

Class 6:

Language processing: Noisy channel / Locality

Ted Gibson
9.59J/24.905J

Noisy-channel models of comprehension

Summary:

Evidence for a noise model:

1. People are more likely to infer the plausible alternative if it involves inferring fewer errors.
2. People are more likely to infer the plausible alternative if it is one deletion away compared to one insertion.
3. Increasing the noise increases the reliance on plausibility.

Evidence for priors:

1. Plausibility Prior: Increasing the likelihood of implausible events decreases the reliance on semantics.

(Gibson, Bergen & Piantadosi, 2013, *PNAS*)

Agreement errors: the result of noisy-channel in comprehension? (Bergen & Gibson, 2012)

A classic finding in the sentence production literature (Bock & Miller, 1991) and comprehension (Pearlmutter, Garnsey & Bock, 1999):

Agreement error *asymmetry*:

1. The key to the cabinets was / **were** on the table.

(Many errors for plural local noun)

2. The keys to the cabinet were / ??**was** on the table.

(Very few errors for singular local noun)

Standard explanation: there is a *markedness* difference between singular vs. plural nouns, in memory retrieval / sentence planning.

Stipulation

Agreement errors: the result of noisy-channel in comprehension? (Bergen & Gibson, 2012)

Our claim:

Agreement errors result from **rational misidentification of the preamble.**

The asymmetry between singular and plural head-nouns is explained by 2 factors:

- Deletions are much more likely than insertions (Gibson et al, 2013). *Thus agreement errors will occur more often when the head noun is singular.*
- Prior distribution of NP sequences: the singular-singular is much the most common sequence. *Thus there will be few errors confusing sing-sing as plural-sing.*

Plural-head/Singular-local

The keys to the cabinet...

Given the plural head noun, it is unlikely that the comprehender will infer that the plural-marking was produced by mistake, so unlikely to be pulled to the sing-sing.

Singular-head/Plural-local

The key to the cabinets...

Given the singular head noun, it is possible that the comprehender will think that the producer intended a plural / plural, hence producing an error.

Agreement errors: the result of noisy-channel in comprehension? (Bergen & Gibson, 2012)

Predictions:

1. Misidentification of the sentence preamble will lead to repetition errors in the preamble (*The actors in the commercials* when the true input is *The actor in the commercials*). These should pattern with typical agreement errors.
2. Additional cues to head-number will decrease the agreement error rate. Multiple cues are likely to have been used intentionally.

Agreement errors: the result of noisy-channel in comprehension? (Bergen & Gibson, 2012)

Experiment 1:

Validation of slightly modified methodology.

Task: 1.5 sec visual presentation of preambles, 13 sec to retype and complete sentence.

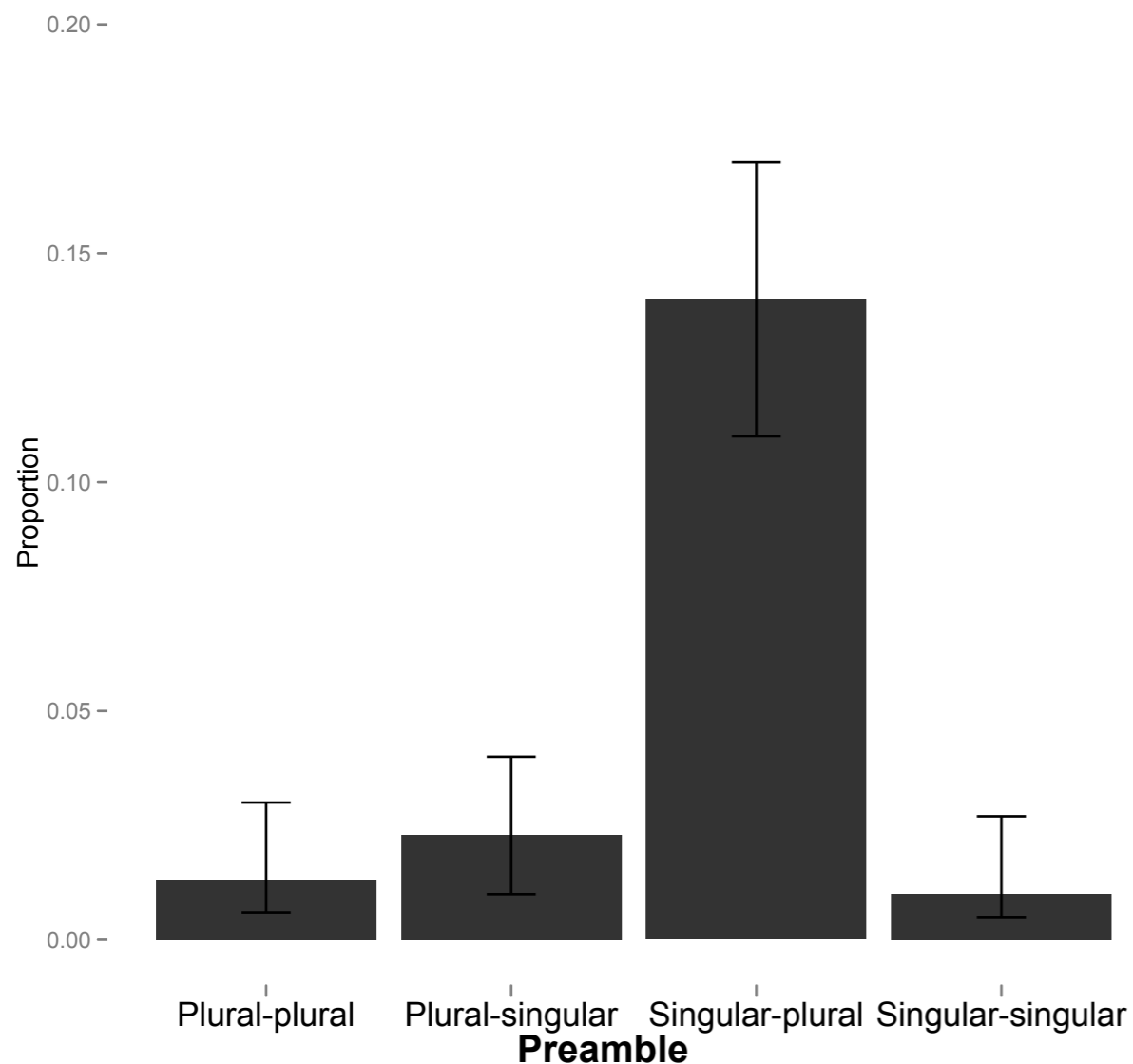
32 items, using materials similar to (Bock & Miller, 1991):

Plural-head/Plural-local	The actors in the commercials
Plural-head/Singular-local	The actors in the commercial
Singular-head/Plural-local	The actor in the commercials
Singular-head/Singular-local	The actor in the commercial

Agreement errors: the result of noisy-channel in comprehension? (Bergen & Gibson, 2012)

