

Class 7:

Noisy channel / Memory in sentence processing

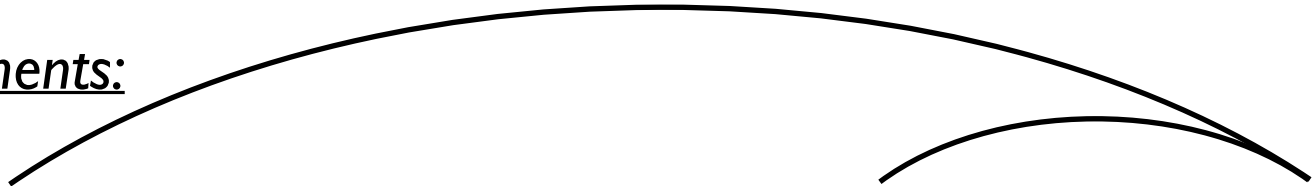
Ted Gibson
9.59J/24.905J

Information processing: Working memory

Working memory: Local connections are easier to make than long-distance ones (Gibson, 1998, 2000; Grodner & Gibson, 2005; Warren & Gibson, 2002; Lewis & Vashishth, 2005; Hawkins, 1994)

Ambiguous attachments:

The bartender **told** the detective that the suspect **left** the country **yesterday**.



yesterday is preferred as modifying **left** rather than **told**

(Frazier & Rayner, 1982; Gibson et al., 1996; Altmann et al., 1998; Pearlmutter & Gibson, 2001)

Unambiguous connections:

The **reporter wrote** an article.

The **reporter** from the newspaper **wrote** an article.



The **reporter** who was from the newspaper **wrote** an article.



Retrieval / Integration-based theories

Integration: connecting the current word into the structure built thus far: Local integrations are easier than longer-distance integrations

- The Dependency Locality Theory (DLT) (Gibson, 1998; 2000): intervening **discourse referents** cause retrieval difficulty (also in production)
- Activation-based memory theory: similarity-based interference (Lewis & Vasishth, 2005; Vasishth & Lewis, 2006; Lewis, Vasishth & Van Dyke, 2006): **intervening similar elements** cause retrieval difficulty
- Production: Hawkins (1994; 2004): **word**-based distance metric.

Consequence: Nested structures are difficult crosslinguistically

English:

The reporter [who the senator attacked] admitted the error.

The reporter [who the senator [who I met] attacked] admitted the error.

I met the senator who attacked the reporter who admitted the error.

Japanese:

Obasan-wa [bebiisitaa-ga [ani-ga imooto-o ijimeta] to itta] to omotteiru
aunt-top babysitter-nom older-brother-nom younger-sister-acc bullied that said that
thinks

“My aunt thinks that the babysitter said that my older brother bullied my younger
sister”

Easier: Bebiisitaa-ga [ani-ga imooto-o ijimeta] to itta] obasan-ga to omotteiru

Locality account of nesting complexity

Nested structures have **longer distance dependencies** than non-nested structures.

The reporter [who the senator [who John met] attacked] disliked the editor.

John met the senator [who attacked the reporter [who disliked the editor]].

An alternative account of nesting complexity: Nested structures have parse states with **more incomplete dependencies** (e.g., Yngve, 1960; Chomsky & Miller, 1963).

Locality account of nesting complexity

Problematic cases for incomplete-dependency approaches: Relative clauses (RCs) and sentence complements (SCs) (Cowper, 1976; Gibson, 1991):

RC within SC: difficult, but processable

The fact [that the employee [who the manager hired] stole office supplies] worried the executive.

SC within RC: much harder to process

The executive [who the fact [that the employee stole office supplies] worried] hired the manager.

Same maximal number of incomplete dependencies, parsing left-to-right: 3 incomplete subject-verb dependencies, plus one incomplete filler-gap

Solution: Distance-based integration accounts. The RC filler-gap dependency between “who” and its role assigning verb (“hired”) in (1) is more local than the RC filler-gap dependency between “who” and its role assigning verb (“worried”) in (2).

Locality account of nesting complexity

The lower complexity of examples **nested pronouns** (Bever, 1974; Kac, 1981)

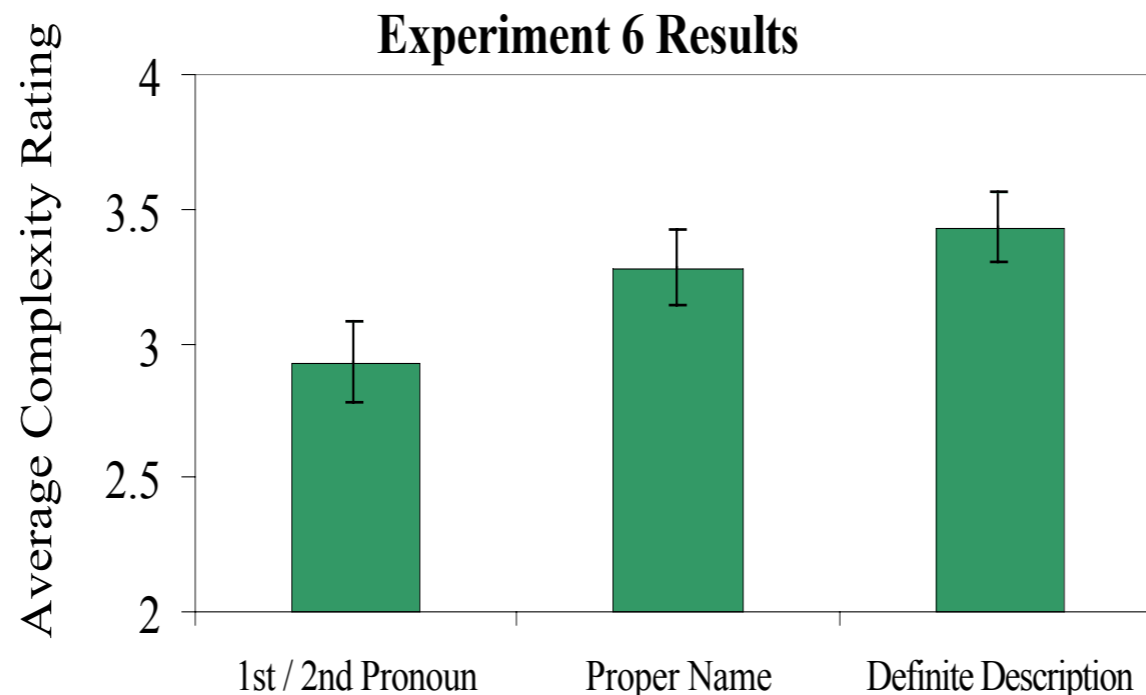
The reporter who everyone that I met trusts said the president won't resign yet.

