

Ling 566
Nov 5, 2020
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.

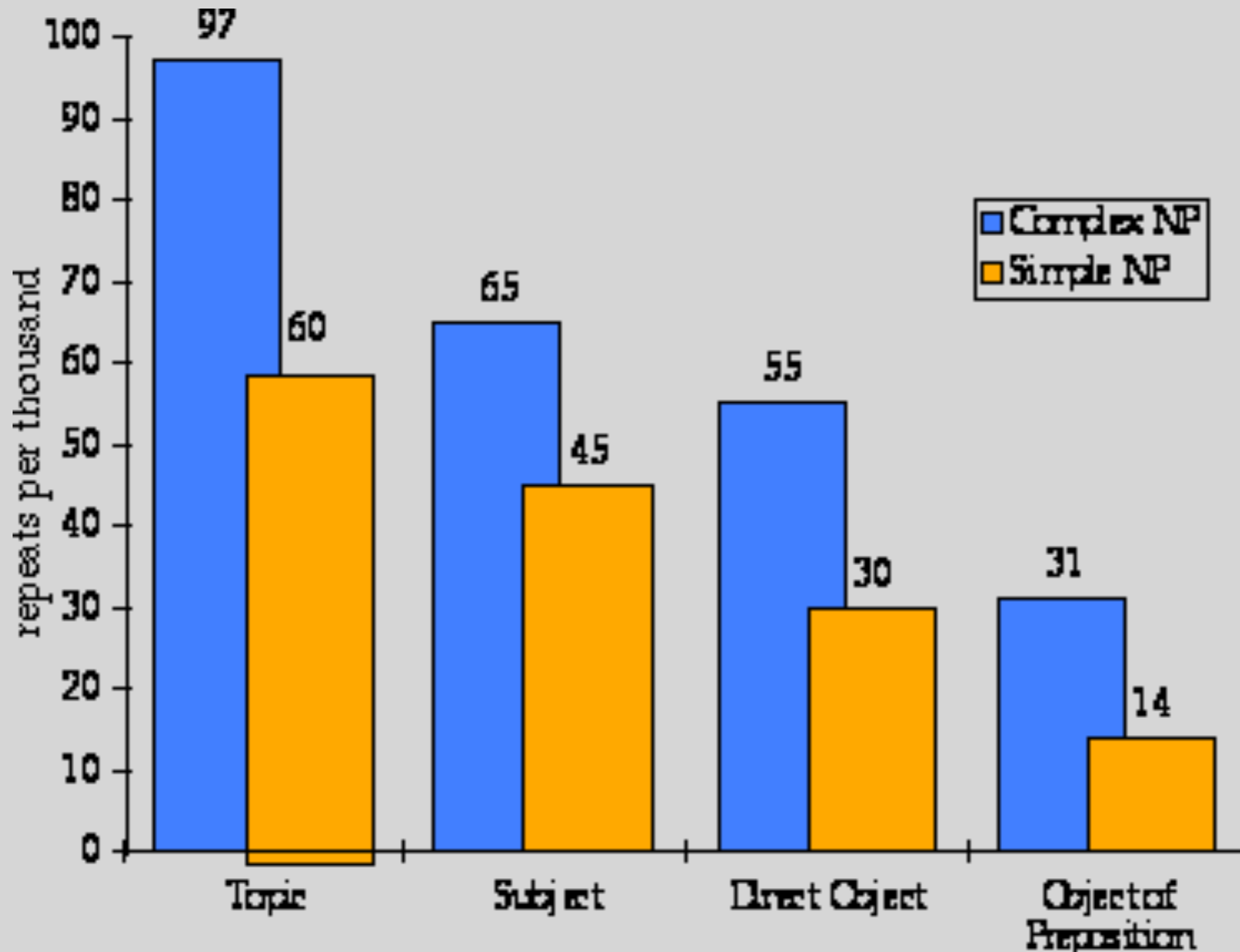
Poll!

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.