Experiment 1: n=80, MTurk

Experiment 1 Agreement Errors



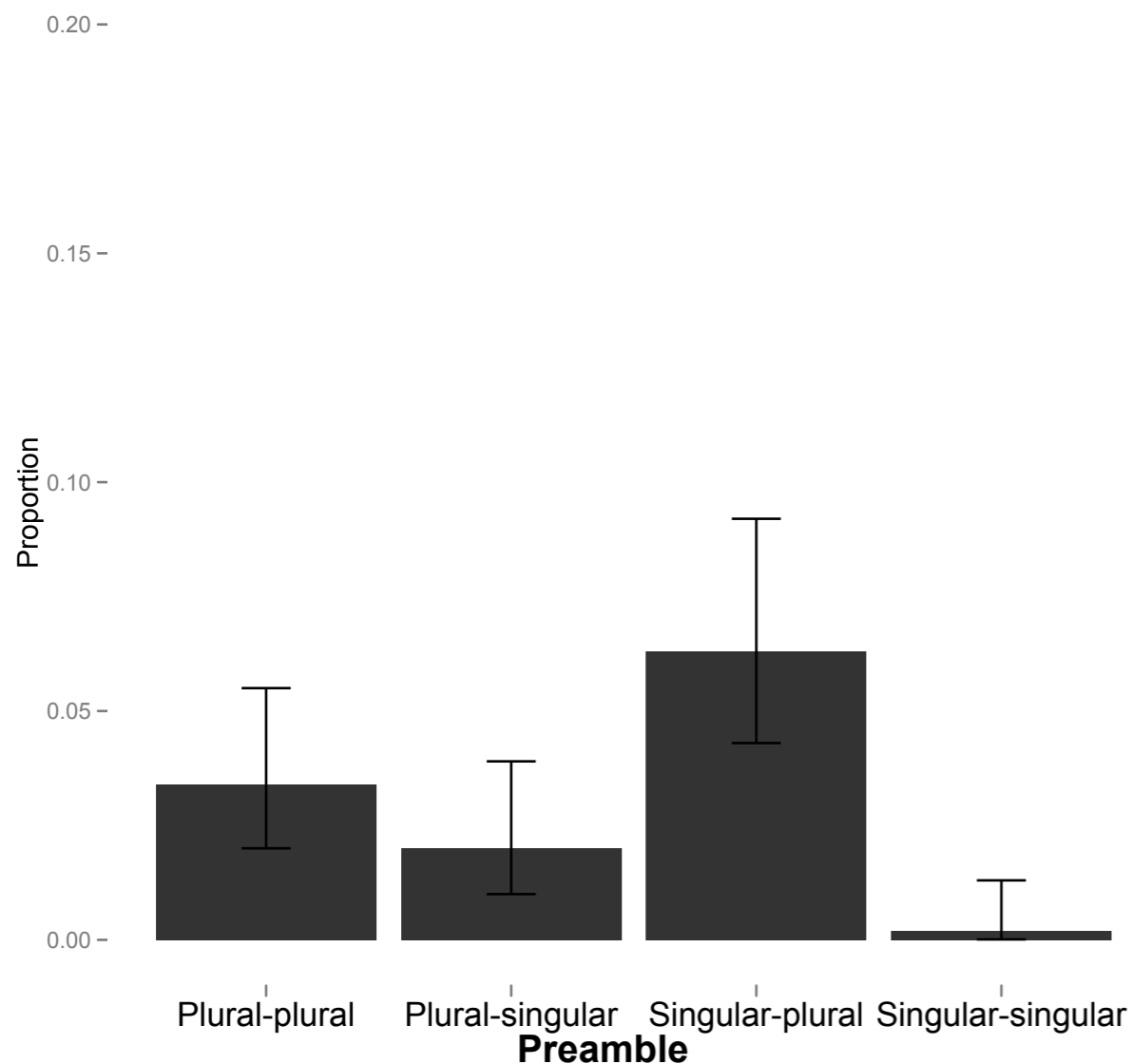
Results:

Replication of two major effects: asymmetry between singular and plural head-nouns, and effect of mismatch between head and local nouns: Reliable interaction

Agreement errors: the result of noisy-channel in comprehension? (Bergen & Gibson, 2012)

Experiment 1: n=80, MTurk

Experiment 1 Repetition Errors

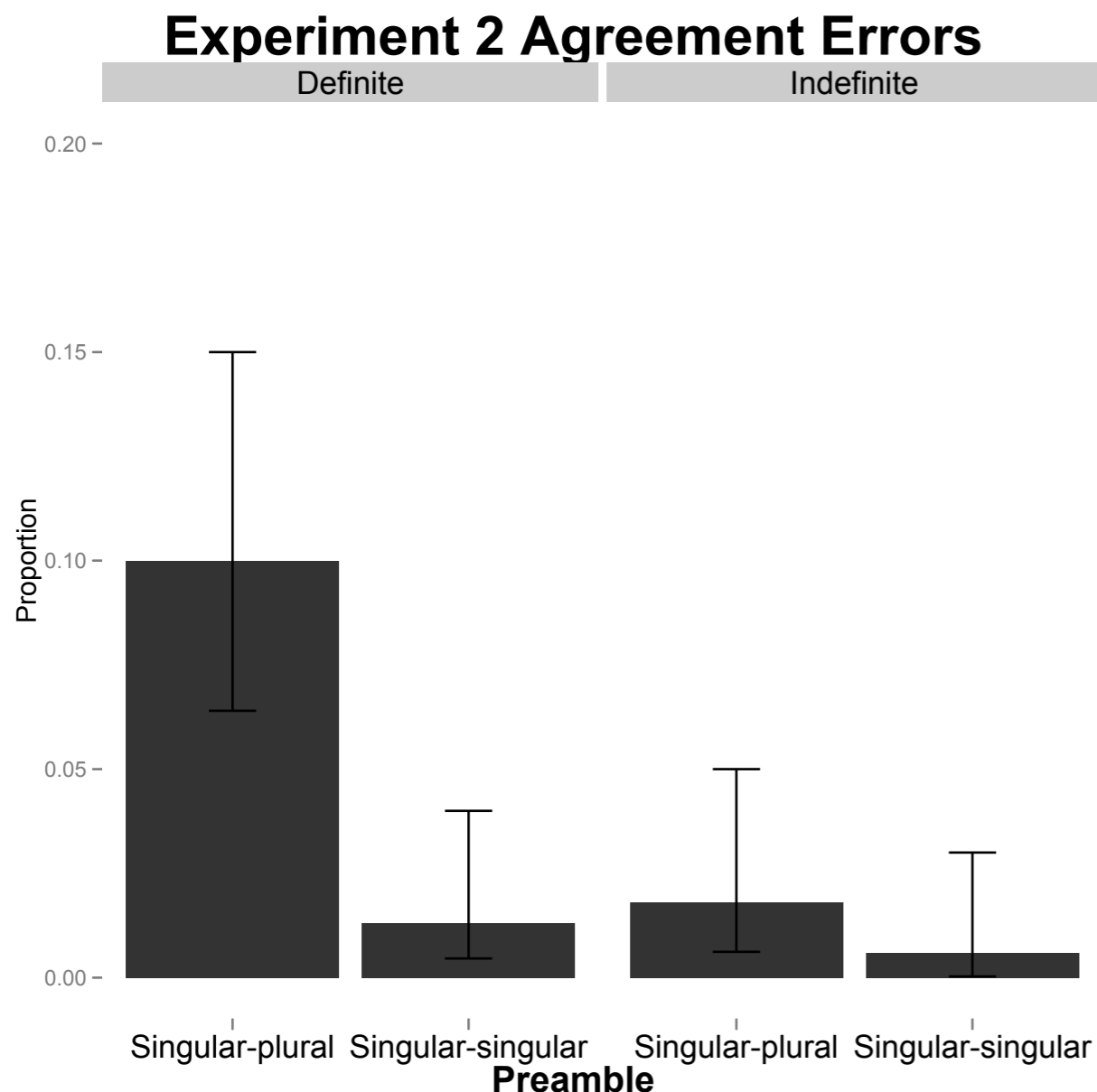


Results:

Repetition errors patterned with agreement errors, occurring most frequently in the singular-plural condition: Reliable interaction

Agreement errors: the result of noisy-channel in comprehension? (Bergen & Gibson, 2012)

Experiment 2: n=80, MTurk



Design: MORE NUMBER CUES

An additional cue to the number marking of the head-noun: an indefinite article (see Hartsuiker et al., 2003, for a similar experiment in Dutch).

Definite/singular-local(plural-local)

The actor in the commercial(s)

Indefinite/singular-local(plural-local)

An actor in the commercial(s)

Prediction: Fewer agreement errors in the indefinite conditions.

Results:

Significant interaction between the article and the number of the local-noun.

Error rates were significantly lower for the singular-plural items in the indefinite condition than in the definite condition.

A noisy-channel explanation of agreement errors (Bergen & Gibson, 2012)

These experiments argue for a ***rational inference*** account (cf. traditional views, i.e., syntactic planning or memory retrieval).

Advantages of the rational inference account:

1. A principled explanation of the sing-plural / plural-sing asymmetry.
2. Repetition errors pattern with typical agreement errors.
3. A unitary explanation for three additional aspects of the data:
 - Additional head-number cues decrease the error rate.
 - Increasing the presentation time decreases the error rate.
 - Increasing the base rate of singular-singular NP-P-NPs decreases the error rate.

Noisy channel: Verb omission errors

Bergen, Levy & Gibson (2012)

Noisy channel account predicts that the phenomenon is not tied to agreement.

