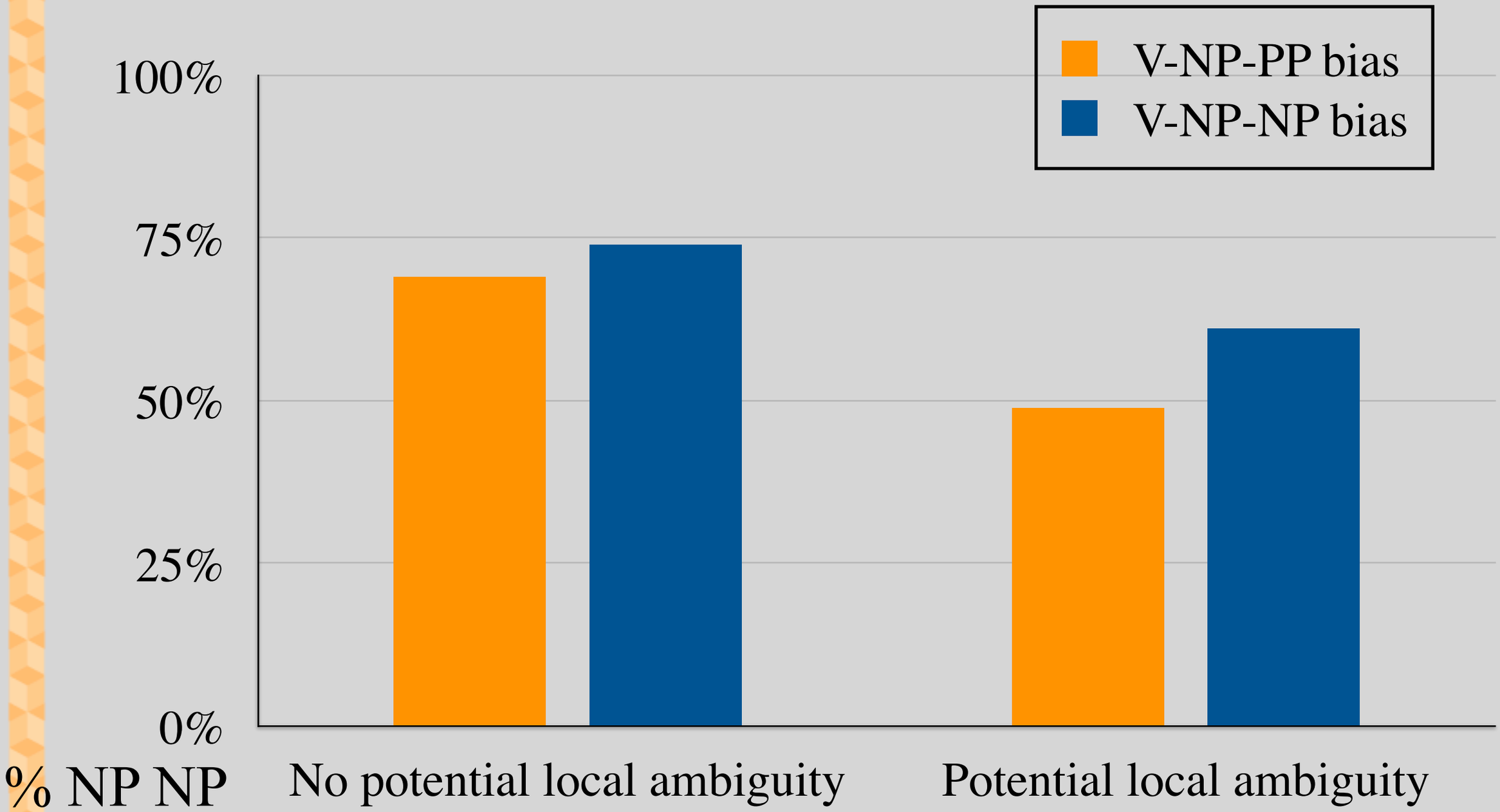


Poll!

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?

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

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 tomorrow
- Due next Saturday (11/12)
- No collaboration
- Send questions to me by email
- Check Canvas read-only midterm Q&A discussion for Q&A :)
  - Questions answered through 11/10

Reading Questions

- Is the discussion on page 303 about our grammar being "process-neutral" referring to the same idea as "the speed of math" and the idea of features/values being shared up and passed down simultaneously?
- Also on page 303, what is a "destructive operation"? I think this term came up in passing in Chapter 8 as well.

Reading Questions

- SOTA language models are very well performing statistical machines. I'm sure some may have made the argument that our brains are processing language in a similar way. This chapter argues that characteristics of this grammar work well with some existing theories about how people process and use language. I'm curious to know if there's a theory that marries some of the ideas from both sides and if computational implementations of this grammar incorporate probabilities. Also, to what extent can we say that the "artificial" methods we develop to enable technology to use language support how our linguistic processes actually work?

Reading Questions

- Do we ever build models (voice assistance for example) to predict the types of 'errors' people intentionally make to better understand users' requests? Or this will be too risky?

Reading Questions

- Do current models of computer language processing have the ability to process language incrementally similarly to humans or do computers need the whole utterance before they can start processing?
- With the current state of parsing, do computers struggle as much as humans do with garden path sentences, or can they actually resolve the ambiguity for some of them better than us now?

Reading Questions

- When I first encountered the discussion of incremental parsing on page 295 I thought it sounded a bit at odds with our desire to treat all components of an expression as sharing information "at the speed of math". While I understand the arguments outlined in the text regarding the utility of developing a constraint-based grammar that is direction-neutral, it doesn't seem to align with the human experience of processing linguistic information. Has any work been done to adapt the grammar so that it takes, for example, a probabilistic approach to layer on constraints incrementally? Also, what order are constraints dealt with in the ERG?

Reading Questions

- Is language processing in difficult circumstances like garden path sentences working in a similar manner to NLP systems that use probabilities to predict results? With no outside context are we essentially just choosing to interpret raced in the way we do because we've seen it in a higher frequency in that form than the form which leads to the correct/grammatical interpretation of the sentence?

Reading Questions

- Although the specific grammar the book is building is specific to English, are the mechanisms underlying linguistic performance (surface-oriented, constraint-based, strongly lexicalist) considered cross-linguistically universal? Do garden path sentences exist in all languages? I would imagine some languages have far less syntactic ambiguity than English. Are typological traits considered in psycholinguistics when approaching how to articulate the "Universal Grammar"?

Reading Questions

- The text says, "Our grammar rules (with the exception of the Imperative Rule) are sufficiently general that, aside from their linear ordering of the constituents, they are natural candidates for universality." I was wondering if this points to any future desire to make the Imperative Rule more general?--as so to align with universality. Or, is it more reasonable to simply accept the Imperative Rule as an idiosyncrasy of English syntax?

- How would our grammar account for polysynthetic languages?

Reading Questions

- Is it actually important that a theory of syntax be something close to how the brain processes language? Couldn't someone claim "I don't particularly care if this theory matches how humans actually understand speech, it's just useful to me as a tool for (insert situation here)"?
- I've heard that there's mild beef between you (Emily) and Chomsky? Is this actually true or is it just a professional disagreement about linguistics?