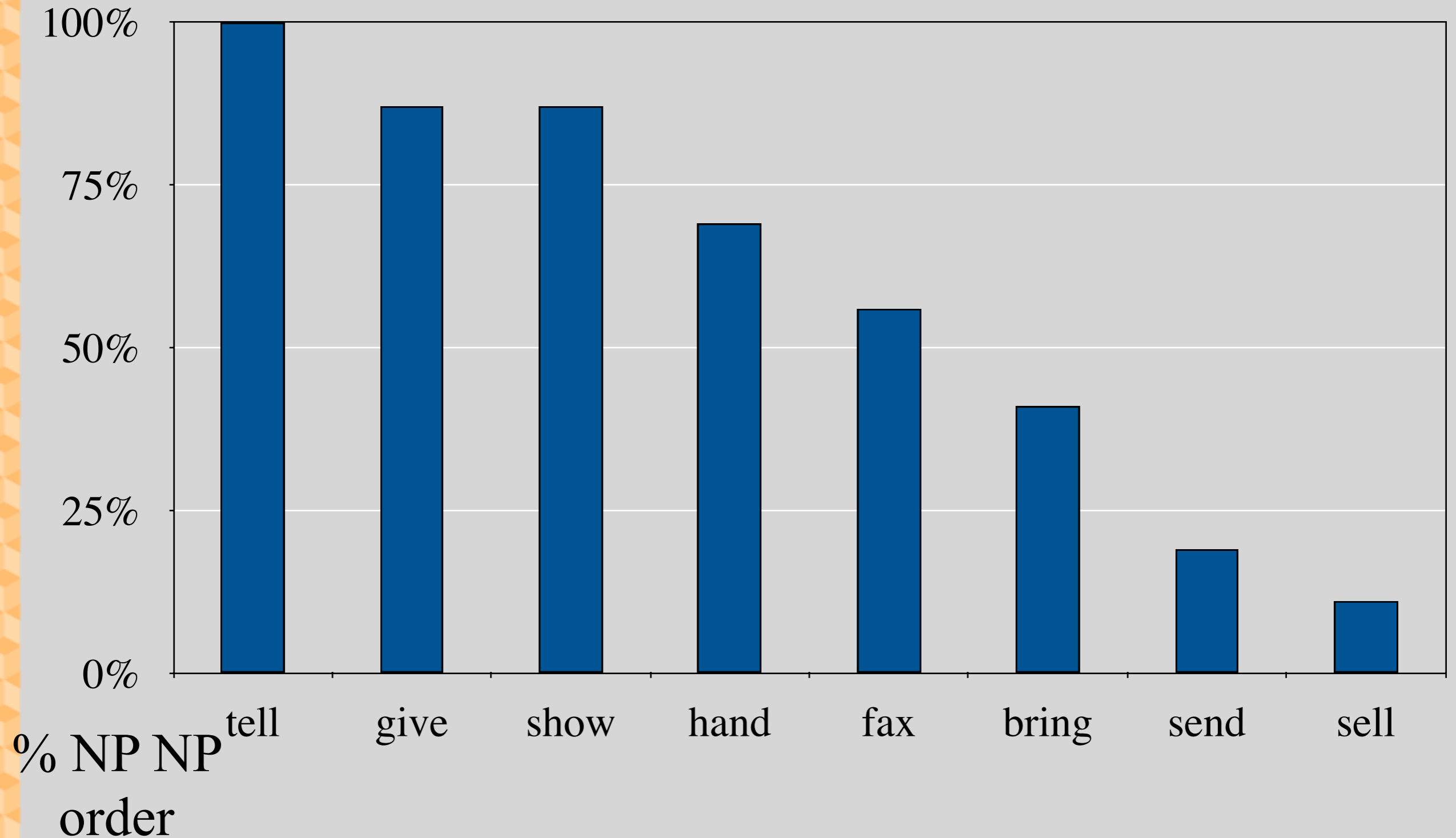
General Conclusion of Eye-Tracking Studies

- 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.

