Ling 566
Nov 5, 2015

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

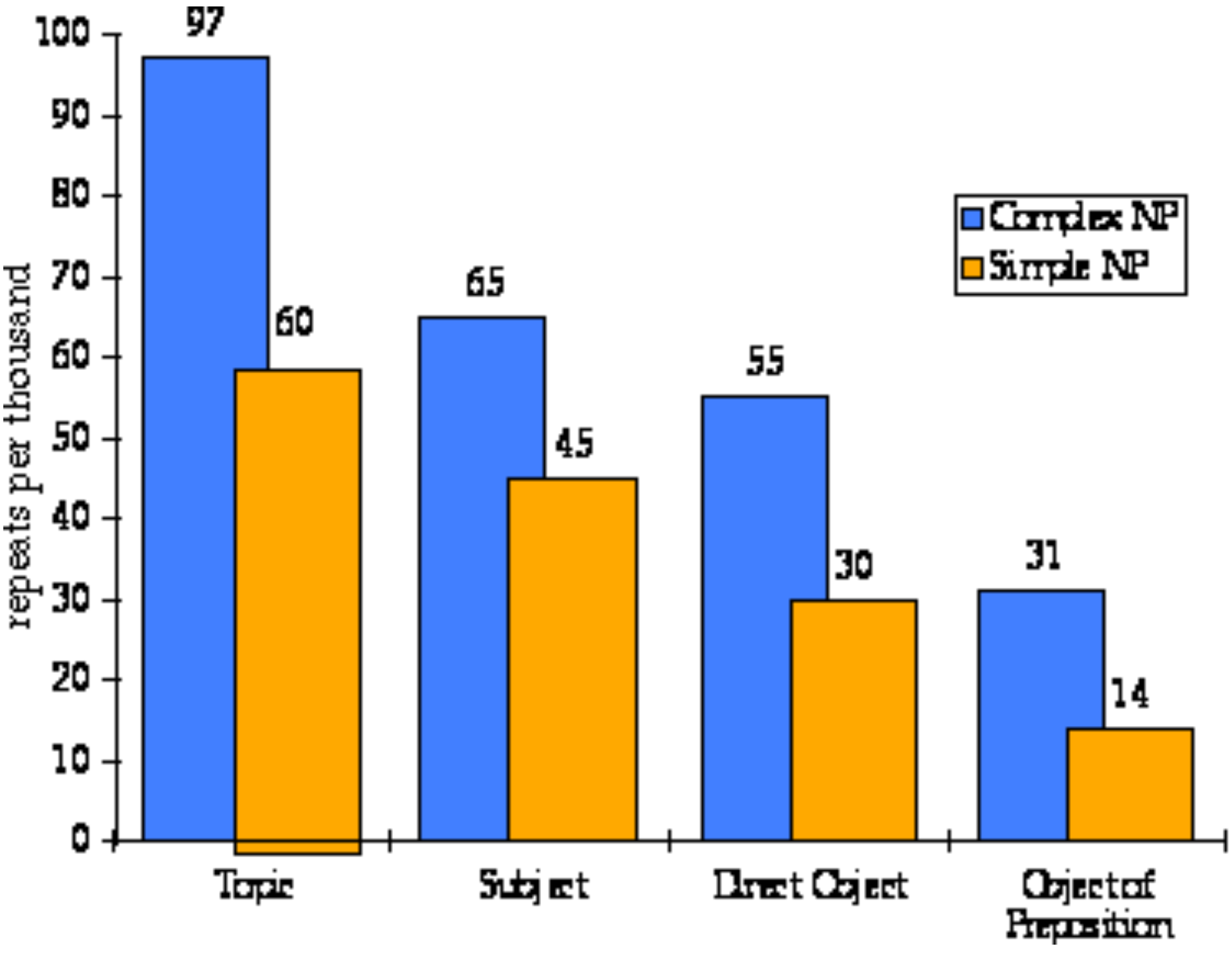
- Chomsky's nativism is very controversial
  - It is based on the "poverty of the stimulus" argument, and a model of learning as hypothesis testing.
  - The environment may be more informative than he assumes.
  - There may be more powerful learning methods than he assumes.
- There has not been much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