Experiment I: Completion of NN/NV preambles

NV biased: NN unambig: *The immigrant fear ...*
 NV unambig: *The immigrant feared ...*

NN biased: NN unambig: *The almond roll ...*
 NV unambig: *The almond rolled ...*

Predicted of noisy-channel approach: deletions more likely than insertions, so people should produce most errors in NV-biased NN

Noisy channel: Verb omission errors

Bergen, Levy & Gibson (2012)

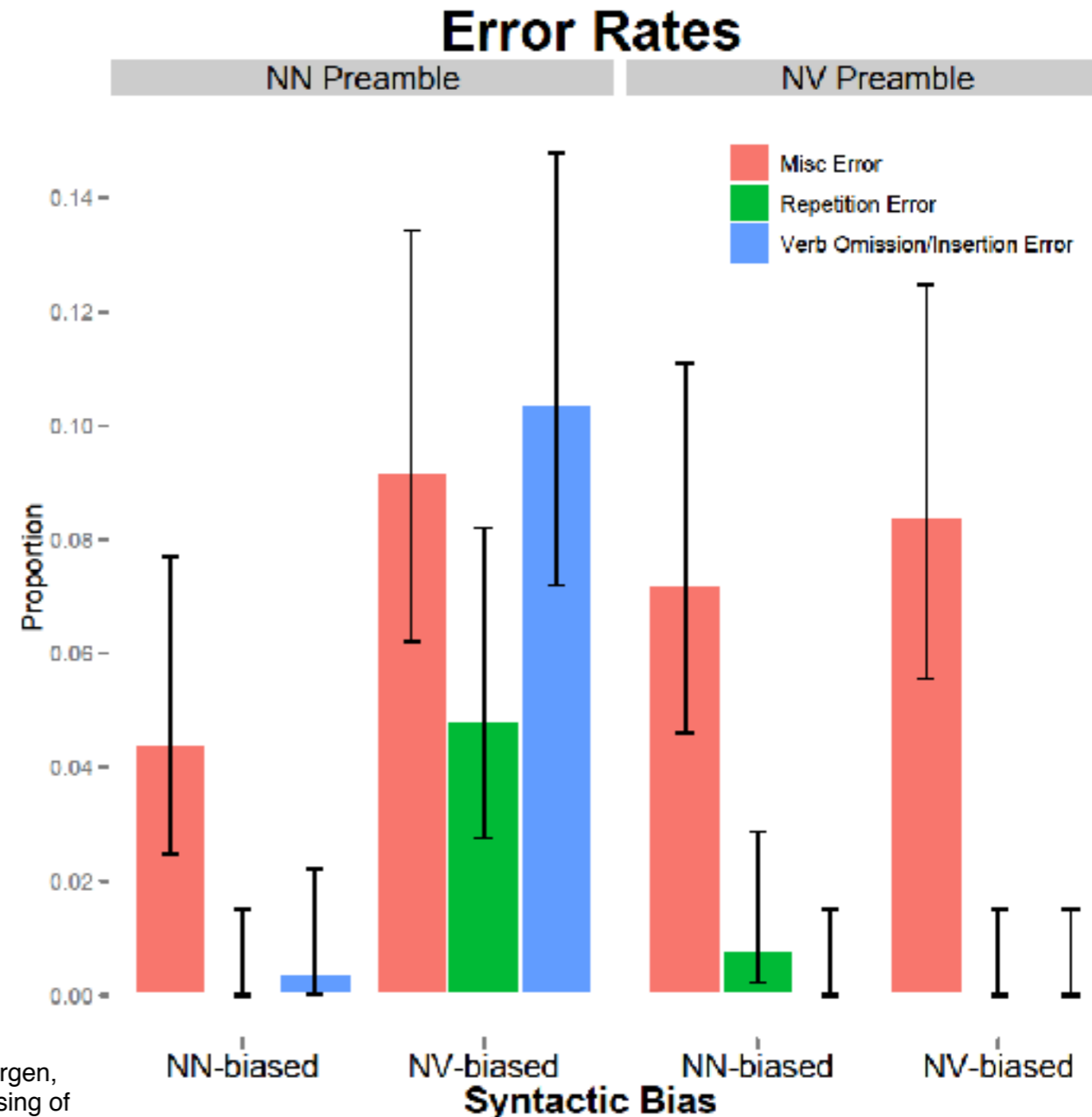
Experiment I

NV biased, NN unambig: *The immigrant fear ...*
NV biased, NV unambig: *The immigrant feared ...*
NN biased, NN unambig: *The almond roll ...*
NN biased, NV unambig: *The almond rolled ...*

As predicted people produce most errors in NV-biased NN:

The immigrant fear ... being deported. (Infer deletion of “s / ed”): Verb omission error

Not: *The almond rolled ... was tasty (Infer insertion: very unlikely)*



Noisy channel: Verb omission errors

Bergen, Levy & Gibson (2012)

Prediction: people will adopt incorrect syntactic analysis if there exist similar phrases that could have easily generated them.
People will follow this incorrect interpretation, and be confused later when it turns out to be wrong.

Experiment 2: Self-paced reading

NN / dense: The intern chauffeur for the governor **hoped for** more interesting work.

NV / dense: The intern chauffeured for the governor but **hoped for** more interesting work.

NN / sparse: The inexperienced chauffeur for the governor **hoped for** more interesting work.

NV / sparse: Some interns chauffeured for the governor but **hoped for** more interesting work.

Noisy channel: Verb omission errors

Bergen, Levy & Gibson (2012)

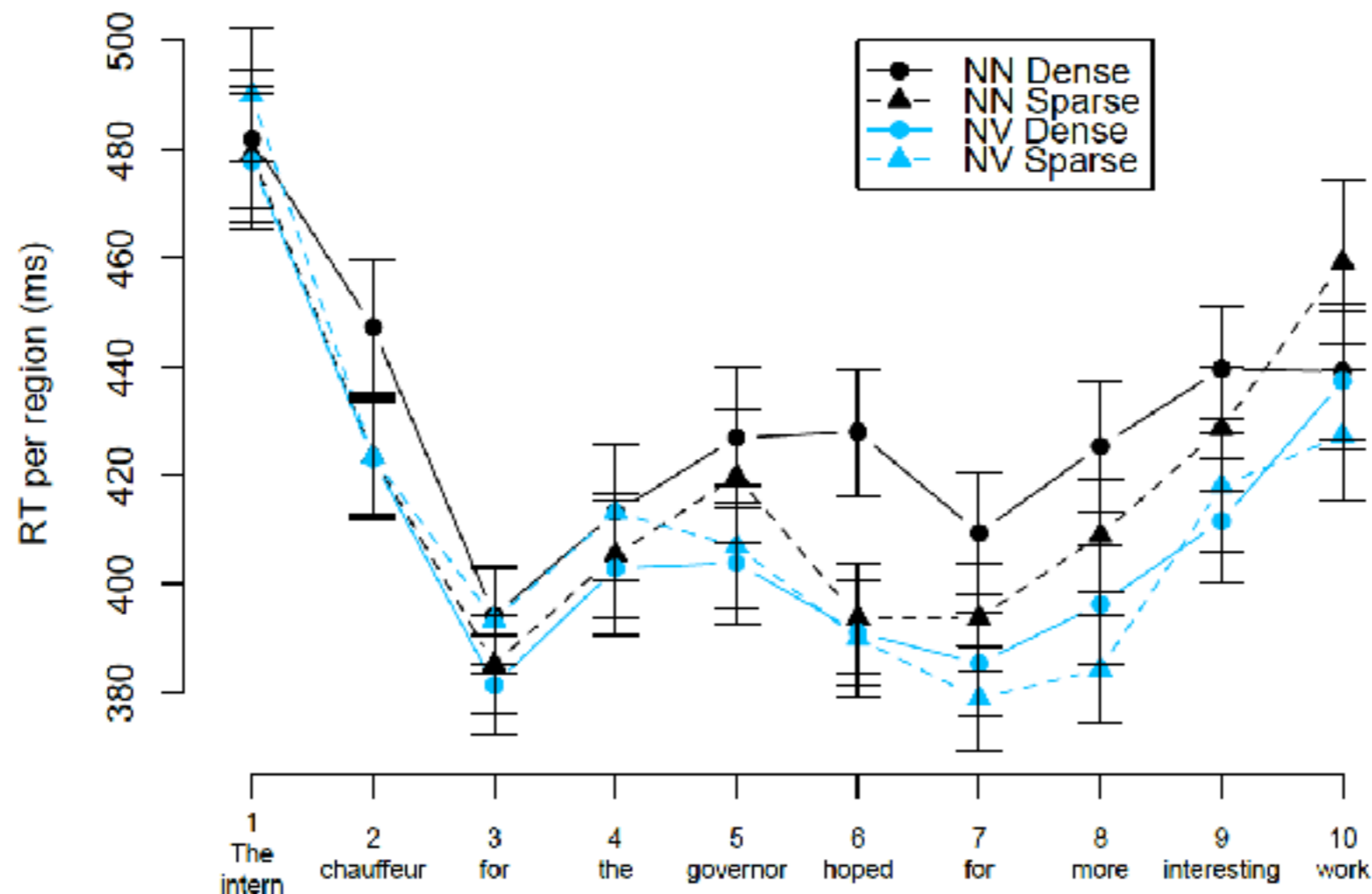
Experiment 2: Self-paced reading

NN / dense: The intern chauffeur for the governor **hoped for** more interesting work.