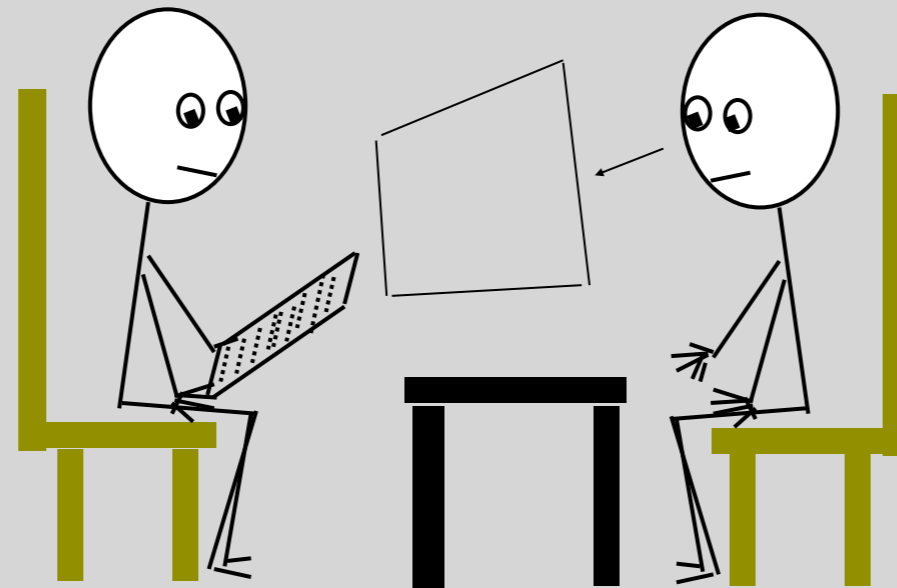
V-NP-NP vs. V-NP-PP Frequency in the *NYT*