A book that some Italian who I've never heard of wrote will be published soon by MIT Press.

The reporter [who the senator [who John met] attacked] disliked the editor.

Warren & Gibson (2002), Experimental evidence:

The reporter who the senator who { **you / John / the professor** } met attacked disliked the editor.



Locality account of nesting complexity

Gibson (1998, 2000): Decay

Discourse-based decay hypothesis: The difficulty of integrating a new word h_2 to h_1 is proportional to the number of discourse objects and events (nouns and verbs, roughly) which were introduced since h_1 was last processed. (cf. Warren & Gibson, 2002)

Hawkins: word-based decay hypothesis

Interference of similar elements in the intervening structure:

NP types: Gordon & colleagues

Phrase structure similarity: Lewis, Vasishth, McElree and colleagues

The syntactic / semantic similarity of intervening NPs: More similar NPs, slower processing

Prediction: Same kinds of NPs as head noun and embedded NP in an objected-extracted RC will lead to most processing difficulty, independent of the NP type

Clefts:

It was (the barber / John) that (the lawyer / Bill) saw in the parking lot.

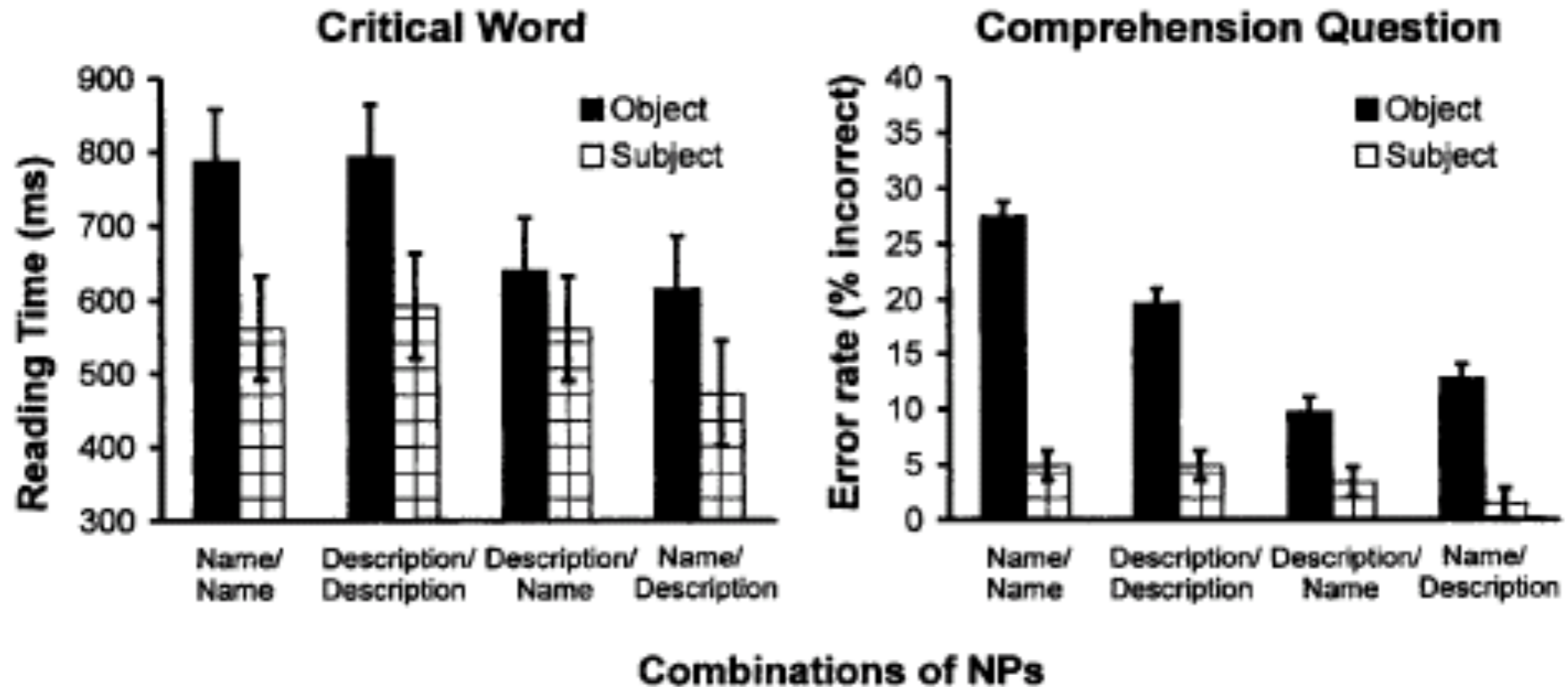
It was (the barber / John) that saw (the lawyer / Bill) in the parking lot.

Gordon et al. 2001, Experiment 4

Clefts:

It was (the barber / John) that (the lawyer / Bill) saw in the parking lot.

It was (the barber / John) that saw (the lawyer / Bill) in the parking lot.



Courtesy of Journal of Experimental Psychology. Used with permission. Source: Gordon, Peter C., Randall Hendrick, and Marcus Johnson. "Memory interference during language processing." Journal of Experimental Psychology: Learning, Memory, and Cognition 27, no. 6 (2001): 1411.