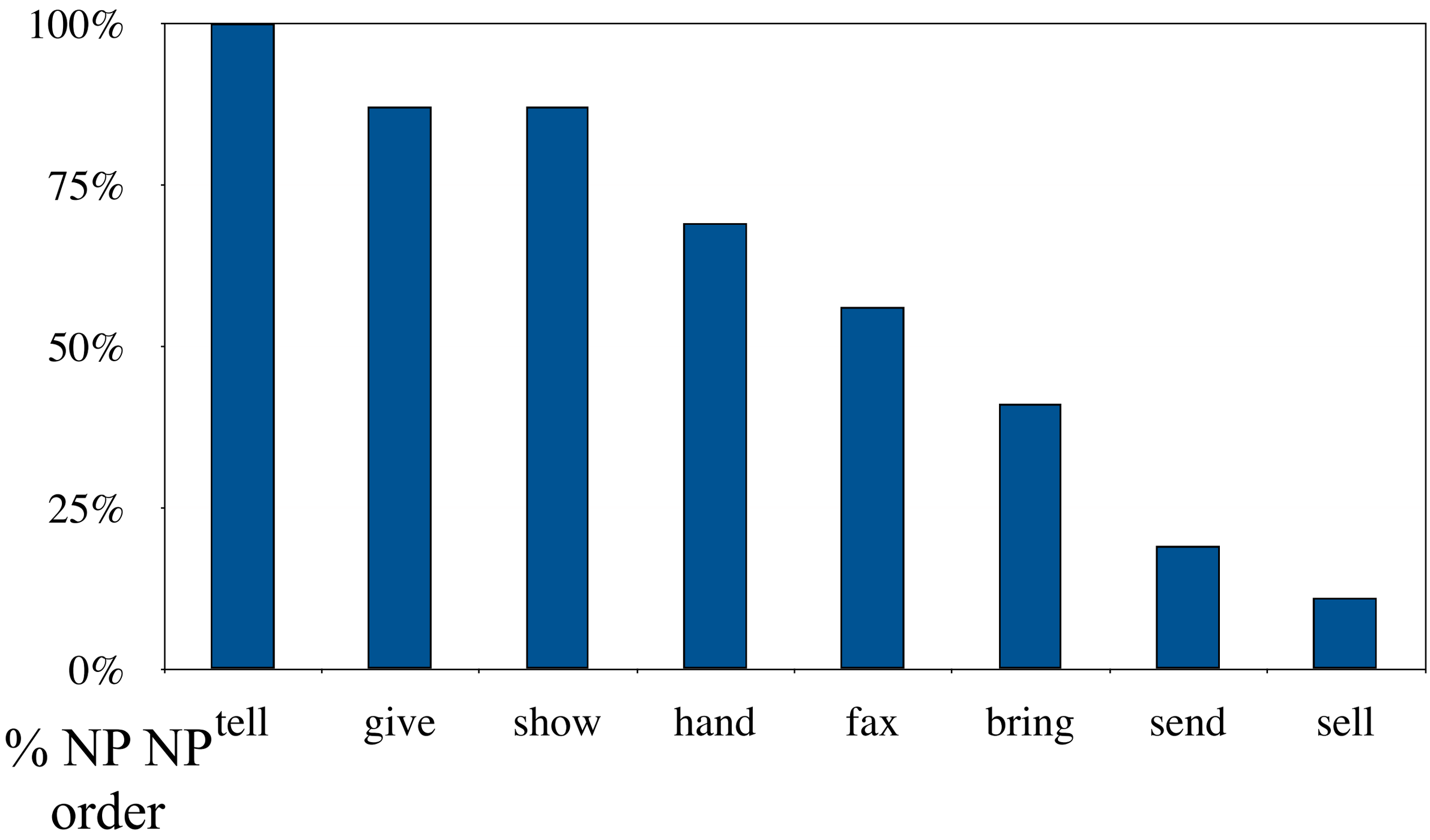
Speakers know a great deal about individual words

