Language comprehension

Lecture 5: Experience / frequency and ambiguity resolution; the serial / parallel issue

9.591; 24.945

October 18, 2004

Ted Gibson

9.591 Course so far

- Lecture 1: Experimental methods; Informational constraints affecting sentence comprehension: Lexical frequency, plausibility, context and syntax; Modularity in sentence comprehension.

- Lecture 2: Resources and sentence complexity. The complexity of unambiguous sentences.

- Lecture 3: Working memory and sentence comprehension.

- Lecture 4: Ambiguity resolution: Resources; structural frequencies.

- **Lecture 5: Experience / frequency and ambiguity resolution; the serial / parallel issue.**

Sentence processing: Recap thus far

- Multiple factors are involved in processing unambiguous sentences and in ambiguity resolution.

  - How exactly the factors are represented and processed are open questions: Minimal Attachment & Late Closure vs. Storage and Integration (DLT)?
    Evidence supporting the DLT in ambiguity resolution

  - **Frequency: What kind of frequency is the human sentence processor sensitive to? We don't know yet.**
    Generalization: Corpus frequencies correlate with RTs, when the appropriate level of corpus analysis is used. (e.g., Desmet & Gibson, 2003)

- What is the time course of information integration?
  - Modular (syntax-first)? Or non-modular?

- **Is the parser serial or constrained parallel?**

To do today:

- The serial / parallel question

- Lexical / structural frequencies and sentence comprehension:
  - ➢ Is the human parser sensitive to structural frequencies?
  - ➢ Is the human parser sensitive to syntactic / lexical frequencies contingent on a syntactic / discourse / world knowledge context?

The serial / parallel question

- Does the human parser retain exactly one structure for the input at each parse state **(serial),** or does it sometimes retain more than one **(parallel)**? See e.g., Gibson & Pearlmutter (2000); Lewis (2000) for discussion.

Serial versus parallel parsing

Frazier (1978), Frazier & Rayner (1982), etc.:
Comprehenders show clear preference asymmetries.

Bill argued the position...

...forcefully.
...was incorrect.

An unlimited parallel processor predicts no asymmetry.

Alternative to unlimited parallel parsing: Multiple alternatives are ranked according to various criteria (**Ranked-Parallel Processing**)

Evidence for (Ranked-) Parallelism (?)

Effects on reading time at disambiguation, of lexical frequencies and plausibility of inappropriate alternatives.

E.g. Garnsey et al. (1997): NP/S ambiguity:
Bill argued the position…

...forcefully.

...was incorrect.

The more plausible the NP interpretation, the greater the ambiguity effect for the S continuation.

Similarly, the more frequent the NP subcategorization, the greater the ambiguity effect for the S continuation.

Evidence for Serialism / Parallelism

Ranked parallel explanation: Both interpretations are calculated in parallel, and size of the ambiguity effect reflects the effort to re-rank the structures.

Alternative, probabilistic serial explanation: Multiple constraints are considered initially, and one is selected stochastically, according to the combination of the constraint weights. The ambiguity effect at disambiguation reflects reanalysis (Traxler et al., 1998; van Gompel et al., 1999).

The serial / parallel question

- If the parser were parallel, then there should be evidence of multiple parses during ambiguous regions.

- E.g., we should see storage costs for each structure (Lewis, 2000)

- But we don't: the ambiguous versions are generally at least as fast as the unambiguous versions (Traxler, Pickering & Clifton, 1988; van Gompel, Pickering & Traxler, 2001; Chen, Gibson & Wolf, in press).

- No evidence of the alternative structures (cf. Spivey & Tanenhaus, 1998; Pearlmutter & Mendelsohn, 1999).

Evidence for serialism:
Traxler, Pickering & Clifton, 1998

The steak with the sauce that was tough / runny / tasty didn't win a prize.

- "Tough": disambiguates to "steak"
- "Runny": disambiguates to "sauce"
- "Tasty": ok with either.

- RTs are numerically **fastest** for the predicates that don't disambiguate.
- These results suggest:
  - ➤ **No competition** between readings (otherwise there would be slowdown for "tasty", which is compatible with both)
  - ➤ One representation is being chosen stochastically (probabilistically), and is followed until information is processed which disambiguates away from this structure.

More, similar evidence for serialism

Traxler, Pickering & Clifton, 1998;
Van Gompel, Pickering & Clifton, 2001;
Chen, Gibson & Wolf, in press

Chen, Gibson & Wolf (in press)
Experiment 2

Do predictions of empty categories in wh-dependencies incur storage costs?

Sentential complement of a noun:

The claim (alleging) that **the cop who the mobster attacked** ignored the informant might have affected the jury.