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



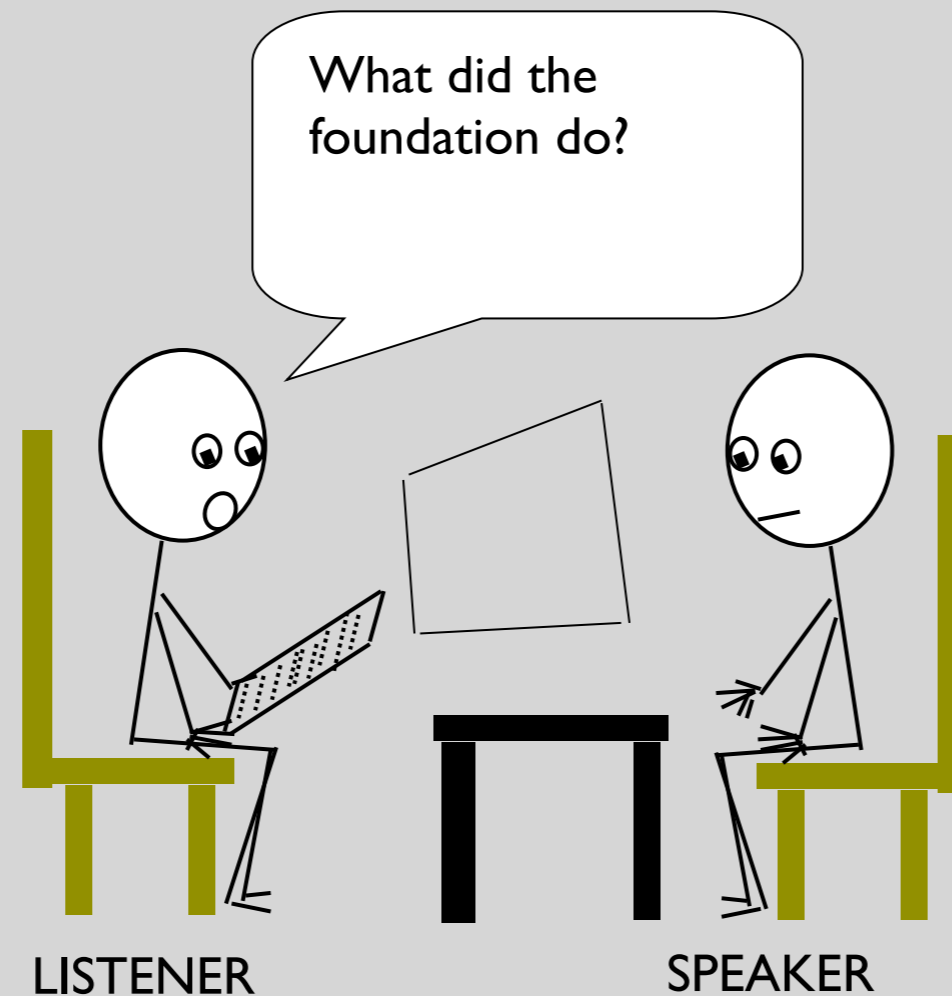
LISTENER

SPEAKER

1. Speaker silently reads a sentence:

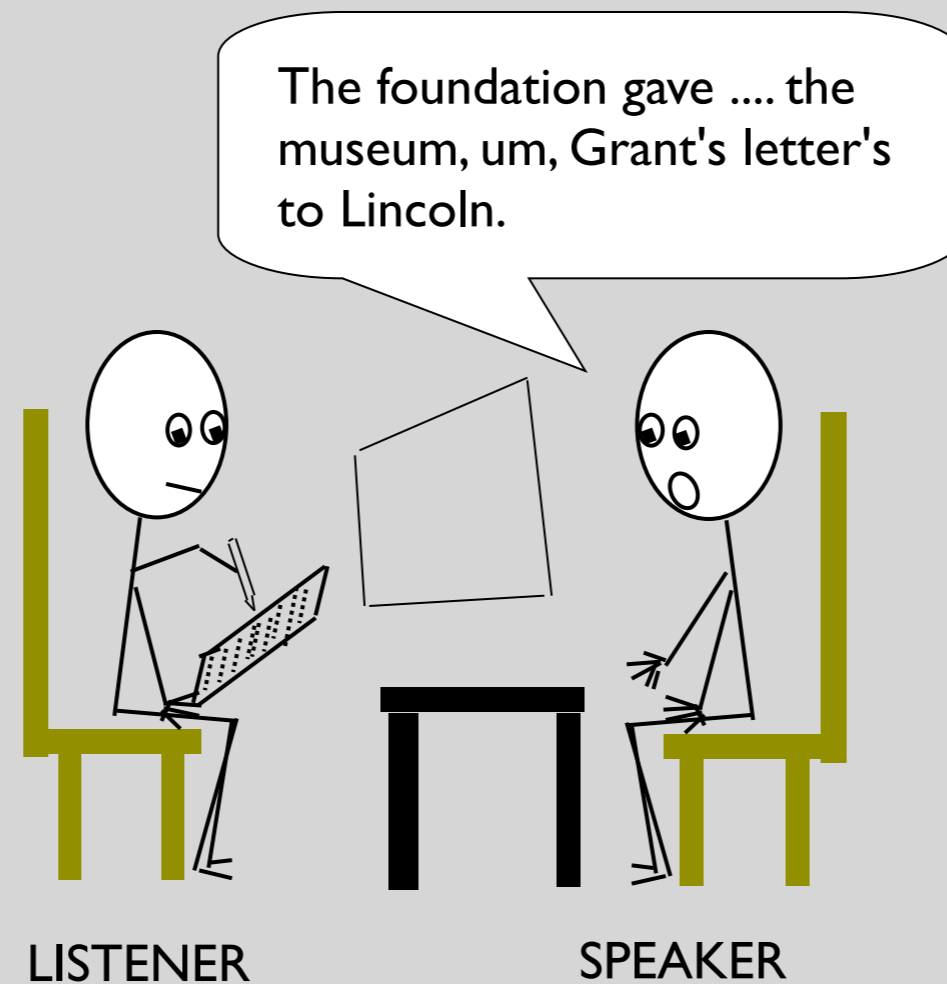
A museum in Philadelphia received Grant's letters to Lincoln from the foundation.