Gibson & Thomas (1999): Structural forgetting: The “missing verb” effect

1. *The **apartment** that the **maid** who the **cleaning service sent over was well-decorated.** 👍
2. The **apartment** that the **maid** who the **cleaning service sent over cleaned was well-decorated.** 👎

(Frazier, 1985, reporting an intuition from Janet Fodor)

Gibson & Thomas (1999): Structural forgetting: The “missing verb” effect

Gibson & Thomas (1999): acceptability judgements on:

All 3 VPs: *The ancient manuscript that the graduate student who the new card catalog had confused a great deal was studying in the library was missing a page.*

Missing VP1: *The ancient manuscript that the graduate student who the new card catalog [] was studying in the library was missing a page.*

Missing VP2: *The ancient manuscript that the graduate student who the new card catalog had confused a great deal [] was missing a page.*

Missing VP3: *The ancient manuscript that the graduate student who the new card catalog had confused a great deal was studying in the library []*

Ratings (high is difficult):

All three VPs	Missing VP1	Missing VP2	Missing VP3
2.90 (.12)	3.58 (.14)	2.97 (.14)	3.53 (.14)

Vasishth et al. (2010): Replication & extension to self-paced reading & German

Vasishth et al. (2010): SPR:

All 3 VPs: *The carpenter who the craftsman that the peasant carried hurt supervised the apprentice.*

Missing VP2: *The carpenter who the craftsman that the peasant carried [] supervised the apprentice.*

Faster RTs, better comprehension accuracy on Missing VP2 example

Christiansen & Chater (1999)

Simple recurrent network explanation of G&T missing verb effect

Is this a good explanation? What is an explanation?

Orthogonally, how would you test this hypothesis?

Gibson, Fedorenko & Mahowald (2014): New paradigm for assessing syntactic WM: sentence completions of complex materials

Expt 2: 60 participants on Mechanical Turk

Design: 4 critical conditions (6 items each) + 2 control conditions

Task: Complete each preamble to make a complete sentence.

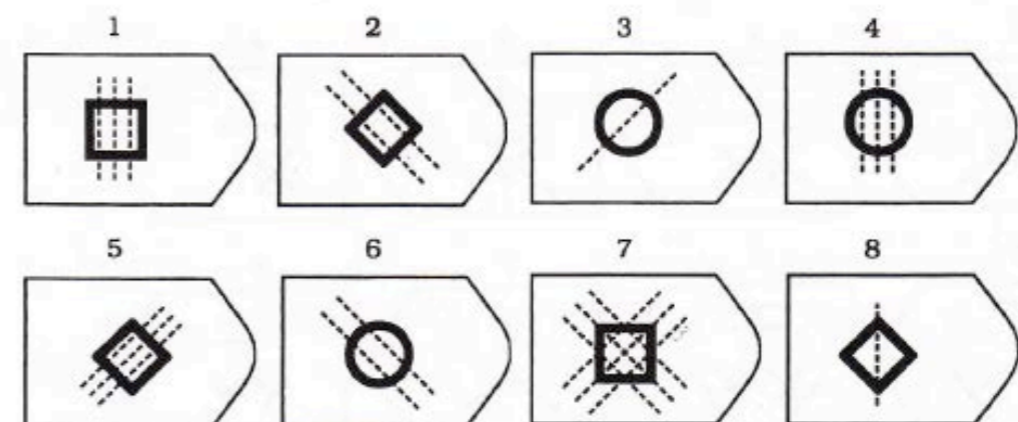
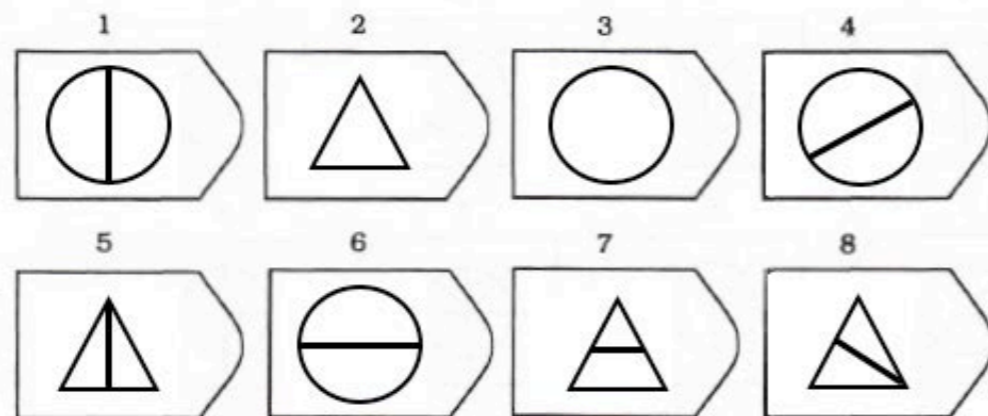
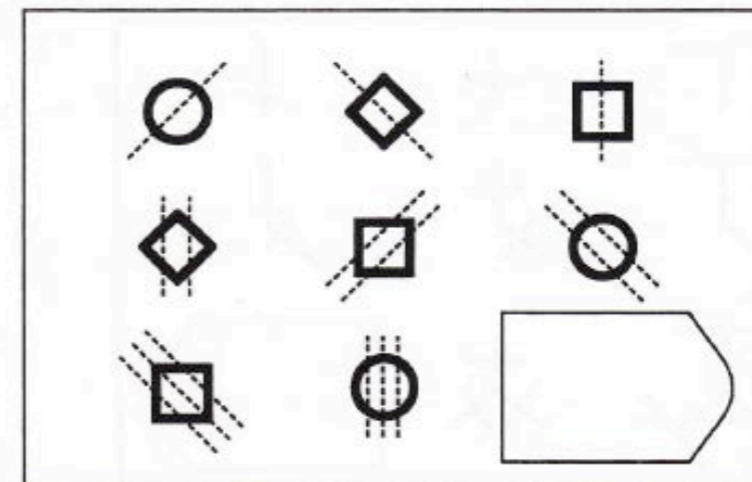
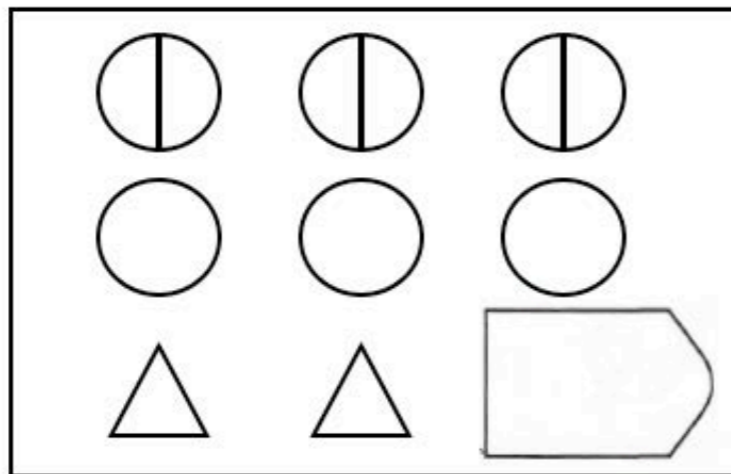
ORC-anim/SRC	The reporter who the professor who...
ORC-inan/SRC	The manuscript which the student who...
SC/ORC	The fact that the professor who the diplomat...
SC-verb/ORC	The rumor stating that the suspected mobster who the media...
<u>Control 1: ORC</u>	The veterinarian who the...
<u>Control 2: SRC</u>	The fencer who...