Relative clause modifying a noun:

The claim which / that **the cop who the mobster attacked** ignored might have affected the jury

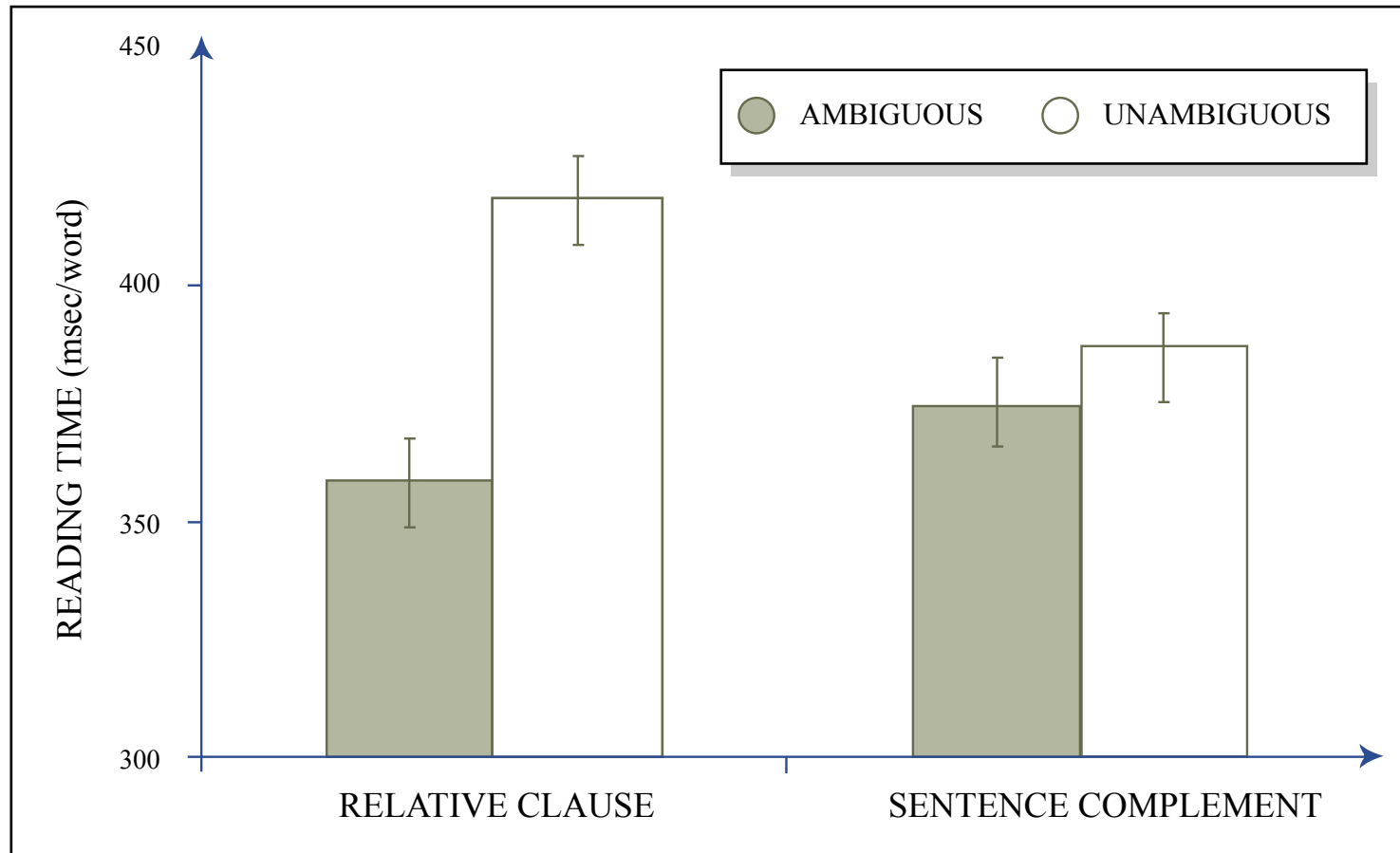# Chen, Gibson & Wolf (in press)
Experiment 2 results

Figure by MIT OCW.

Potential kinds of evidence for a serial model

Faster RTs during the ambiguous region for a temporary ambiguity than for either of the unambiguous controls.

Bimodal patterns of reaction times during disambiguation: Yet to be tested.

Tentative conclusion: Stochastic Serial

Multiple constraints are considered initially, and one is selected stochastically, according to the combination of the constraint weights.

Ambiguity effects throughout the ambiguous region (ambiguous faster than unambiguous) reflect the parser's stochastic choice of one interpretation.

Hold on:
Evidence for (Ranked-) Parallelism (?)
Pearlmutter & Mendelsohn (1999)

To find convincing evidence of parallel parsing, we need to look for effects of a non-preferred alternative during processing of a preferred alternative.