Experimental Method, continued



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

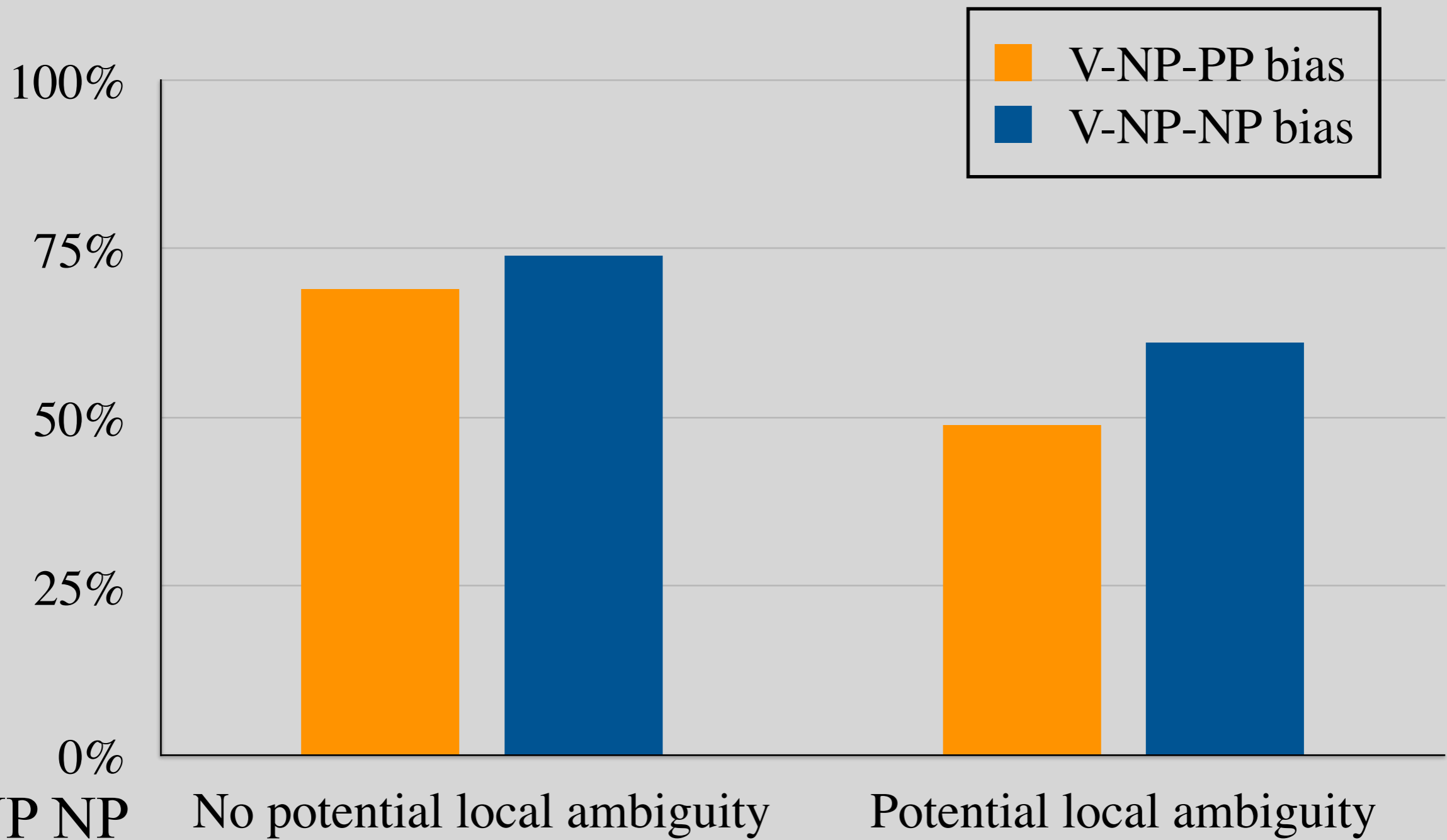
Experimental Method, continued



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



% NP NP

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

Midterm

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

Reading Questions

- Are we saying that the grammar developed in this textbook is "better" than other grammars because it has properties that seem to correspond to the way the brain handles language? We can't say that this grammar is definitely correct in its correlation to the human brain, so this chapter feels a bit speculative to me. Why is this speculation so valued by computational linguistics?
- So is our grammar based off of how humans process language (psycholinguistic aspect) or how to represent grammar in a more "structured way" that can easily be translated programmatically or both?

Reading Questions

- What is the relation between organizing a grammatical sentence and expressing a (semantic) thought? Someone can definitely speak ungrammatical English while still being semantically understood, which is rather common in L2 speakers. On a side note, relating to the debate covered in this chapter, Chomsky's UG seems to be more about organizing grammatical sentences while Tomasello's communication seems to be more about expressing thoughts (in social contexts).
- One thing that was emphasized was how syntactic ambiguity is often resolved using context, and how repeating uncommon structures can prime processing for those structures in later sentences. Is this nonlinguistic information (i.e. general knowledge, times a structure has occurred, etc.) something that can potentially be embedded into our grammar, in a similar fashion to how we added SEM as a feature?