Experiments 1 & 2

60 participants on Mechanical Turk

The non-language task: Ravens Advanced Progressive Matrices (RAPM; Raven et al., 1988) (a very general fluid intelligence task; has been shown to correlate with many WM and cognitive control tasks, like Stroop & digit span and many others)

Task: Choose a picture (out of 8 possibilities) that best completes the set.
40 trials.



Experiment 2 results: Sentence completions

For analyses of the critical conditions we only included participants (n=55) who grammatically completed at least 5/6 of each of the control conditions (simple SRCs, simple ORCs).

As in Experiment 1, most incorrect completions involved omitting the middle VP:

The fact that the professor who the diplomat ...

... had lunch with was true.

The manuscript which the writer who ...

... is well-known wrote was excellent.

... hangs out at Denny's sent in, was rejected.

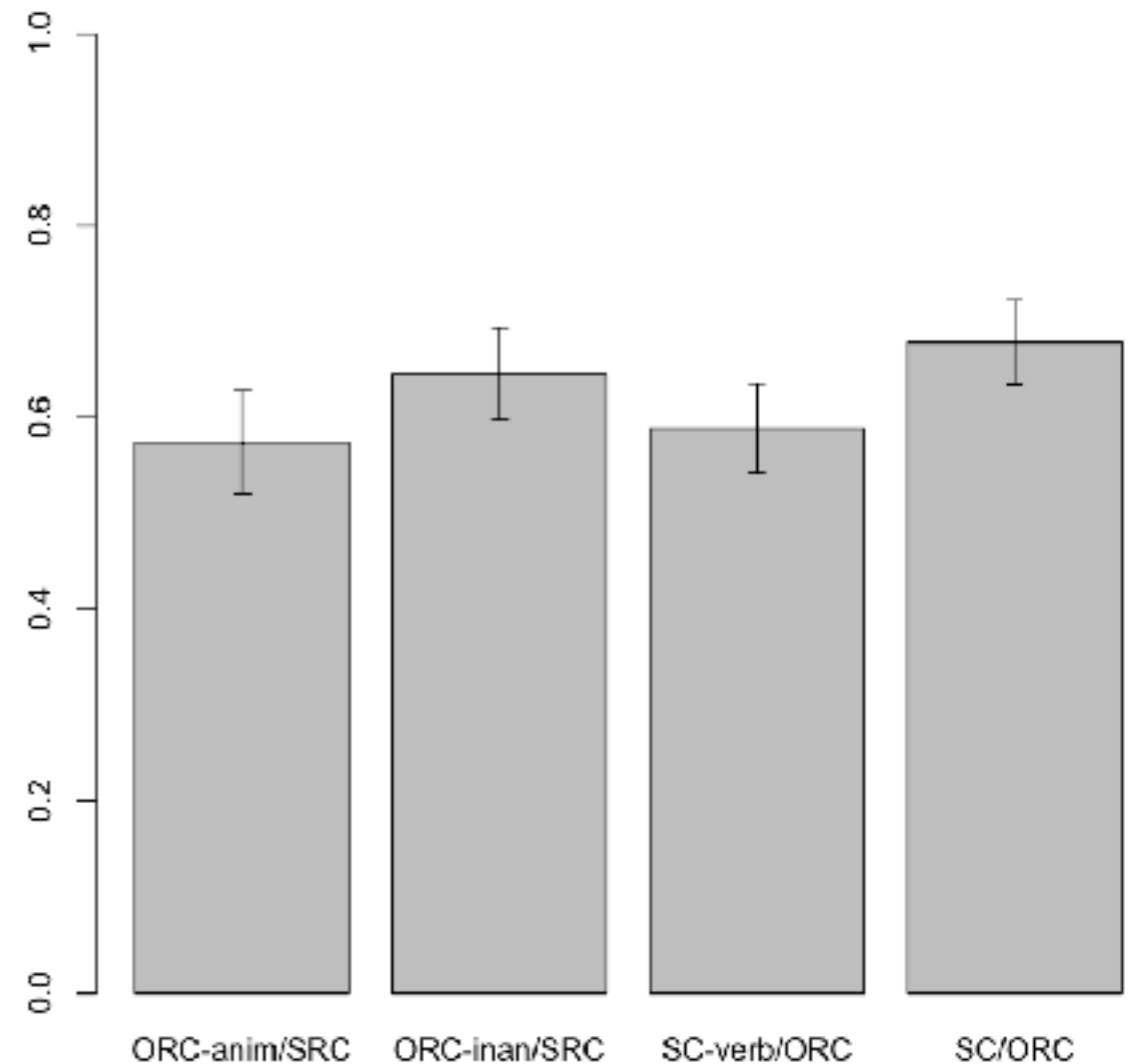
... was sad wrote, won the prize.

... was sick was overdue.

... was poor was rejected.

... enjoyed opium was really very confusing.

... collaborated with the original book's author, was finished ahead of schedule.