Serial vs. Ranked-Parallel Models

1. Select an ambiguity.
2. Determine preferred interpretation. (Experiment 1)
3. Test for presence of non-preferred interpretation. (Experiment 2)

Sentential Complement (SC) vs. Relative Clause (RC) Ambiguity

The report that the dictator described the country was not evaluated. (SC)

The report that the dictator described was not evaluated. (RC)

Experiment 1 Stimuli

Sentential Complement (SC)

Ambiguous: The report that the dictator described <u>the country</u> was not evaluated…

Unambiguous: The report showing that the dictator described <u>the country</u> was not…

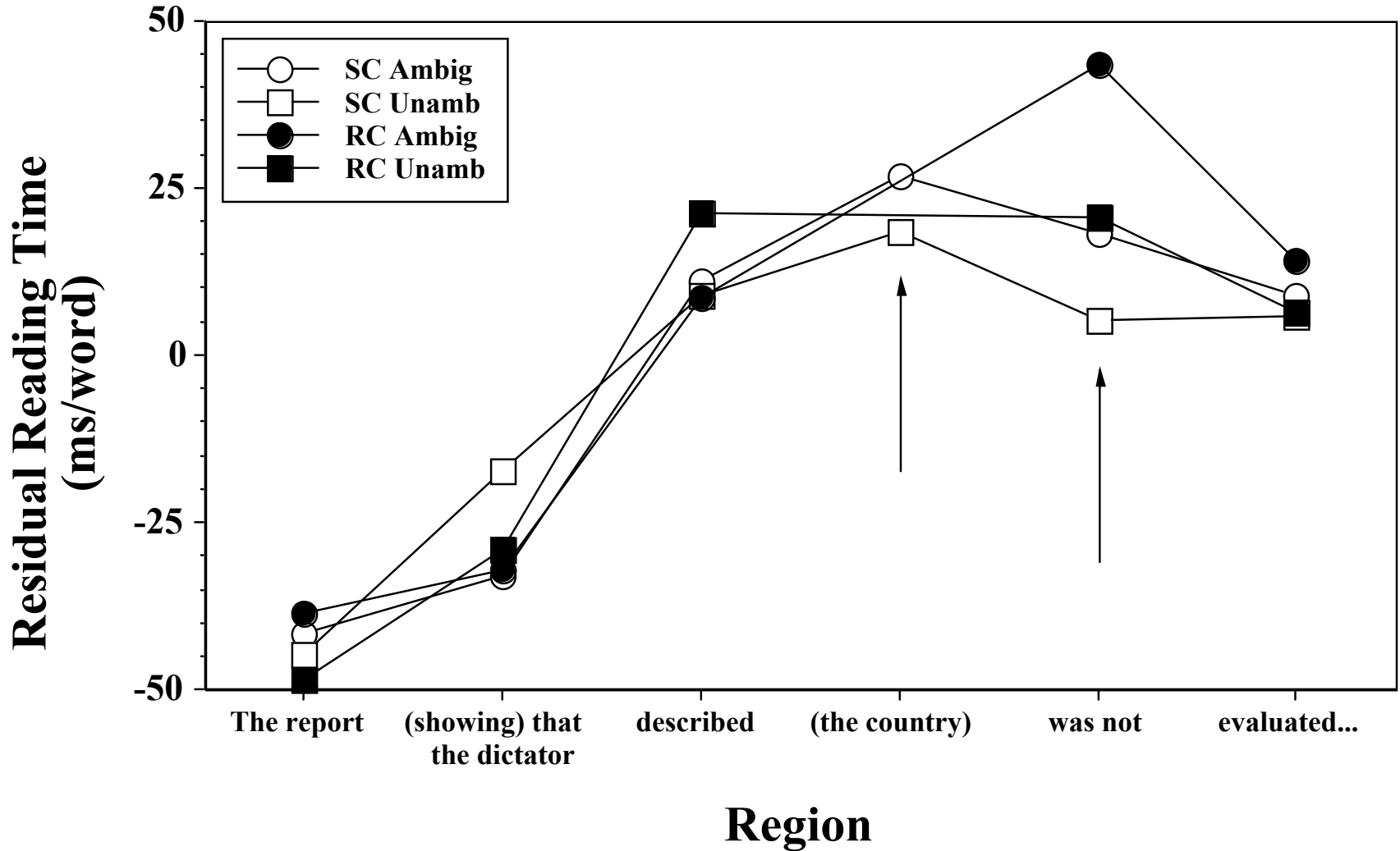Relative Clause (RC)

Ambiguous: The report that the dictator described <u>was not</u> evaluated…

Unambiguous: The report which the dictator described <u>was not</u> evaluated…

Expect difficulty (ambig vs. unamb) for non-preferred interpretation.

Experiment 1 Reading Time by Region

Legend:
- ○ SC Ambig
- □ SC Unamb
- ● RC Ambig
- ■ RC Unamb

X-axis (Region): The report | (showing) that the dictator | described | (the country) | was not | evaluated...

Y-axis: Residual Reading Time (ms/word), from -50 to 50

Experiment 1 Summary

Comprehenders prefer the SC alternative:
- little/no difficulty when SC is forced
- significant difficulty when RC is forced.