NV / dense: The intern chauffeured for the governor but **hoped for** more interesting work.

NN / sparse: The inexperienced chauffeur for the governor **hoped for** more interesting work.

NV / sparse: Some interns chauffeured for the governor but **hoped for** more interesting work.



Courtesy of Proceedings of the Cognitive Science Society. License CC BY-NC. Source: Bergen, Leon, Roger Levy, and Edward Gibson. "Verb omission errors: Evidence of rational processing of noisy language inputs." In Proceedings of the Cognitive Science Society, vol. 34, no. 34. 2012.

8 Possible Projects (9 more to come)

1. The potential context sensitivity of acceptability judgments.

How are acceptability judgments affected by the context? that is, what happens when the same materials are embedded in different sets of filler materials? Are the results the same? or does the ease / difficulty / similarity of the filler materials affect the judgment task in important ways?

Replication: a replication of some standard results from the literature. E.g., island effects in extractions; nested complexity; agreement.

2. Acceptability judgments in a language other than English.

This project would take Mahowald et al (2016) and Sprouse et al (2013) as a baseline, and evaluate some judgments from the syntax literature in another language, like the work of Linzen & Oseki (2015). In order to do this project, you would need to be a native speaker of this language.

8 Possible Projects

3. Information theory: word length (Mahowald et al. 2013)

Mahowald et al. 2013 compared meaning-matched word pairs like chimp/chimpanzee and found that a more supportive sentence context is more predictive of the shorter form. This result holds in both a corpus analysis using Google n-grams and a behavioral experiment on Turk in which people were asked to choose either the long or short form of a word. This is consistent with the findings of Piantadosi et al. (2011), who found that, consistent with predictions from information theory, surprisal in context was a better predictor of word length than frequency.

4. Information theory: optional elements in the syntax (Jaeger, 2010)

Jaeger (2010) did corpus analyses, showing that people tend to produce the optional complementizer "that" in environments with high surprisal (e.g., following a verb like "saw", which often takes an NP complement but rarely takes an S complement), and they tend to omit "that" in environments with low surprisal (e.g., following a verb like "know", which often takes an S complement but rarely takes an NP complement).

Project: replicate this effect in acceptability ratings over materials where the verb subcategorization frequencies are varied from an S being highly expected, to an S being much less expected.

Uniform Information Density (UID)

Levy & Jaeger (2006); Jaeger (2010)

One property of good codes for communication:
Information that is conveyed per unit time is constant

Communicate at the channel capacity

- if you go over, you are overwhelming processing mechanisms
- if you go under, you are not being efficient

So peaks and dips in entropy (average surprisal)
should be “smoothed out”

Uniform Information Density (UID)

Levy & Jaeger (2006); Jaeger (2010)

UID in phonetics:

People lengthen unpredictable words and shorten predictable ones:

Lieberman: the vowel in “nine” is dependent on the preceding context:

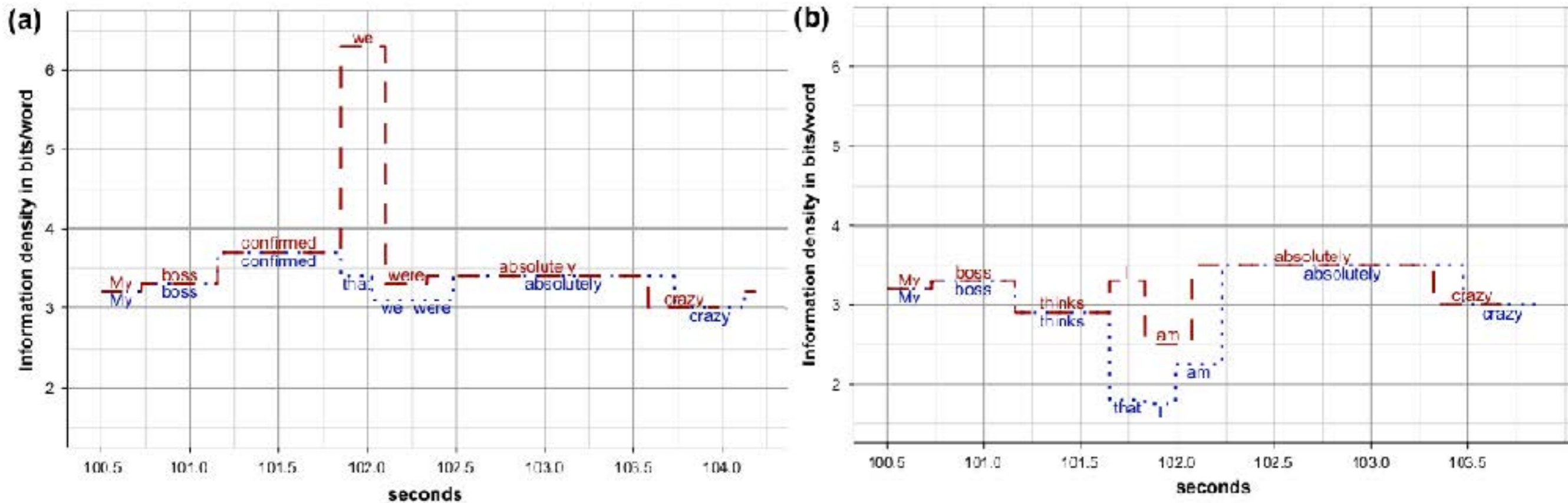
- “A stitch in time saves nine.”
- “The number you will now see is nine.”

UID in Syntax (Jaeger, 2010)

Optionality in syntax: the complementizer “that” is optional in many circumstances in English:

- My boss thinks (that) we were absolutely crazy.
- My boss confirmed (that) we were absolutely crazy.

UID in Syntax (Jaeger, 2010)



- People insert “that” when there would otherwise be a peak in log probability.

8 Possible Projects

5. Mondegreens

We have talked about mishearing song lyrics as noisy-channel inference. Project: Obtain 10-20 mondegreen song examples from popular culture, along with controls from the same songs by the same artists, which don't induce the mondegreen effect. "Replicate" the mondegreen effect by having people write down what they hear for these lyrics.

Attempt to explain the observed effect by doing a language model, such that the "mondegreen" examples are less likely in real language and / or are less plausible in M Turk experiment on the written materials.

6. Information theory: understanding the noise model in a noisy-channel model of sentence comprehensions (Gibson, Bergen & Piantadosi, 2013)

Replication / extension of Gibson et al. and/or: Poppels & Levy (2016); Gibson et al (in press) (accents paper)

One extension of Gibson et al (in press): look at the effect of different accents

An extension of Poppels & Levy: examine features of the noise model

The noisy-channel proposal applied to aphasic comprehension

Old observation: aphasics' comprehension relies more on world knowledge than non-brain-damaged controls. (e.g., Caramazza & Zurif, 1976)