- 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

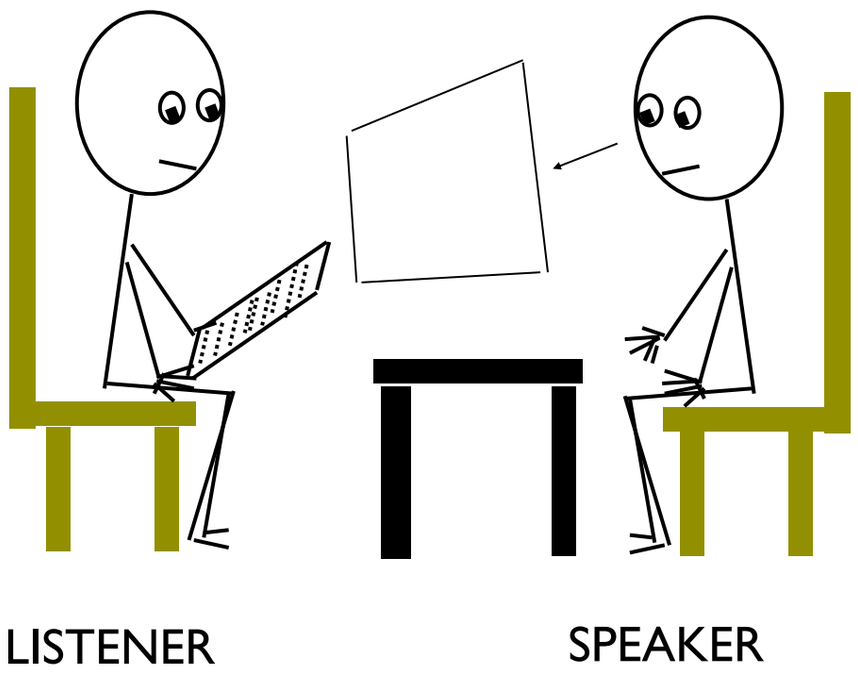
- Wasow et al ran a production experiment to test whether ambiguity avoidance would influence speakers' choice between (1) and (2):

  (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).