Experiment 2

Test for presence of non-preferred (RC) interpretation:
Manipulate RC plausibility while keeping SC plausible.

Correct interpretation is always SC and always plausible.

Plausible (same as Experiment 1 SCs)
The report (showing) that the dictator described <u>the country</u> was not evaluated...

Implausible
The report (showing) that the dictator bombed <u>the country</u> was not evaluated...

Experiment 2 Predictions

Deterministic Serial Model:
- SC is preferred, plausible, and correct throughout.
- RC should never be considered.
- RC plausibility should not influence comprehension.

Ranked-Parallel Model:
- SC is preferred.
- RC is also considered.
- RC (im)plausibility should influence comprehension.
- Implausible ambiguous condition should be hard.
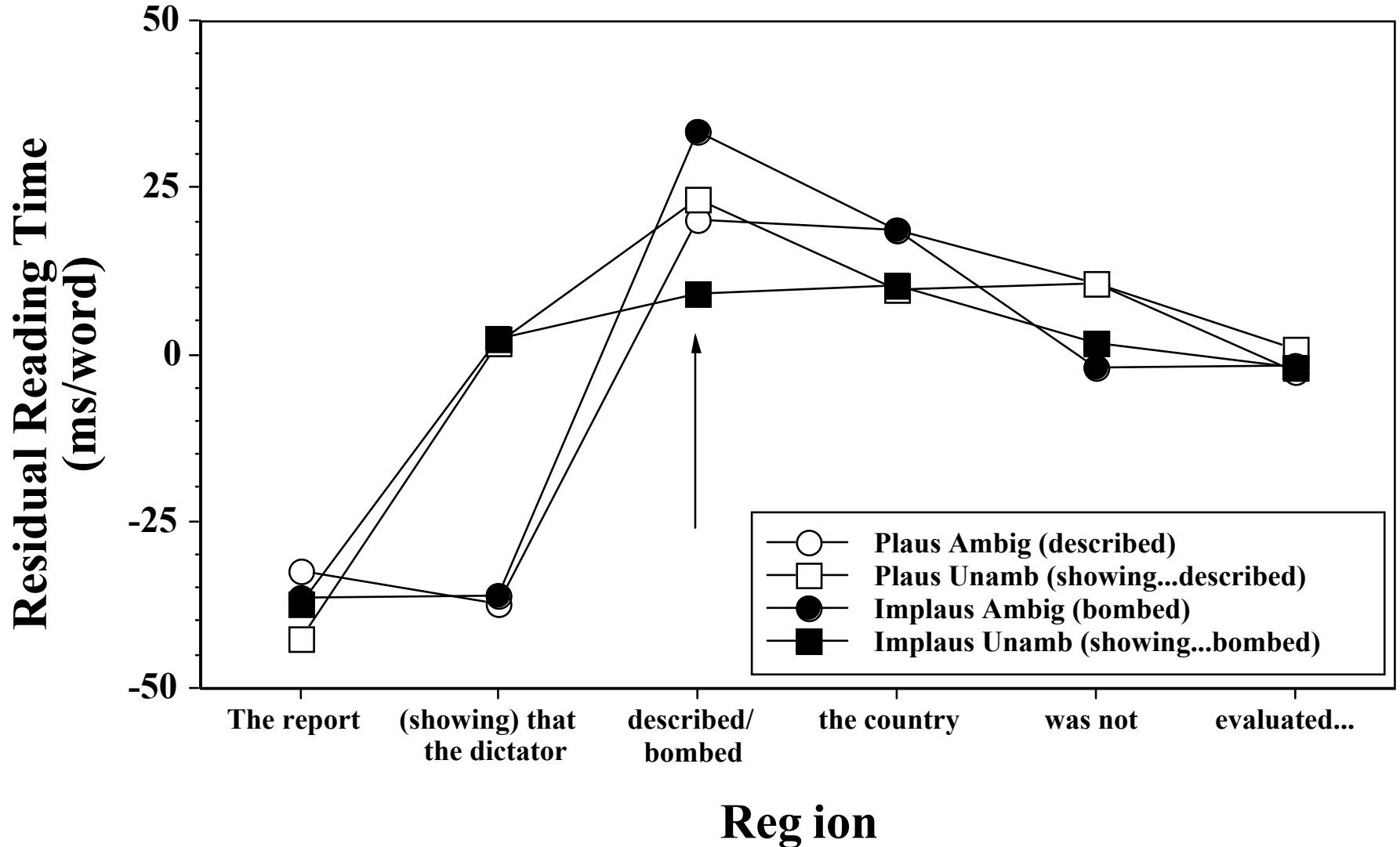
Experiment 2 Plausibility Norming

The report that the dictator described (the country)...
The report that the dictator bombed (the country)...

Rate complete sentences; 1-7 scale, 7 = most plausible.