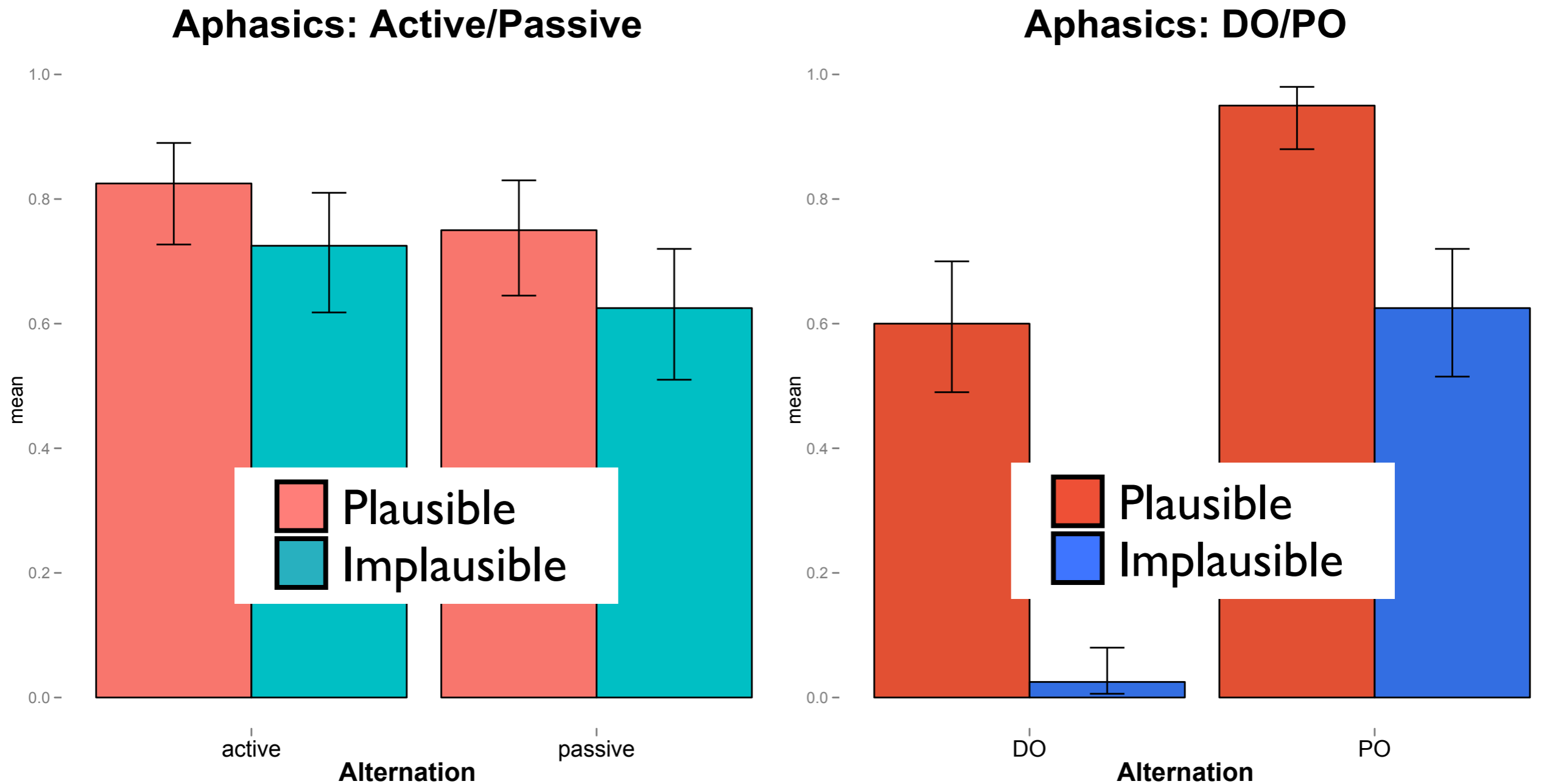
Hypothesis: Aphasics' perception is noisier than that of healthy individuals. In maximizing $P(s_i | s_p)$, aphasics will rely more on their prior distribution $P(s_i)$ over plausibly intended sentences.

(Gibson, Sandberg, Fedorenko, Bergen & Kiran, 2015, *J of Aphasiology*)

Prediction:

Aphasics will rely on semantics more than healthy individuals, in both major-edit (active-passive) and minor-edit alternations (DO-PO).

Results: Active / Passive vs. DO / PO



Aphasics rely more on semantics in minor-edits (DO/PO) than in major-edits (active-passive): $z = 2.93, p < .005$

Similar results for other populations (replicating Gibson, Bergen & Piantadosi, 2013)

Challenges faced by L2 speakers

- L2 speakers are embarrassed by their accents and the errors they make (Gluszek & Dovidio, 2010)
- L2 speakers are perceived to be:
 - less credible (Bourdieu, 1991; Lev-Ari & Keysar, 2010; Livingston et al., 2014)
 - less educated (Fraser & Kelly, 2012)
 - less intelligent (Fuertes, Potere & Ramirez, 2002; Anderson et al., 2007).

L2: One potential advantage

Imagine you are at a cocktail party where you want to make business connections.

Suppose someone asks you about a Marketing Technologist position.

If you have an L2 accent, you could say “Marketing Technologist was hired SEO Consultant.”

With a foreign accent, they may interpret this in the most plausible way. Without a foreign accent, you cannot get away with this uncertainty.

L2: One potential advantage

Arianna Huffington, Smith College commencement address in 2013:

“I moved to New York in 1980 and met Henry Kissinger, who told me not to worry about my accent, because you can never, in American public life, underestimate the advantages of complete and total incomprehensibility.”

Current work: Investigate whether her idea is true: that there is a potential benefit to being misunderstood

Framework for investigating this idea: **rational inference / noisy channel** models of sentence comprehension

L2 vs. L1 Speakers

Current work: Investigate whether the Kissinger / Huffington idea is true: that there is a potential benefit to being misunderstood, using rational inference framework: **Maybe people will make more plausible inferences for L2-accented speakers**

Konieczny, Hemforth & Scheepers (1994): People are likely to infer a higher likelihood *linguistic prior* when interacting with non-native speakers:

One experimenter is was German; another was English with German accent:

- When the German experimenter spoke to self-paced reading participants, they interpreted NPV NP sequences as Object-Verb-Subject, because of the appropriate morphology.
- When the English experimenter spoke to them in accented German, they interpreted NPV NP sequences as Subject-Verb-Object in spite of inappropriate morphology

L2 vs. L1 Speakers: New Experiments

Interpretation of implausible materials, spoken by the same person (each of 2 Speakers), +accent or -accent

3 sets of implausible materials, from the PNAS paper:

1. DO/PO

The mother gave the candle the daughter.

The mother gave the daughter to the candle.

2. Transitive/intransitive

The businessman benefited the tax law.

The tax law benefited from the businessman.

3. Active/Passive

The ball kicked the girl.

The girl was kicked by the ball.

L2 vs. L1 Speakers: New Experiments

Interpretation of implausible materials, spoken by the same person (each of 2 Speakers), +accent or -accent

3 sets of implausible materials, from the PNAS paper:

Fillers: Filler items from Gibson et al., spoken with no accent by the other speaker

Speaker 1: accented / no-accent target items

Speaker 2: no-accent filler items

L2 vs. L1 Speakers: Predictions

If participants think there is more noise in the accented productions, then we predict higher rates of inferences for the DO/PO materials and the transitive/intransitive materials

where Gibson et al. (2013) had seen more inferences when noise was added to their filler materials

but not necessarily for the active/passive materials

*where Gibson et al. (2013) had **not** seen more inferences when noise was added to their filler materials, possibly because there are too many edits to get from the implausible to a plausible version*

L2 vs. L1 Speakers: Methods / Participants

3 experiments, each consisting of four groups of 80 workers, on Amazon.com's Mechanical Turk

Instructions:

This is a set of 80 auditory sentences. Answer the questions immediately following, according to what you think the speaker intended.

10-15 minutes for each participant to complete the task.

Evaluating the comprehensibility of the materials

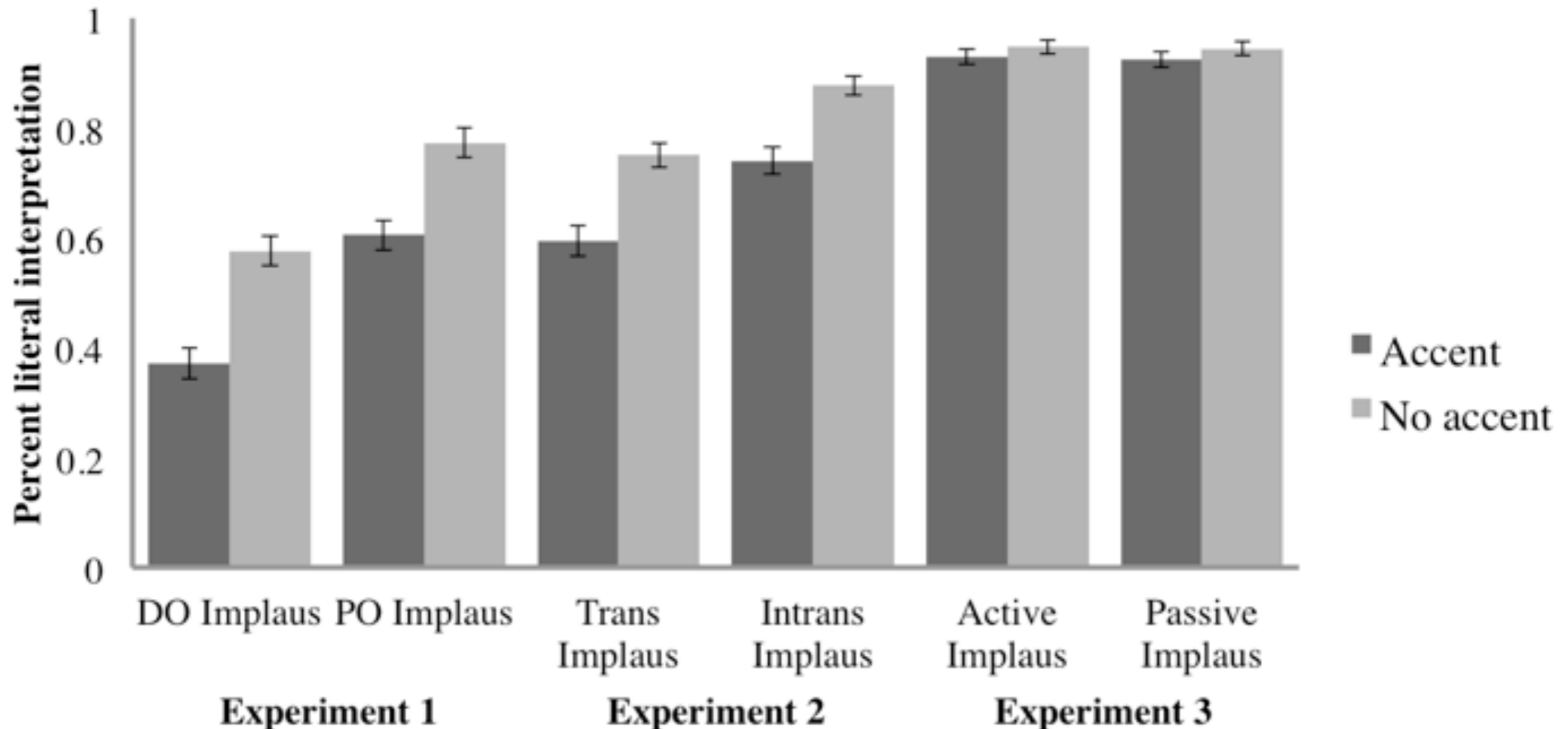
A higher rate of plausibility-based inferences in the +accent condition could result if participants simply cannot discern the words in the utterance.