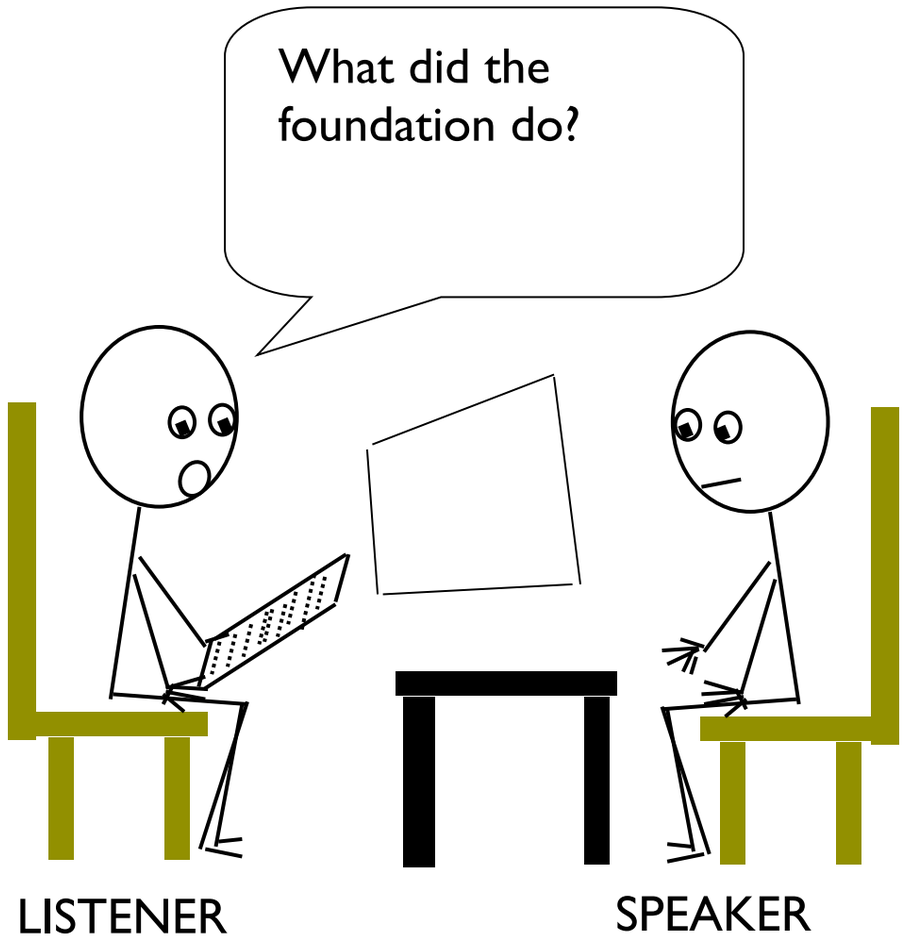
Experimental Method



1. Speaker silently reads a sentence:

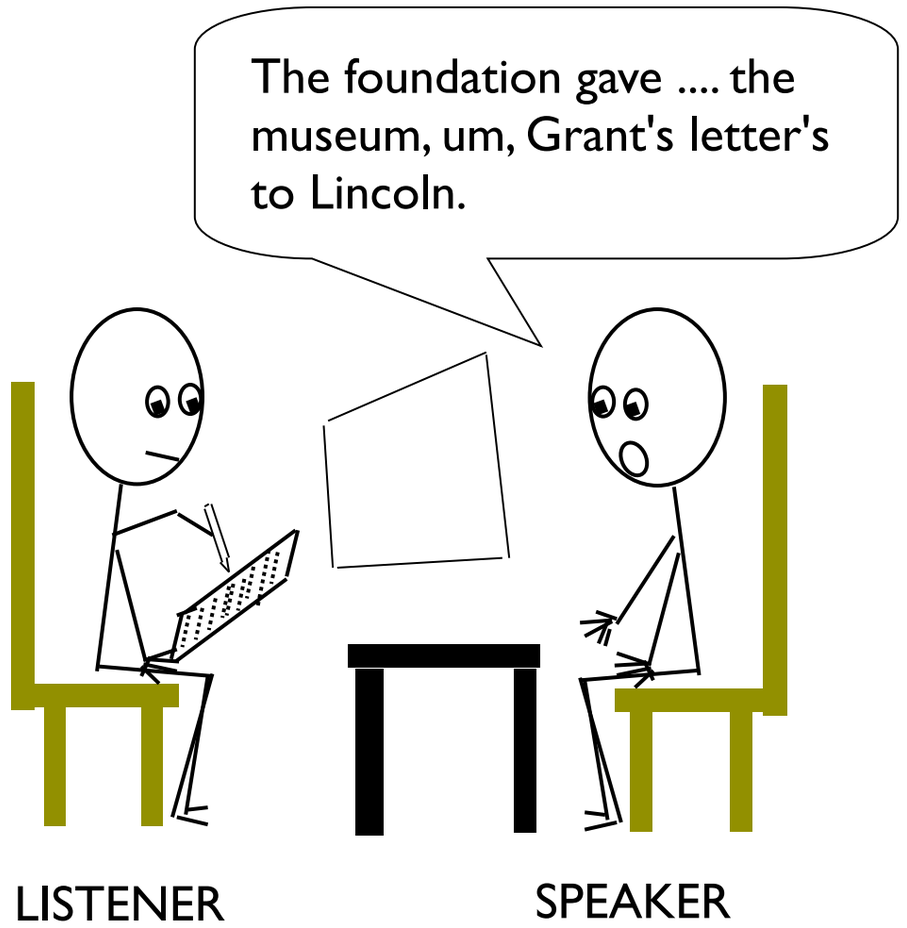
A museum in Philadelphia received Grant's letters to Lincoln from the foundation.