Plausible RC   Mean   (SD)
SC: The dictator described the country. 5.7   (.76)
RC: The dictator described the report.   5.4   (.75)

Implausible RC
SC: The dictator bombed the country.   5.8   (.86)
RC: The dictator bombed the report.   2.6   (.74)

Experiment 2 Reading Time by Region

Legend:
- ○ Plaus Ambig (described)
- □ Plaus Unamb (showing...described)
- ● Implaus Ambig (bombed)
- ■ Implaus Unamb (showing...bombed)

Y-axis: Residual Reading Time (ms/word), ranging from -50 to 50

X-axis (Region): The report | (showing) that the dictator | described/bombed | the country | was not | evaluated...

Experiment 2 Summary

Implausibility of RC creates difficulty.
RC is being considered even though SC is preferred.
These data rule out deterministic serial parsing.

Alternative: **probabilistic serial parsing**:

SC is preferred most of the time, but occasionally RC is preferred.
Implausibility of RC creates difficulty only when RC is preferred.

Predictions:
• Difficulty of SC at disambiguation depends on lexical preferences.
• Difficulty from RC implausibility depends on lexical preferences.

Sentence Completion Norming

The report that

Count SC completions out of SC and RC completions.

Bill was pleased by the report

Count SC completions out of SC and RC completions.

Count SC completions out of all completions.

Experiment 2 %SC vs. Disambiguation Ambiguity Effect
(Plausible Conditions)
This effect is predicted by both serial and ranked parallel models

$y = 43 - 0.73x$ ($r^2 = 0.226$, $p < .005$)

Ambig - Unamb Reading Time (ms)

%SC Out of All SC+RC
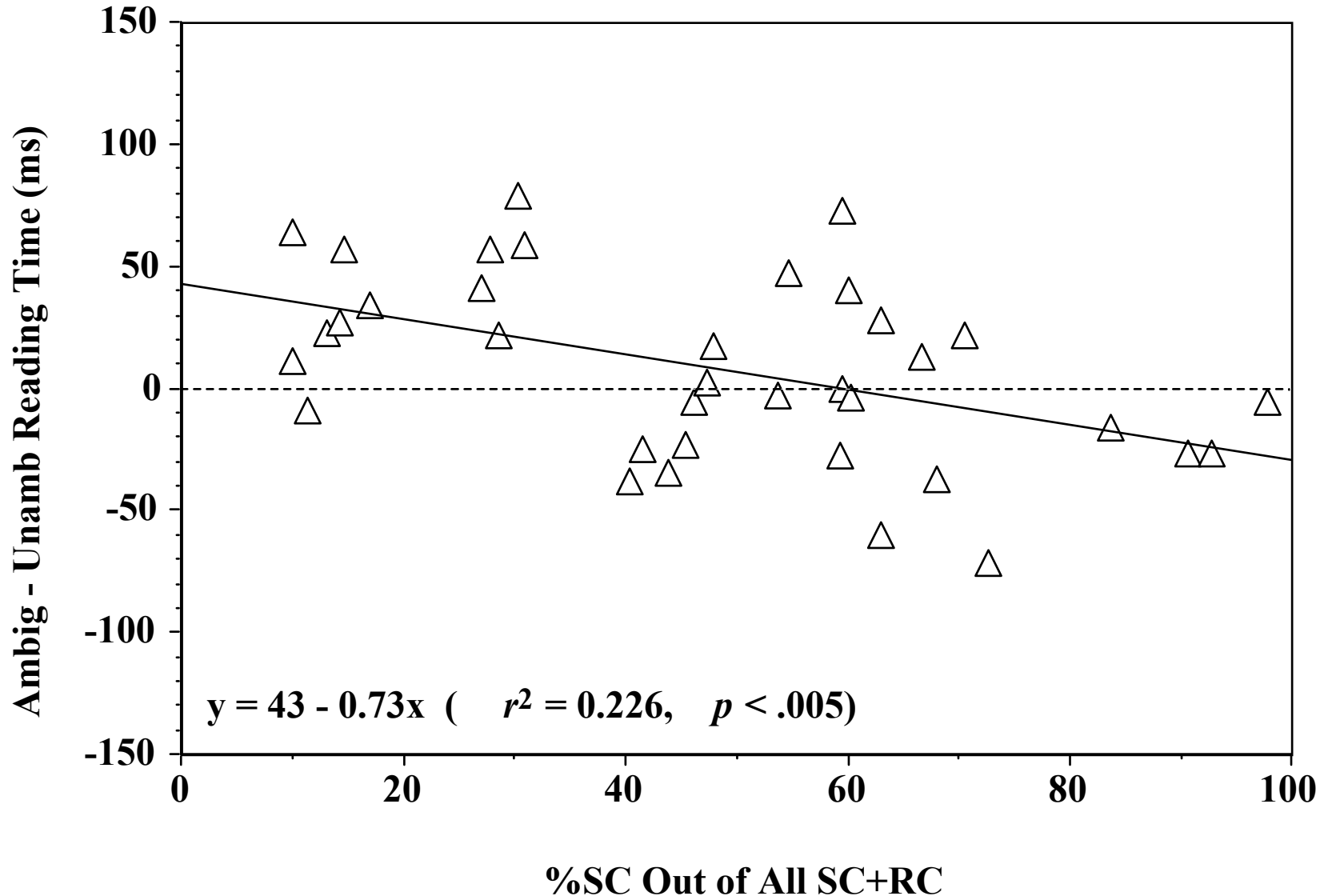
Experiment 2 %SC vs. Embedded Verb Ambiguity Effect
(Implausible Conditions)
The lack of an effect here is a surprise to the serial model
Ranked parallel??