Norming: An additional 480 Mechanical Turk participants were asked to transcribe what each speaker said (even if it was implausible) across four surveys: 120 participants for the implausible target sentences from each of Speaker 1 and Speaker 2, for each of their accent and no-accent productions.

L2 vs. L1 Speakers: Results

1. DO, PO: ~20% inference effect
2. Transitive, Intransitive: ~15% inference effect
3. Active, Passive: no significant difference.

4. Main effect of speaker: more inferences for S1 vs S2;
5. Replication of Gibson et al. 2013 differences between constructions



Discussion

As suggested by Kissinger / Huffington, there is a potential benefit to being misunderstood: *With a foreign accent, others may be generous in interpreting your speech.*

Explained by rational inference models of sentence comprehension

Does the effect depend on the particular accent?

Does the effect change when all materials are in the accent?

8 Possible Projects

7. Sentence completion errors as rational inference

Bergen & Gibson (2012) proposed a rational inference hypothesis for explaining the asymmetry between "the key to the cabinets are ..." and "the keys to the cabinet is ...".

Bergen, Levy & Gibson (2012) showed a similar effect, which did not involve agreement:

NV biased, NN unambig: *The immigrant fear ...*

NV biased, NV unambig: *The immigrant feared ...*

8. Locality vs. surprisal in online reading.

Replicate subject- vs object-extractions in English (e.g., Gibson, 2000):

subject-extracted relative clause: The reporter who attacked the senator admitted the error.

object-extracted relative clause: The reporter who the senator attacked admitted the error.

Extension: look at genitive extractions, dative extractions

Compare frequency in the input vs. the RTs that are observed.

Language:

Information sources and constraints

Working memory: Longer distance dependencies are harder to process than more local ones

Dependencies between a verb and its post-verbal objects:

Short NP object:

Local Particle: Joe threw **out** the documents.

Non-local Particle: Joe threw the documents **out**.

Long NP object:

Local Particle: Joe threw **out** the very important documents that he brought home.

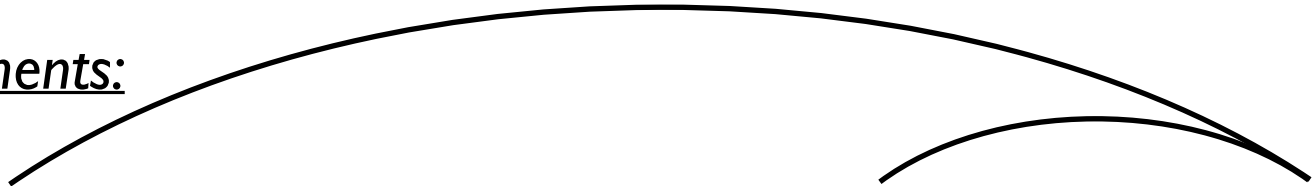
Non-local Particle: Joe threw the very important documents that he brought home **out**.

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; Pearlmuter & 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:

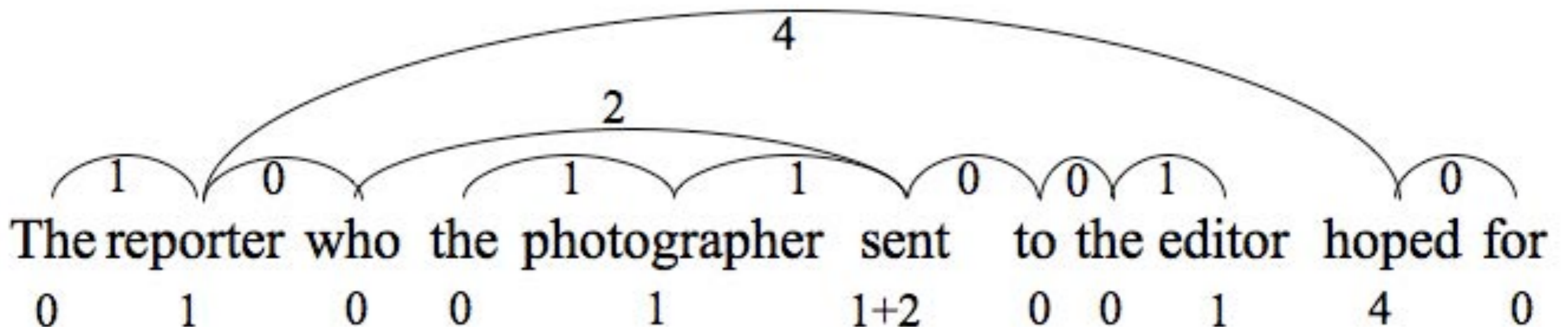
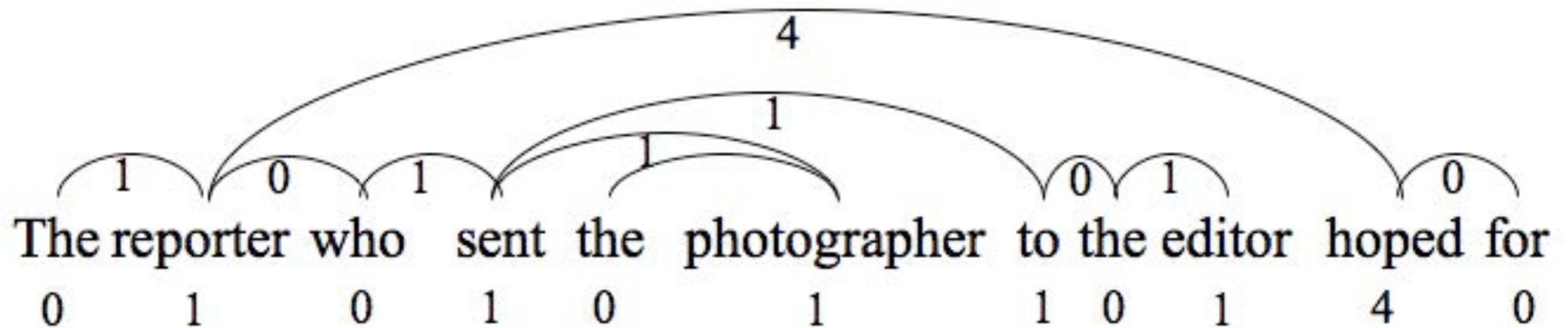
Obasan-wa [bebiisitaa-ga [ani-ga imooto-o ijimeta] to itta] to omotteiru
aunt-top babysitter-nom older-brother-nom younger-sister-acc bullied that said that
thinks