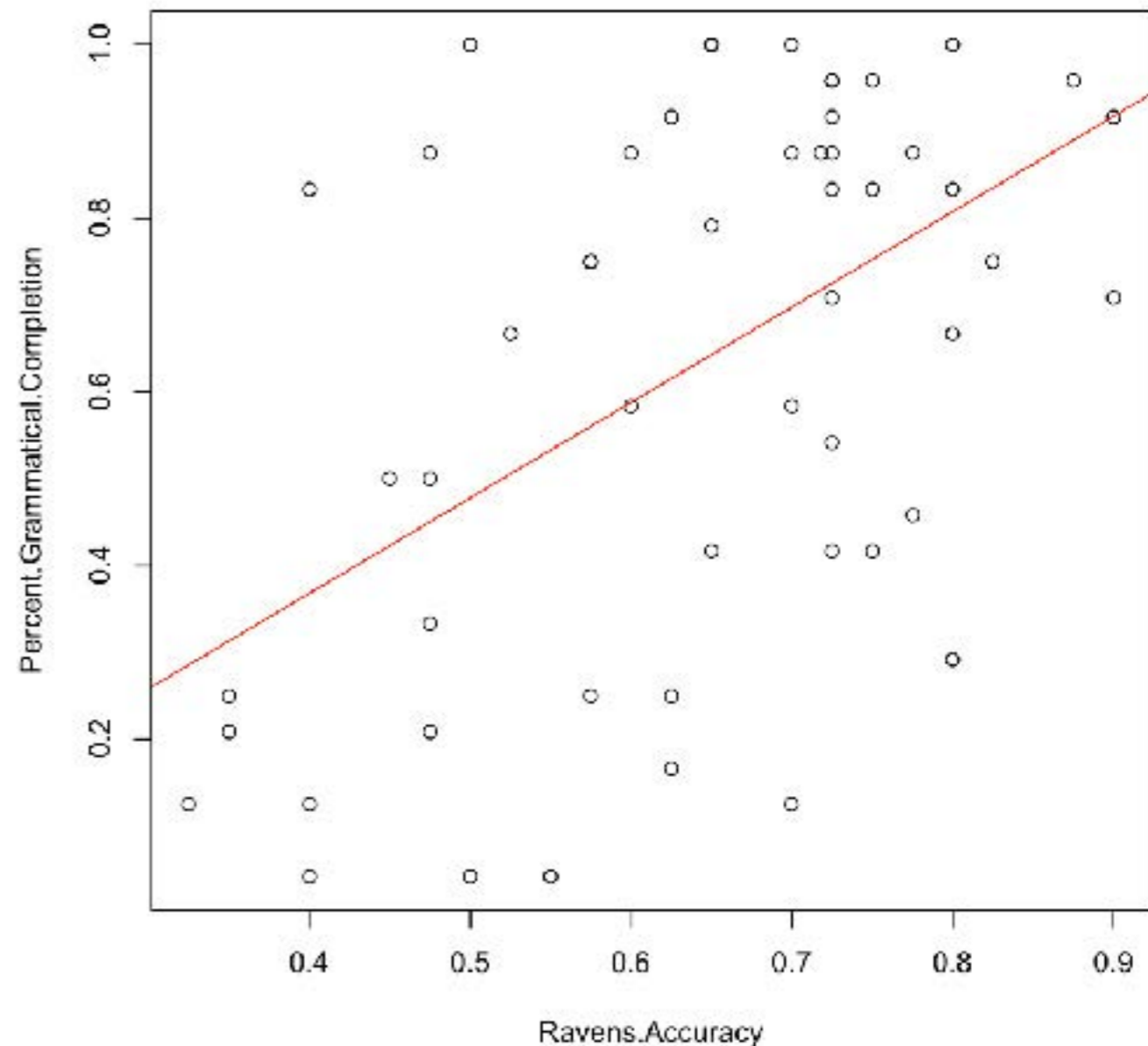
Experiment 2 results:

Correlation between completions and Ravens

Correlation between proportion of grammatical completions and Ravens accuracy:

$n=55$; $r=.505$; $p<.001$

Significant effect of RAPM score predicting the completion score ($\beta=5.16$; $t=3.31$; $p=.0017$), but no reliable effect of work time ($\beta=-0.17$; $t=-0.68$; $p=.50$) nor average filler length ($\beta=0.30$; $t=1.31$; $p=.20$).



Gibson, Fedorenko & Mahowald (2014)

Conclusions

- Performance on a non-linguistic demanding task (a task assessing general fluid intelligence) explains variance in people's ability to grammatically complete syntactically complex materials.
- In some cases **language draws on highly domain-general resources** (contra claims by Caplan & Waters, 1999, that syntactic working memory is language-specific).

Vasishth et al. (2010): Replication & extension to self-paced reading & German

Vasishth et al. (2010): German SPR:

All 3 VPs: *The carpenter who the craftsman that the peasant carried hurt supervised the apprentice.*

Missing VP2: *The carpenter who the craftsman that the peasant carried [] supervised the apprentice.*

German readers have better comprehension the grammatical version, with all 3 VPs!

Why the difference?

Frank et al. (2015): a neural network model of syntactic predictions that gets the German results when trained with German syntax, and the English results when trained with English syntax.

Vasishth et al. (2010): Replication & extension to self-paced reading & German

Why might German be different?

Futrell & Levy (2017): **head-final** embedded clauses are much more common in German

Futrell & Levy (2017): Information Locality

- **Information locality** generalizes dependency locality effects and predicts that words that predict each other (have high **mutual information**) should be close.
- Derived from a general theory of language understanding as rational inference with limited memory resources.