Experimental Method, continued



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

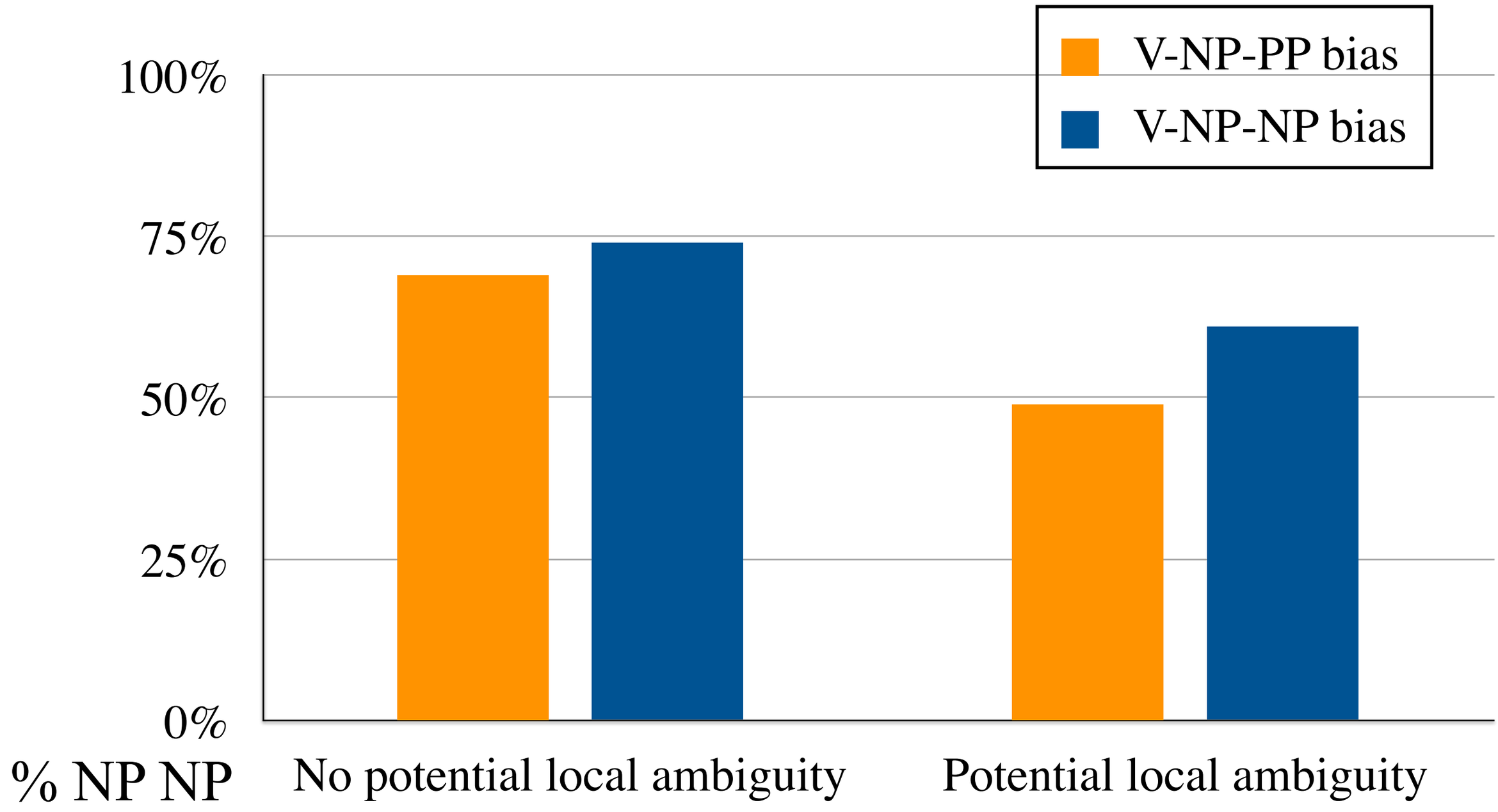
Experimental Method, continued



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

- Why are defeasible constraints not addressed in the satisfaction of feature structure descriptions?
- What is the purpose of modeling a grammar based on psycholinguistic evidence? Creating a grammar to represent natural syntax seems to be an artificial process.

Reading Questions

- What is a performance model and what would it look like?
- How can HPSG be modeling for performance? It doesn't disallow extremely long sentences and garden path sentences, if anything it's seems to be modeling the competence part of the language only.
- If it could be more about modeling performance, it would need some notion of "fuzziness" in accepting or rejecting a structure. Mainly because in realistic natural language some mistakes are more "wrong" than others, and some acceptable constructs are usually avoided more than others (like garden path sentences). And this cannot be modeled in the current HPSG accept/reject process.