“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

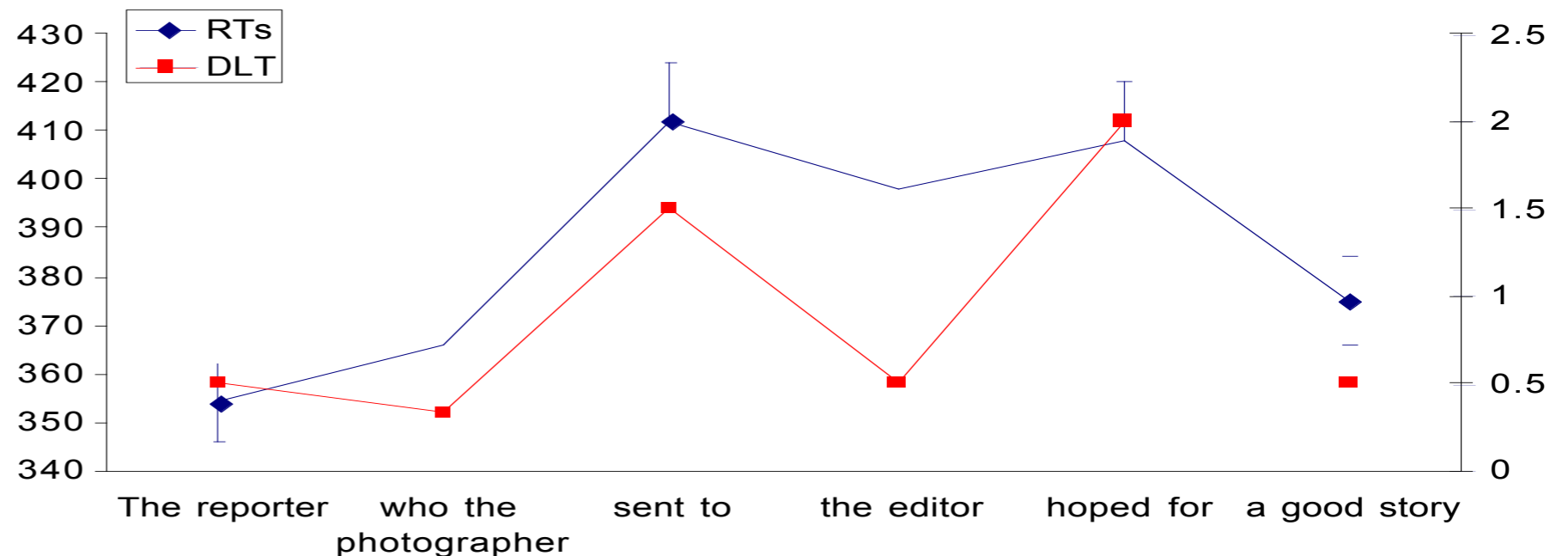
Locality effects in unambiguous structures: Gibson & Grodner (2005) Experiment I

English: Subject- vs. object-extracted relative clauses

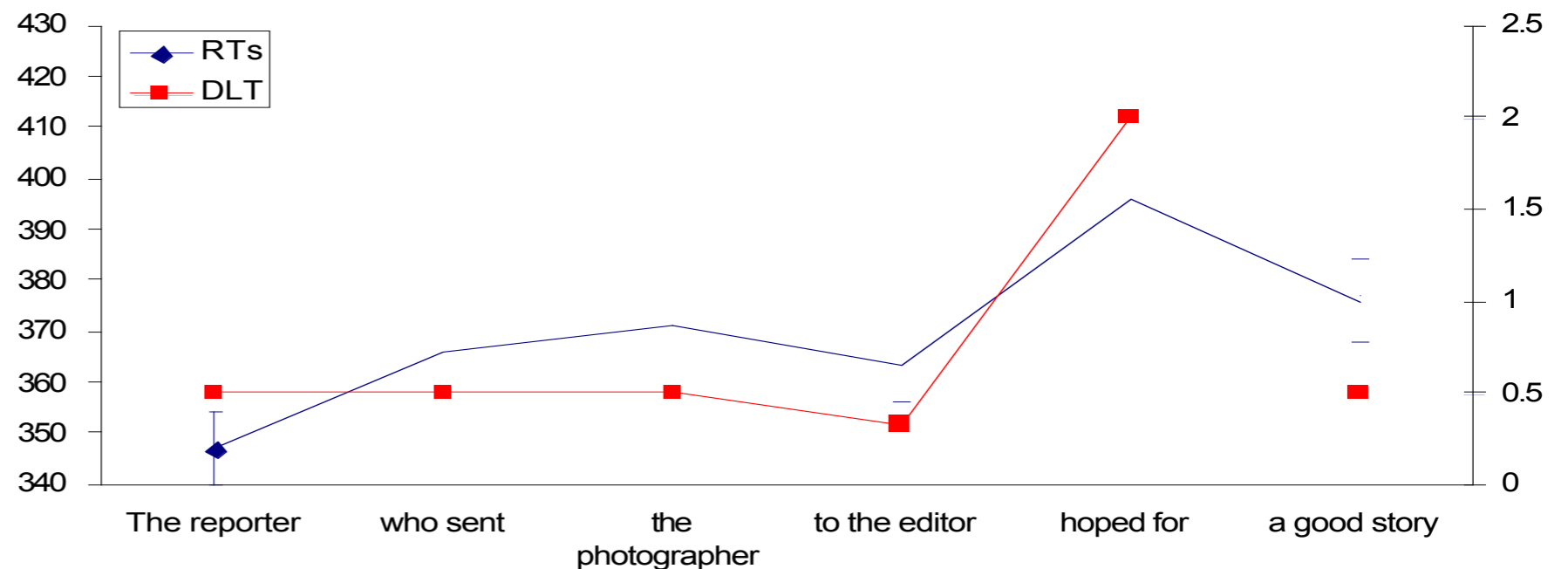


Locality effects in unambiguous structures: Gibson & Grodner (2005) Experiment I

object-extracted relative clauses



subject-extracted relative clauses



Locality effects in unambiguous structures: Gibson & Grodner (2005) Experiment 2

Matrix – Unmodified Subject

The nurse supervised the administrator while ...

0 1 1 0 1 1

Matrix – PP Modified Subject

The nurse from the clinic supervised the administrator while ...

0 1 0 0 1 2 0 1 1

Matrix – RC Modified Subject

The nurse who was from the clinic supervised the administrator while ...

0 1 0 1 0 0 1 3 0 1 1

Embedded – Unmodified Subject

The administrator who the nurse supervised scolded the medic while...

0 1 0 0 1 3 3 0 1 1

Embedded – PP Modified Subject

The administrator who the nurse from the clinic supervised scolded the medic...