$y = 40 - 0.32x$ ($r^2 = 0.015$, $p > .55$)

Ambig - Unamb Reading Time (ms)

%SC Out of All SC+RC
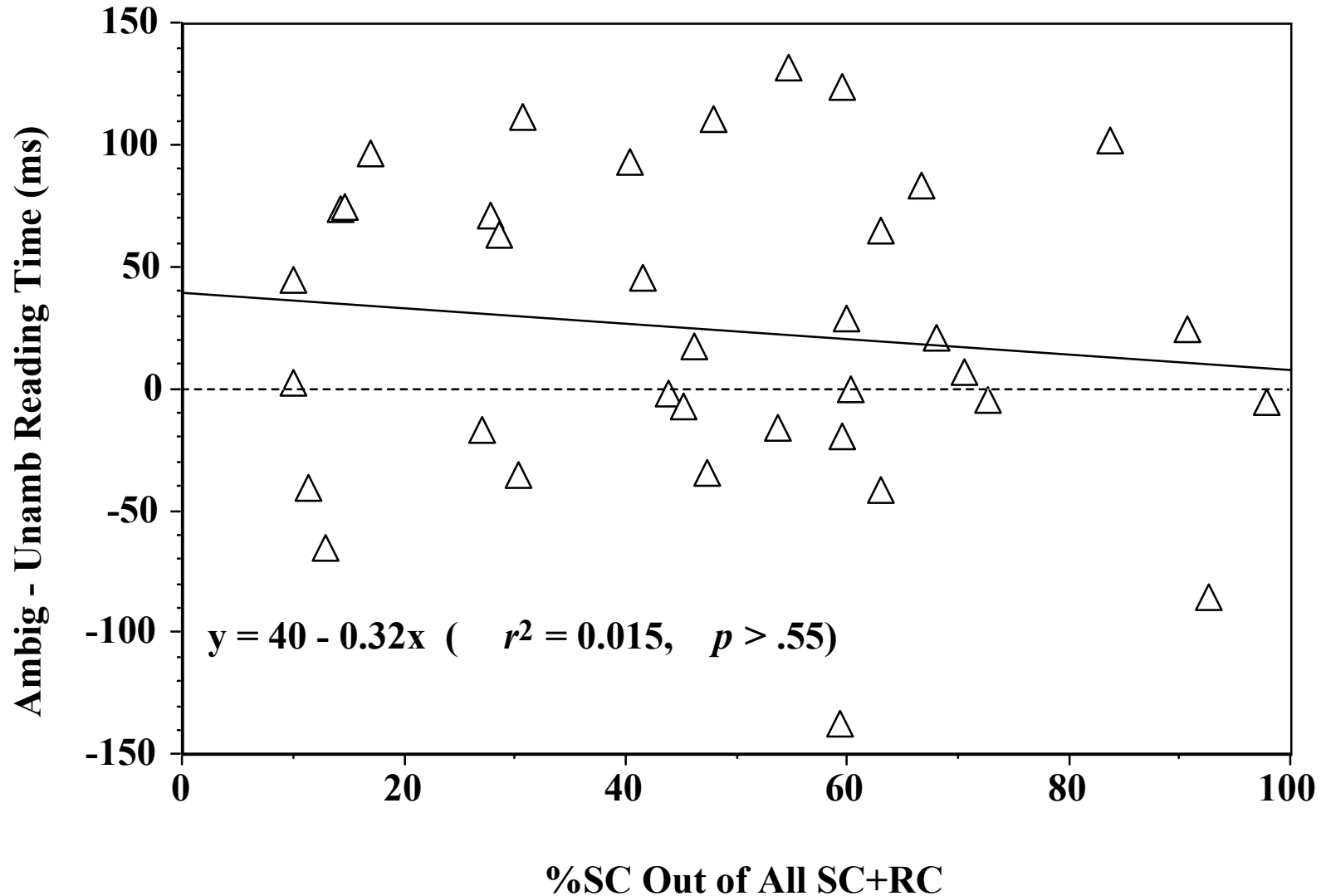
P&M's Conclusions

RC implausibility created difficulty even though SC was preferred.

Both SC and RC were being considered.

SC preference depended on lexical bias.

RC implausibility effect did not.

P&M's Tentative conclusions: Processing is ranked-parallel:

Multiple alternatives are considered simultaneously.