Mutual information between two events: high when they co-occur frequently

Futrell & Levy (2017):
Length-Dependent Noisy Context Surprisal

noisy context

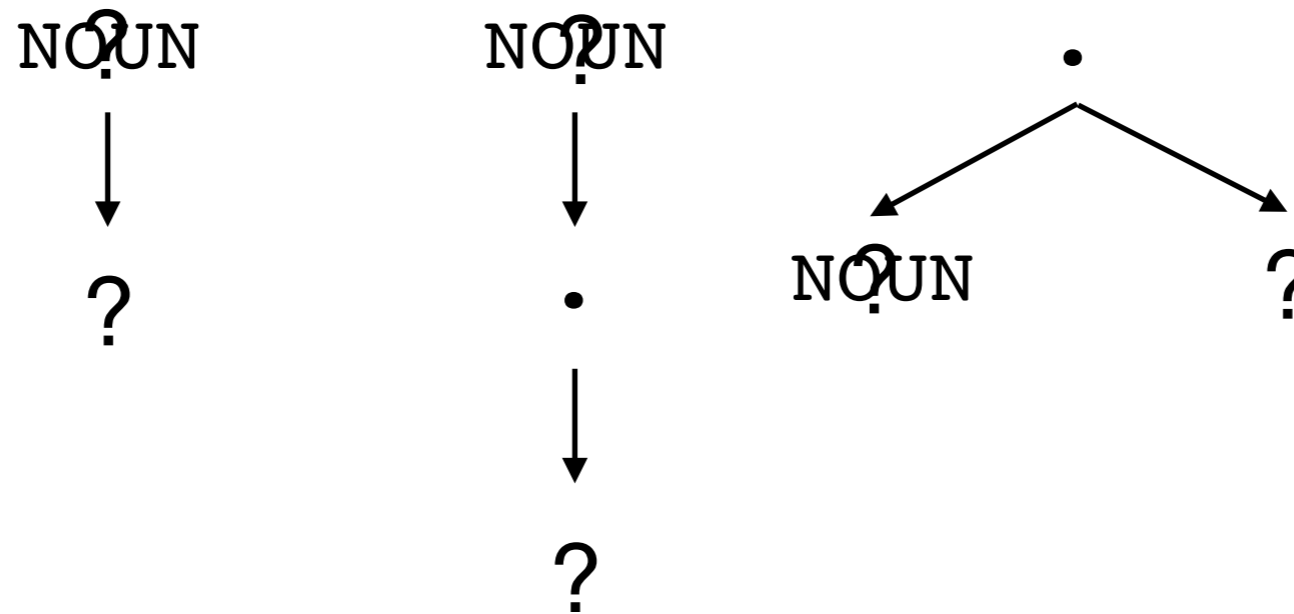


- Suppose we have an **increasing noise rate** the longer a word has been in memory.
- When "threw" is far from "out", then it is less likely to reduce the surprisal of "out": more likely to be affected by noise.
 - When it's closer, it's more likely to be available.
- Noisy-context surprisal increases when **words that predict each other are far apart**.
- We call this **information locality**.

Information Locality

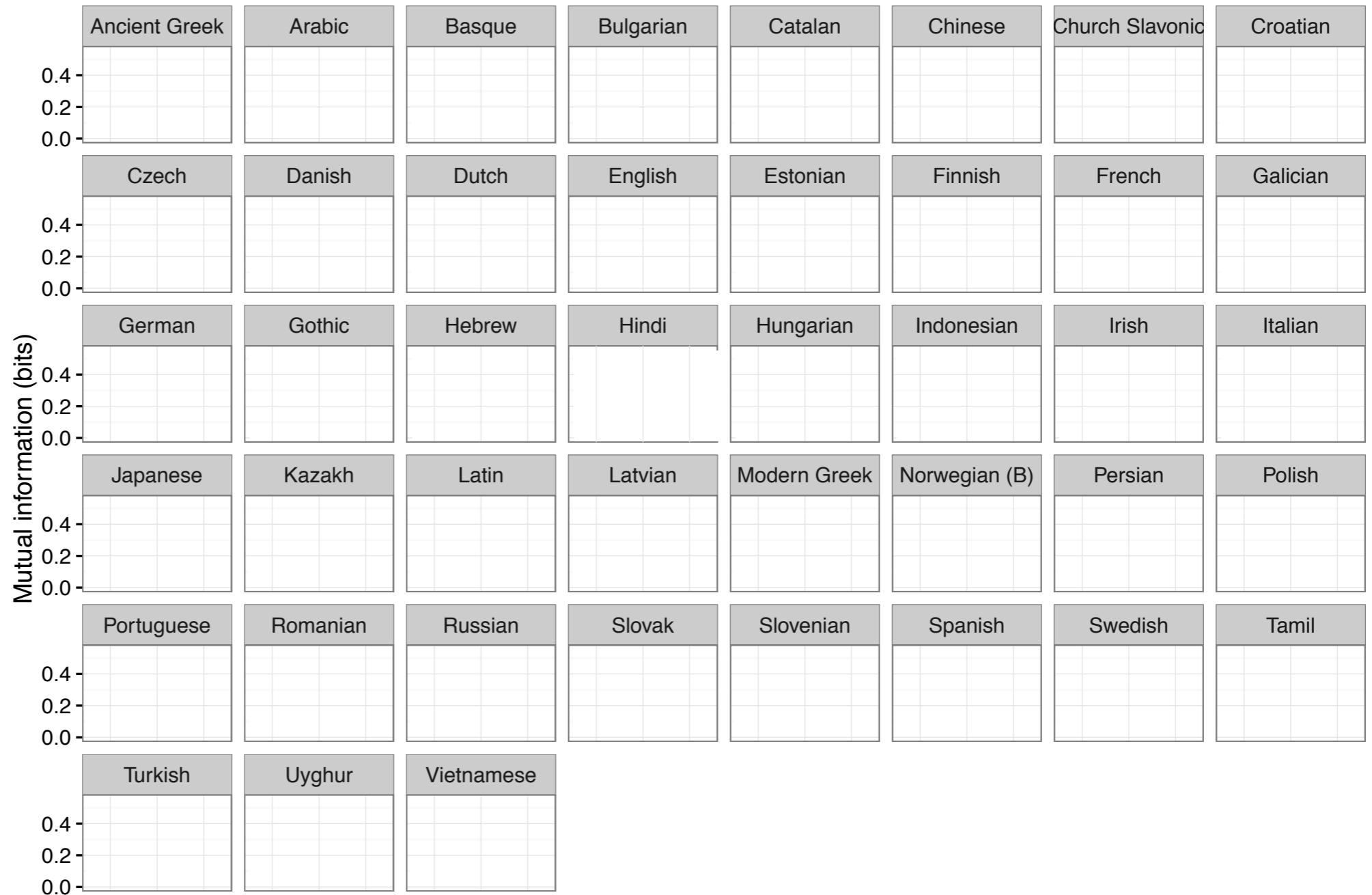
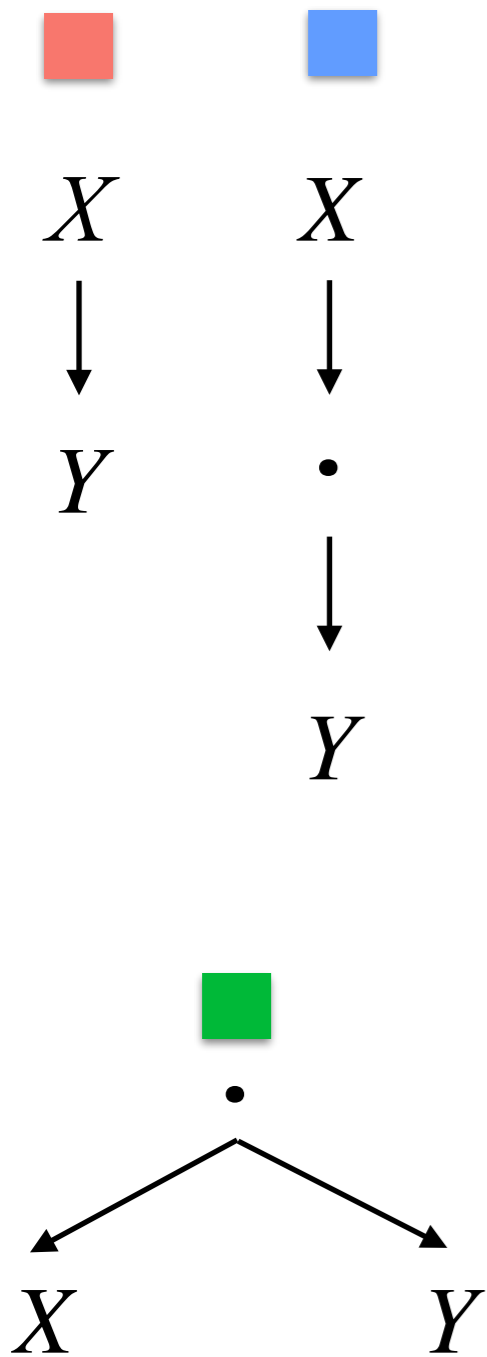
- **Information locality:** predicts processing difficulty when words that predict each other (have high mutual information) are far apart.
- How does this relate to **dependency locality**?
- Hypothesis: **Words in syntactic dependencies have high mutual information.**
 - If this is true, then we can see dependency locality effects as a subset of information locality effects.
- This is true in dependency corpora.