Reading Questions

- Is non linguistic information mentioned in (52) related to SEM in this grammar? It seems to me that we are not consider meaning as a linguistic feature.
- In the chapter, the term 'nonlinguistic' knowledge is used frequently. I was wondering can there always be a clear distinction between linguistic and non-linguistic knowledge? It seems to me that they sometimes overlap, and a lot of the times nonlinguistic knowledge is used to refer to semantics.

Reading Questions

- I had trouble understanding 9.6 and its breakdown of the theory's components relating to task-specificity. The book explains how the structure hierarchy isn't in its usage for phase structure rules. From what I understand, this is support for language as a side effect, thus in opposition of task-specificity? How does the differentiation between constraints belonging to a universal grammar versus particular languages as seen in types and the type hierarchy contribute to this conversation?

Reading Questions

- I am curious about how our minds resolve ambiguous sentences quickly. We have seen many phrases/sentences with ambiguous meanings, but it seems that we are able to select a meaning that "makes more sense" when we hear or see such phrases/sentences. Is it purely a statistical process that we resolve, or is there deeper psycholinguistic explanation to it?

Reading Questions

- In 9.5 the text says "More generally, language understanding appears to be a process of constraint satisfaction." Even though that seems to be what we've been working up to until this point by combining constraints, I still hadn't thought of this way. Is this to say, in other words, that when we're listening to a sentence and interpreting it, we're understanding its meaning to some extent based on the things that it could not mean?

Reading Questions

- One thing I'm curious about is to what extent (if at all) lexical frequency information is used to resolve ambiguity when developing a grammar for a language. Is it useful for a grammar to have access to the most likely definition for a given word (either a priori, or within the context of a sentence)? Or, relatedly, I was thinking about garden path sentences - are there applications of grammar engineering in which we'd want to know that humans will have some difficulty parsing "The horse raced past the barn fell"?

Reading Questions

- Chapter 9 talks a lot about modeling performance. I am curious whether or not a second language speaker having similar or same performance, or having additional steps of modeling language compare to native speakers.
- Footnote 16 (which is another absolute banger of a footnote) also points out that comprehension can extend beyond production because it's easier for hearers to "allow grammatical principles to relax" in order to understand an utterance than it is for speakers to be understood when they produce ungrammatical sentences. Is that why it's easier to understand a foreign language (that you're in the process of learning) than it is to speak it, or is there something else going on there?

Reading Questions

- I am also intrigued by the assertion on page 302 that "By 'fairly uniformly' we mean that the set of sentences reliably producible by a given speaker-hearer is similar – in fact bears a natural relation (presumably proper inclusion) – to the set of sentences that that speaker-hearer can comprehend." I'm thinking about how this works in the case of "heritage" speakers who grew up with a language at home and are largely able to understand what is spoken to them, but who are unable to produce much themselves. I don't think this phenomenon disproves what the text is saying about a similar process working in various types of language activities, but I am interested in what causes it.

Reading Questions

- Can we talk about how incremental and rapid language processing apply to second language speakers? I think my ability to process English or other second languages is noticeably different/slower from people who speak English as their first language. Language comprehension is still working incrementally as other people talk to me, but my perception and "real-time analysis" on the sentence meaning could be different. One example is that I did not find the Garden Path sentence "The horse raced past the barn fell" to be ungrammatical at all. It actually took me a while to realize that it's a weird sentence. Even now when I hear the sentence again, I still feel the same way. Or maybe, one can argue that I am not able to find the Garden Path example weird because I do not process English as rapidly as English native speakers.

Reading Questions

- I was wondering if the semantic parsing of sentences in languages that have different sentence structure (SVO vs. SOV) than someone's native language (lets say a native English speaker learning Japanese for example) have more of a interpretational delay? Based on some of the examples of the sentences in this chapter, word order shouldn't entirely matter since we interpret semantic representations with each lexical component in the sentence right?
- This chapter discusses language learnability, and I was wondering how the theories for/against universal grammar relate to second language acquisition? Does the difficulty of learning a second language (after a certain age) weaken the argument for universal grammar, or is there a plausible explanation compatible with universal grammar?