0 1 0 0 1 0 0 1 5 4 0 1

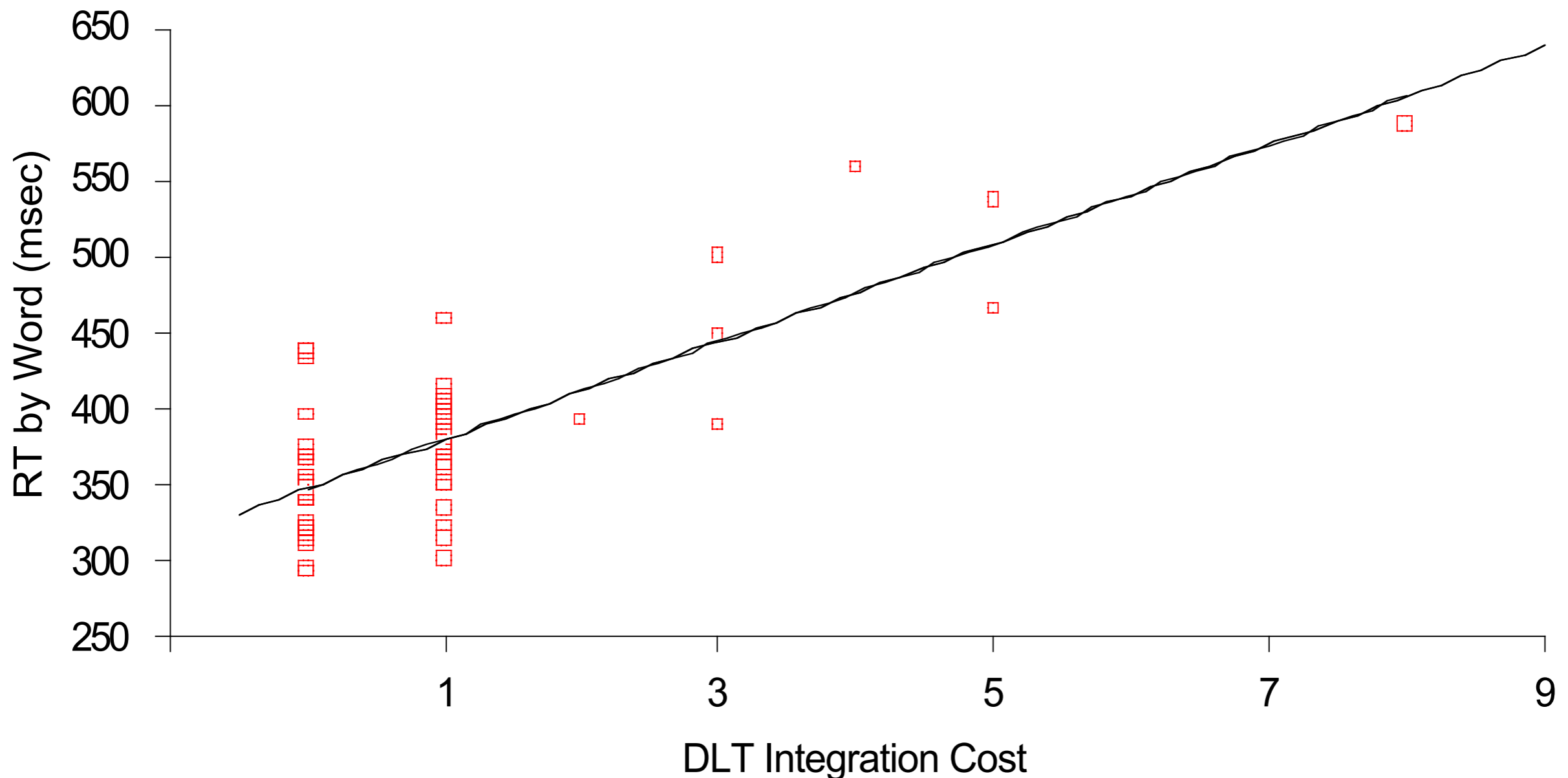
Embedded – RC Modified Subject

The administrator who the nurse who was from the clinic supervised scolded the medic...

0 1 0 0 1 0 1 0 0 1 7 5 0 1

Locality effects in unambiguous structures: Gibson & Grodner (2005) Experiment 2

Experiment 2: DLT vs. RTs by Words



Potential project

Result to replicate: Subject-extractions in Relative clauses (RCs) are easier to process than object-extractions:

Subj-RC: The reporter who attacked the senator admitted the error.

Obj-RC: The reporter who the senator attacked admitted the error.

RTs faster at “attacked” in SRC than in ORC

Two explanations: ORCs are rare, and longer-distance

Extension: evaluation other kinds of extraction in English:

Dative extractions: infrequent, long-distance

The boy who the girl gave the book to admitted the error.

The boy to whom the girl gave the book admitted the error.

Genitive extractions: infrequent, short-distance

The girl whose friend invited the kids to the party was kind.

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)

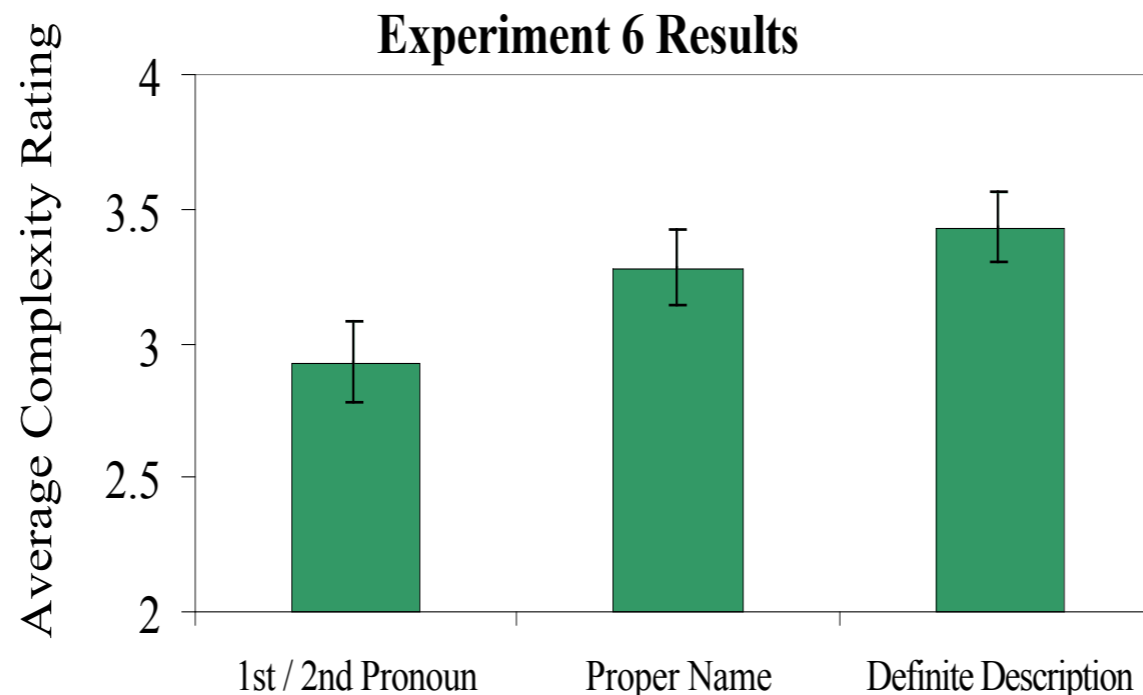
The reporter who everyone that I met trusts said the president won't resign yet.

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.



© Elsevier. All rights reserved. This content is excluded from our Creative Commons license. For more information, see <http://ocw.mit.edu/help/faq-fair-use/>
Source: Warren, Tessa, and Edward Gibson. "The influence of referential processing on sentence complexity." *Cognition* 85, no. 1 (2002): 79-112.

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:

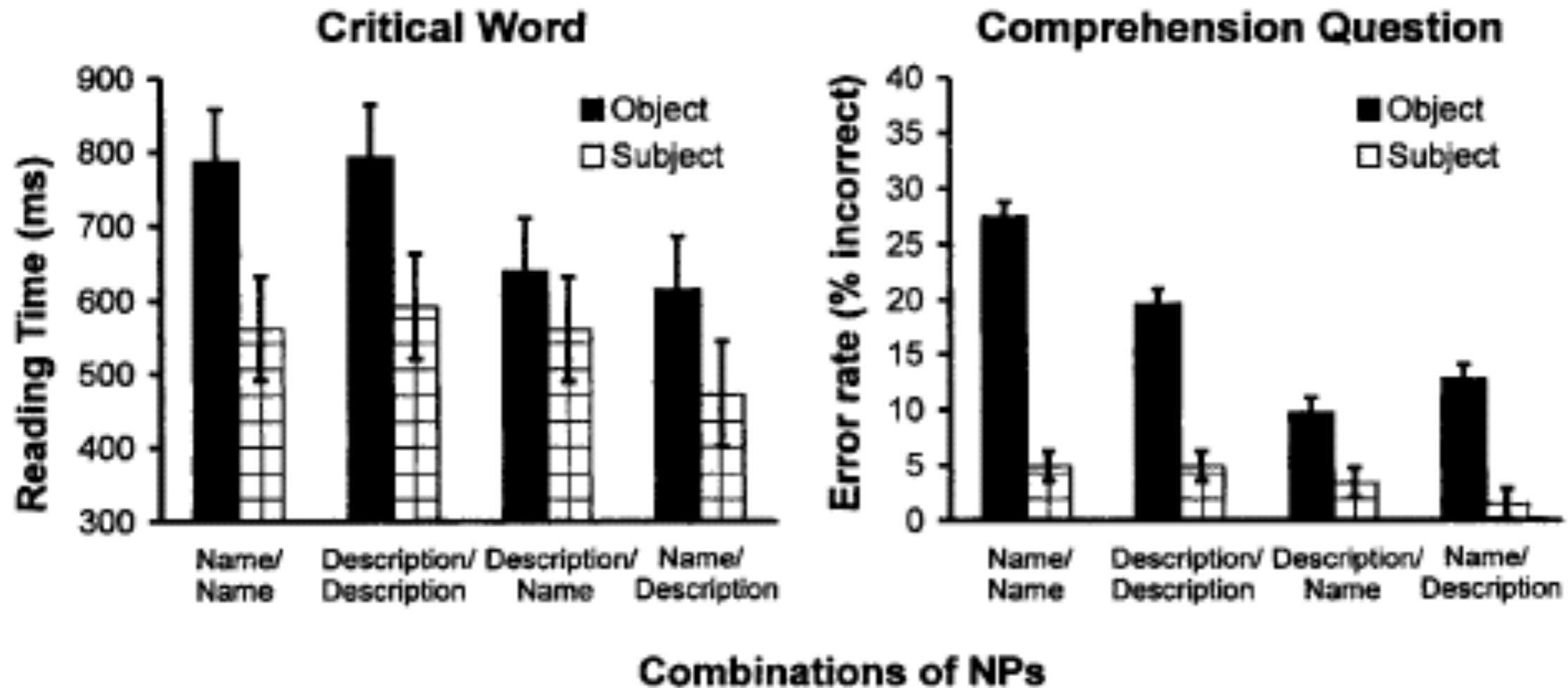
It was (the barber / John) that (the lawyer / Bill) saw in the parking lot.

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.

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>.