Do Dependencies Have High Mutual Information?



- Futrell & Levy calculated mutual information values over part-of-speech tags for pairs of words in the UD corpora.

Do Dependencies Have High Mutual Information?



Structural Forgetting

1. *The **apartment** that the **maid** who the **cleaning service** **sent over** **was well-decorated**. 👍

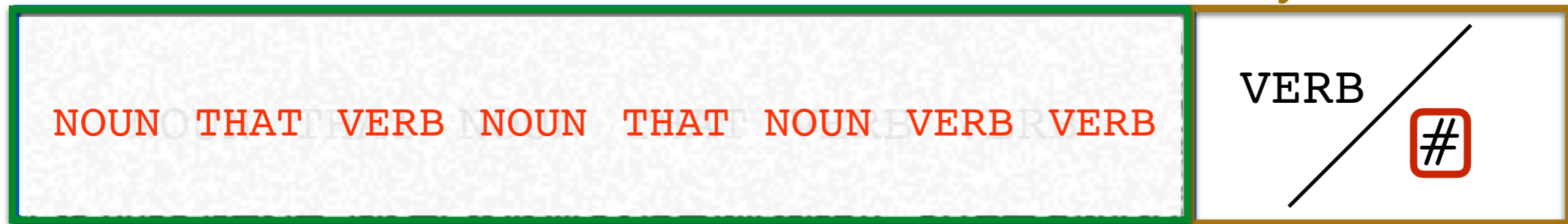
2. The **apartment** that the **maid** who the **cleaning service** **sent over** **cleaned** **was well-decorated**. 👎

- Object-extracted RCs are uncommon in English (Roland et al., 2007).
- English: the maid [that cleaned the apartment] **80%**
the apartment [that the maid cleaned] **20%**

Noisy-Context Surprisal Account of Structural Forgetting English

noisy context

key word



- Correct for noise based on prior about the language.
- Low probability for verb-final RCs in English,
 - so likely to make the wrong prediction.

Structural Forgetting

1. *The **apartment** that the **maid** who the **cleaning service sent over was well-decorated.** 👍

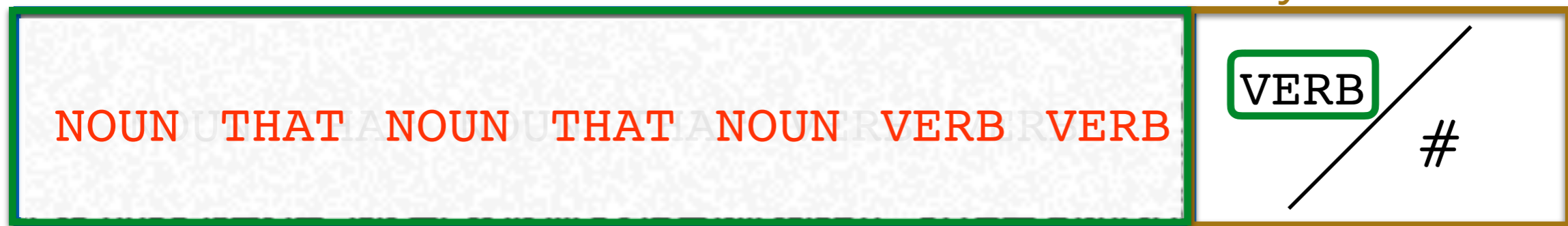
2. The **apartment** that the **maid** who the **cleaning service sent over cleaned was well-decorated.** 👎

- These contexts are more common in German than in English
 - German: das Dienstmädchen, [das die Wohnung reinigte]
die Wohnung, [die das Dienstmädchen reinigte]
- **All RCs are head-final in German**

Noisy-Context Surprisal Account of Structural Forgetting German

noisy context

key word



- Correct noise based on prior about the language in German
- Higher probability for verb-final RCs in German,
 - so more likely to make the right prediction.