Reading Questions

- I am not quite seeing how this grammar is well suited to the theory of universal grammar. It seems that that elements of the grammar that may be universal are mainly features and types, while the structures formed from these and the ways words fit together are mostly language specific. To me, this feels more like an argument against universal grammar, rather than fitting well with it.

Reading Questions

- In 9.4, there is a good explication on how humans process text incrementally as new words are said/written. I wonder why we have things choose their specifiers in the grammar, rather than the other way around, which would seem to follow the order our brains process text more closely. The order of the head complement rule seems to be more aligned with our mental ordering.

Reading Questions

- How do we know if these partialities are similar enough? I guess I am wondering if there is any rule governing a speaker's partial interpretation and if our grammar picks the same increments as a speaker would to make a complete interpretation. The example of "echo questions" might be a simple one since it is a repetition, but in other contexts with so many linguistic and paralinguistic pieces of information, can we ensure that our grammar represents the same "parts" for a complete interpretation as a speaker would?

Reading Questions

- The chapter makes the case that strong lexicalism is very desirable because it allows us to store a lot of information in the words themselves, and this makes good psycholinguistic predictions. But what would an implemented HPSG parser or a hypothetical human using HPSG to process do with words they don't know? I can imagine a lot of the information can be gleaned from context alone (probably almost all of SYN and most of SEM), but I'm curious what the formalism for that process would look like.

Reading Questions

- Is the Imperative Rule not considered a destructive operation? Doesn't it "prevent" a verb from realizing its specifier?

Reading Questions

- This chapter discusses many strong points in which the grammar that we have been learning about matches well with evidence from psycholinguistics. From the opposite end, are any parts of the grammar that potentially conflicts with psycholinguistic evidence?
- This chapter argues that the theory in this textbook has properties that align well with empirical accounts of linguistic performance. Since this is my first syntax class, I'm curious about what arguments there are in favor of other theories of grammar in terms of this criterion.

Reading Questions

- With an eye towards respecting the difficult nature of inquiry in linguistics and its near neighboring disciplines, where within syntactic theory should we take care not to RULE OUT parts of an explanation of grammar competence? I'm still new to the history of syntax, but if there are any valuable lessons in that area I'd be especially interested. I'm wondering in particular about strong claims of reduction, like what happened in the early 2000s with neuroscience to formal semantics, where even weak claims about computation in the mind were declared explanatorily redundant since everything could be reduced to particular weights in a network.

Reading Questions

- Since the chapter went through a summary of our grammar so far, it got me thinking about things our grammar may not be able to handle yet. For example, how do we model things like when people use just tones to say things like "I don't know" or "Hmmm" with a rising tone to indicate questions. This question is probably just an enrichment question, but I'm interested to how we model these kinds of tonal utterances in English even though English isn't a strictly tonal language.
- Are there serious applications of HPSG towards modeling speech over text? Or are these elements largely irrelevant to syntax theory?

Reading Questions

- For an actual system that aims to implement our grammar introduced in the book, how should we cover most of our lexeme entries in English, for example? To my knowledge, many words in English have two or more entirely different meanings, such as "arrest" which means both "to bring to a stop" and "to catch suddenly and engagingly." How would a real system capture every meaning of every word?

Reading Questions

- Is it also within the scope of psycholinguistics to discuss how what language we speak affect how we process events and the world around us?
- I'm curious about how the nature of a language affects how people queue up information in their brains and organize their thoughts about a topic, such as when switching between SVO and SOV languages. How quickly is the queue populated while mid-speech? How much fundamental difference is there in the mental modeling of the content, aka what feels important to the speaker about the situation being described? Do differences in average word lengths/sentence lengths of a given language affect the speed of speaking/understanding/reading of its speakers when compared to other languages? Do they manifest culturally? How do differences in precision vs ambiguity (e.g. homonym prevalence in a language) affect how quickly listeners of that language disambiguate meaning?

Reading Questions

- Non-linguistic data being used to aid in language disambiguation is something I've thought about for a while. When working on a virtual assistant, we wanted to be able to leverage previous transactions for understanding. For example if a customer asked "how much was my total for my previous visit", the assistant understands previous to be your latest visit last week. What if any other forms of non-linguistic data have people used for language understanding? Is it necessary to have background knowledge of a particular speaker?