Lexical constraints determine strength of preference.

But P&M's conclusions rest on a **null effect**: no correlation at embedded verb region.

And the ranked parallel model that they propose is not defined.

Tentative conclusion: Stochastic Serial (?)

Multiple constraints are considered initially, and one is selected stochastically, according to the combination of the constraint weights.

Ambiguity effects throughout the ambiguous region (ambiguous faster than unambiguous) reflect the parser's stochastic choice of one interpretation.

Sentence processing: Recap thus far

- Multiple factors are involved in processing unambiguous sentences and in ambiguity resolution.

  - ➢ How exactly the factors are represented and processed are open questions: Minimal Attachment & Late Closure vs. Storage and Integration (DLT)?
    Evidence supporting the DLT in ambiguity resolution

  - ➢ **Frequency: What kind of frequency is the human sentence processor sensitive to? We don't know yet.**
    Generalization: Corpus frequencies correlate with RTs, when the appropriate level of corpus analysis is used.

- What is the time course of information integration?
  - ➢ Modular (syntax-first)? Or non-modular?

- **Is the parser serial or constrained parallel? Stochastic serial?**

The interaction of top-down and bottom-up statistics in syntactic ambiguity resolution

Ted Gibson, MIT

Tabor, Juliano & Tanenhaus (1997)

Two syntactic contexts for the word "that":

Context 1: **Sentence initial**

(1) a. That cheap hotel **is clean** and comfortable to our surprise.

b. That cheap hotels **are clean** and comfortable surprised us.

Determiner preference: "are clean" is read slower than "is clean"

Tabor, Juliano & Tanenhaus (1997)

Two syntactic contexts for the word "that":

Context 2: **Post-verbal**

(2)   a. The lawyer insisted that cheap hotel **is clean** and
comfortable.

b. The lawyer insisted that cheap hotels **are clean** and
comfortable.

Complementizer preference: "is clean" is read slower than
"are clean"

Tabor, Juliano & Tanenhaus (1997)

Graph removed for copyright reasons.

Tabor, Juliano & Tanenhaus (1997)

This result was not predicted by any structure-based principles of the time.

E.g., Minimal Attachment favors the determiner reading in both environments.

Tabor, Juliano & Tanenhaus (1997)

The result also cannot be explained by some frequency-based theories: e.g., Jurafsky (1996)

Jurafsky (1996): People are sensitive to the frequency of use of phrase structure rules.

Jurafsky (1996)

Graph removed for copyright reasons.
"Figure 7."

Jurafsky (1996)

- Beam-width parallel parsing (cf. Gibson , 1991)

Table removed for copyright reasons.
"Table 1."

Jurafsky (1996)

- In Jurafsky's model, the rule probabilities are independent of context: The same probabilities apply for ambiguities in different syntactic contexts.

- Thus the model makes no differential predictions for the resolution of the word "that" in sentence-initial and post-verbal positions

Tabor, Juliano & Tanenhaus (1997)

Key claim: any model will need to keep track of **contingent frequencies.**

A two-part model that accounts for the observations:

(1) a simple recurrent network like that of Elman (1991) to simulate learning the relevant syntactic knowledge;

(2) an attractor-based dynamical system to simulate preferences and times in sentence processing.

Tabor, Juliano & Tanenhaus (1997)

Part 1:

The network was trained on a naturalistic corpus consisting of simplified sentence forms, whose structural frequencies matched the frequencies observed in the Brown corpus.

The weights on the network were then fixed so that no more learning would take place.

Tabor, Juliano & Tanenhaus (1997)

Part 2:

The attractors in the second component of the model, the dynamical system, were then calculated based on the resultant hidden units of the network:

Clumps of similar processing.

Target sentences were then fed to the network, and the values for the hidden units' activations after each word was processed were recorded.

Tabor, Juliano & Tanenhaus (1997)

Part 2:

Simulating processing times: **gravitation**

Count the number of iterations that the model requires for a resultant point to be attracted to the center of one of the attractors of the system, based on a gravitational model of attractor speed.

Tabor, Juliano & Tanenhaus (1997)

Graph removed for copyright reasons.
"Figure 7."

Tabor, Juliano & Tanenhaus (1997)

Key claim: any model will need to keep track of **contingent frequencies.**

Tabor et al. proposed a two-part model to account for the observations:

(1) a simple recurrent network like that of Elman (1991) to simulate learning the relevant syntactic knowledge;

(2) an attractor-based dynamical system to simulate preferences and times in sentence processing.

Complex model. Not an analytic model: No way to understand what the model will do without running the model.

Tabor et al. (1997)

Main claim: ***The context-dependent lexical frequency hypothesis*** (Gibson, submitted)

The sentence processing mechanism tracks the frequencies of words (e.g., determiner vs. complementizer for "that") within all different syntactic contexts.