Noisy-Context Surprisal Account of Structural Forgetting

- Probability that a context is remembered depends on its prior probability.
 - Noisy-context surprisal *explains* the behavior of the RNN in Frank et al. (2016): the RNN is using a lossily compressed / noisy representation of context.
- Same model that derives dependency locality effects also derives language-specific memory effects.
- The model has an explicit grammar (competence), but cannot apply it correctly (performance).

7 More Possible Projects

9. Information theory and sentence production (Gibson & Thomas, 1999; Futrell & Levy, 2017)

Materials like (1) are so difficult to understand that people often accept an ungrammatical version with the second verb phrase (VP2) omitted (2) (Gibson & Thomas, 1999):

- (1) The apartment that the maid who the service had sent was cleaning was well decorated.
(2) missing VP2: The apartment that the maid who the service had sent was well decorated.

Replication: replicate the Gibson & Thomas (1999) results using acceptability judgements.
Extension: extend using the ideas from Futrell & Levy (2017).

10. Domain specificity / generality: Individual differences measures of completing complex sentences and IQ (Gibson, Fedorenko & Mahowald, 2014)

A long-standing question in the language literature is whether or not there is a domain-specific pool of resources just for comprehending and producing language. Gibson, Fedorenko & Mahowald (2014) presented a novel paradigm for assessing linguistic working memory based on people's ability to grammatically complete complex sentences. They showed that this measure correlates with a measure of non-verbal IQ, suggesting that there is no domain-specific pool of resources (or minimally, that there is some overlap in the pools of resources).

7 More Possible Projects

11. Discourse Structure and Constraints on English Relative Clauses (Ambridge & Goldberg, 2008)

The acceptability of certain complex relative clauses in English depends on the verbs present. For example, the question "What did John say that Mary stole from the store?" seems well-formed, but the questions "What did John realize that Mary stole from the store?" and "What did John mumble that Mary stole from the store?" are typically rated as less acceptable.

Ambridge & Goldberg (2008) explore the idea that the unacceptability of the latter two sentences arises because of constraints on discourse structure. In the acceptable sentence, the verb "say" makes it seem that the important information in the sentence is contained in the following clause. In contrast, in the unacceptable sentences, the verbs "realized" and "mumbled" make the following clause seem like background information. They perform ratings studies showing that the degree to which verbs imply that the following clause is background information strongly predicts the acceptability of relative clauses formed from those clauses.

While the results of this study are strong, they have not been extended to other grammatical phenomena in English or other languages.

7 More Possible Projects

12. Individual differences in quantifier scope

The sentence “Every student didn’t pass the test” is ambiguous between a meaning where none of the students passed the test, and one in which some but not all of the students passed the test. While one might think that the context could disambiguate the meaning, it appears that some people have trouble overcoming a (possibly primed) meaning bias.

Context biased to “none”: *Six little-known singers were each getting ready to release their second albums. Their record labels expected that none of the unknown singers’ new albums would sell one million copies. As expected, every singer’s album didn’t sell one million copies.*

Context biased to “some but not all”: *Three famous singers and three little-known singers were each getting ready to release their second albums. Their record labels expected that all of the famous singers’ new albums would sell one million copies, but that none of unknown singers’ new albums would sell one million copies. As expected, every singer’s album didn’t sell one million copies.*

How many singers' albums sold one million copies? 1) All 2) Some 3)None

Replicate the lack of sensitivity to context that some participants obtain. Extend this in some interesting way, possibly to try to figure out what kind of priming this is, and how you could make it disappear.

7 More Possible Projects

13. Knowledge and Implicature: Context and the interpretation of quantifiers

Goodman & Stuhlmuller (2013) observed that how one interprets “some” depends on the context. In a context where Laura has checked all vs some of the letters to her company, when she says “some” it changes the likely meaning:

Letters to Laura's company almost always have checks inside. Today Laura needs to find out whether 3 of the letters have checks inside. Laura tells you on the phone: I have now looked at 2 (or all 3) of the 3 letters, and given what I saw, I can tell you that some of today's letters have checks inside.

Estimate (in percent) how likely it is that all 3 of the letters have checks inside.

In a context where Laura has looked at all 3, then “some” usually means some but not all. But if she has only looked at 2 of the 3 letters, then this means some and possibly all.

Replicate this, and extend it in some interesting way.

7 More Possible Projects

14. Sedivy (2003) color naming

Sedivy (2003) shows that speakers use color adjectives only when they are typically informative, and that listeners reason on-line about speakers' intentions in using color adjectives by assuming that speakers are trying to be informative. Her paper shows that people use color adjectives to modify nouns referring to objects (such as "the blue cup") when the object can conceivably come in many colors, even when the color adjective is not necessary to disambiguate objects in immediate context. For objects that do not vary freely in color (such as a yellow banana), people do not use color adjectives often. On the comprehension side, in a visual world paradigm, when subjects hear the word "yellow" they are likely to look AWAY from a yellow banana, since they do not think a speaker would use the color adjective for that object. Project: color labeling over M Turk.

15. People are more rational in L2 than in L1.

Keysar et al (2012) observed that using a foreign language reduces decision-making biases. Four experiments show that the framing effect disappears when choices are presented in a foreign tongue. Whereas people were risk averse for gains and risk seeking for losses when choices were presented in their native tongue, they were not influenced by this framing manipulation in a foreign language. Two additional experiments showed that using a foreign language reduces loss aversion, increasing the acceptance of both hypothetical and real bets with positive expected value.

Costa et al (2014): extension to morality judgments.

9.59J/24.905J Lab in Psycholinguistics
Spring 2017

For information about citing these materials or our Terms of Use, visit: <https://ocw.mit.edu/terms>

.