Reading Questions

- Although we welcome that our grammar is compatible with many aspects of language performance, we still hold on to the distinction of competence and performance. It is undeniable, that the distinction has been productive in grammar theory of the last half a century, but still a grammar that explains natural language performance is here preferable. Why not going the last steps to gap the c/p distinction and incorporate other elements of natural performance? Is the lexicon such an instrument that allows us to shape a language in such a way that can make it unique? For example, I am thinking about claims there are languages (like Piraha) do not have recursion or just limit recursion in a practical way, considered otherwise a universal characteristic of languages.

Reading Questions

- But what if we wanted a computer parser to stumble over garden path sentences? Could we model human incremental linguistic processing so closely that the computer makes the same predictions about the "right" interpretation of a sentence? Besides being an interesting model of psycholinguistic processes, I'm curious if this has been explored in computational linguistics and AI.

Reading Questions

- I understand that we want a process-neutral grammar that can handle both production and comprehension, and that directionality in transformational grammar is problematic when it comes to actual processing. Would it still be problematic if the directionality were parallel to the "direction" of language? incrementally in this way?
- L-to-R processing in HPSG?

Reading Questions

- I'm struggling to understand the debate over task-specificity. Wouldn't the proposal of any generative grammar implicitly support the theory of universal grammar and task-specificity?
- Surface oriented: is the "surface" here referring to words serving as the leaves on the end of all the branches and the order of those words?

Reading Questions

- I'm having trouble understanding how the input/ output schema of HPSG's lexicon works with the idea that a performance-based grammar is surface-oriented. The syntax itself is surface-oriented because it doesn't undergo transformations and have an underlying form like other grammars do. However, our lexicon does have an underlying form (a lexeme), so I'm unsure on how this fits in with being surface-oriented. I think I'm either misunderstanding what surface oriented means, or how the lexicon represents information, but I'm not sure which.

Reading Questions

- Is there research that supports the existence of a mental analogue to, say, the lexeme type hierarchy? Obviously it makes sense to talk about transitive verbs and intransitive verbs as subtypes of verbs. That is a very natural classification when we consciously think about the structure of verbs, but is there evidence that such hierarchical divisions take place subconsciously?

Reading Questions

- I'm wondering what other types of non-linguistic information are used to process language and how this relates to the question of modularity. For instance, we know that eye movement is likely related to language processing. Other non-linguistic information that is used to process language include viewing expressions and body language, hearing tone, awareness of situational context, etc. If a person is using a number of different senses and types of knowledge to process language, then is that an argument for modularity?

Reading Questions

- Much of the work we have done up to this point, we aim for 'well-formed' structures. But, we still understand a four year old's language, a noisy phone call, a "tweet", or a slightly-less than perfect translation, and so on. On one side, these examples strongly support the importance of the lexicon and incremental processing -- mix-up or drop a few words, and we still comprehend the utterance solely based on the available words. But, t makes me wonder how this type of formalism could model spoken language, or if we are only talking about generated language for the near future...

Reading Questions

- In the section on Universal Grammar, the book mentions that, although phrase structure rules depend on ordering, "it would not be hard to factor out the ordering, so that versions of these rules could be posited as part of universal grammar." How could this be done? Does the same apply to the SEM feature - could order be factored out?

Reading Questions

- Following work done by DELPH-IN and others in HPSG on numerous languages, what, if anything, has emerged as universal between them all? An on a related note, wouldn't an attempt to develop HPSG to account for all languages wind up being far too broad and general to actually be useful in describing any particular language?

Reading Questions

- I wonder how the approach that this book is following in understanding and formulating natural language aligns with the current advancements in deep neural networks research for NLP. The current research shows that with a generic representation of words and relations between them computed from a vast amount of data, we can build systems that perform different tasks of NLP (parsing, tagging, etc.) with a very high accuracy. How the approaches and constraints in this book aligns with these observations?