Sentence-initial: more determiners that comps

Post-verbal: more comps than determiners

Tabor, Juliano & Tanenhaus (1997)

Experiment 2: "that" vs. "those" in an unambiguous environment:

(3) a. The lawyer visited **that cheap hotel** to stay for the night.

b. The lawyer visited **those cheap hotels** to stay for the night.

Result: "that cheap hotel" is processed slower than "those cheap hotels"

Tabor, Juliano & Tanenhaus (1997)

This result is also not explained by any models of the time.

The result was captured by Tabor et al.'s model.

Explanation within the context-dependent lexical frequency hypothesis

People treat contexts following all verbs similarly, whatever the verb subcategorization properties might be.

Thus, the word "that" is initially taken to be a complementizer in (3a) – despite the subcategorization properties of "visited" – because most instances of "that" following verbs are complementizers

An alternative approach: An interaction between top-down syntactic expectations and bottom-up lexical frequencies (Gibson, submitted)

Part 1: **Context-independent lexical frequency hypothesis (bottom-up component of the model):** People are sensitive to the lexical frequency distributions of each word, independent of context: the unigram lexical frequencies for a word.

Part 2: **Syntactic probability hypothesis (top-down component of the model):** People are sensitive to the lexical and syntactic expectations at each word, giving rise to a probability distribution (PD) of expectations at each word for upcoming syntactic categories. The set of (probability, feature-structure) pairs following words $x_1 \dots x_n$ will be notated as **PD($x_1 \dots x_n$)**.

Note: the syntactic probability hypothesis is **context-dependent**.

Because of smoothing, there is a non-zero probability associated with matching the most general set of syntactic features. It is initially arbitrarily assumed that this minimal probability is 0.01. All words match at least this feature structure, and possibly more in the list.

Top-down syntactic expectations and bottom-up lexical frequencies

Given a set of (probability, feature-structure) pairs PD (x_1 ... x_n) after processing the input x_1 ... x_n, and an input word w, the **context-dependent lexical bias (CDLB)** associated with lexical entry w_i, one of w's lexical entries, is defined as:

CDLB(w_i) = (the sum of all the probabilities from the (probability, feature-structure) pairs PD (x_1 ... x_n) in which the category of the feature-structure in the pair subsumes the category of w_i) * (the unigram probability of w_i)

The syntactic-probability / lexical frequency ambiguity resolution proposal (serial hypothesis):

The likelihood that the parser will choose lexical entry w_i is:

CDLB(w_i) / the sum of CDLB(w_j), for all w_j, lexical entries for w.

(Note: CDLB here = LLB in the paper! The notation has changed.)

Top-down syntactic expectations and bottom-up lexical frequencies

- Example:

  The word "that": 77.5% complementizer; 11.1% determiner; 11.5% pronoun (from Brown corpus, Kucera & Francis, 1977)

  Syntactic expectations (from samples from the 1989 AP newswire corpus, Marcus et al.,1994): (Note: no serious theory of syntax yet: These are just estimates from corpora!)

Environment	Category of the ensuing syntactic head		
	Complementizer	Determiner	Pronoun
Sentence-initially	.0004	.205	.178
Following an S-bias verb	.496	.081	.009
Following the non-S verbs	0	.318	.058

Top-down syntactic expectations and bottom-up lexical frequencies

- Local Lexical Biases for "that" **following an S-bias verb:**
  - ➢ CDLB (that$_{comp}$) = .775 * .506 = .3922
  - ➢ CDLB (that$_{det}$) = .111 * .091 = .0101
  - ➢ CDLB (that$_{pronoun}$) = .115 * .019 = .0022

  Therefore, **following an S-bias verb** the relative likelihoods for pursuing:
  a complementizer = 0.970
  a determiner = 0.035
  a pronoun = 0.005

  Therefore there is a preference for the complementizer reading.

Top-down syntactic expectations and bottom-up lexical frequencies

- Local Lexical Biases for "that" **sentence-initially**:
  - ➢ CDLB (that$_{comp}$) = .775 * .0104 = .0081
  - ➢ CDLB (that$_{det}$) = .111 * .215 = .0239
  - ➢ CDLB (that$_{pronoun}$) = .115 * .188 = .0216

  Therefore, **sentence-initially** the relative likelihoods for pursuing:
  a complementizer = 0.151
  a determiner = 0.446
  a pronoun = 0.404

  Therefore there is a preference for the deictic determiner / pronoun reading.

Top-down syntactic expectations and bottom-up lexical frequencies

- Local Lexical Biases for "that" **following a verb like "visited"**:

  - CDLB (that$_{comp}$) = .775 * .01 = .0078
  - CDLB (that$_{det}$) = .111 * .328 = .0364
  - CDLB (that$_{pronoun}$) = .115 * .068 = .0078

  Therefore, **following a verb like "visited"** the relative likelihoods for pursuing:

  a complementizer = 0.149

  a determiner = 0.700

  a pronoun = 0.150

  Therefore, there is some likelihood in following the complementizer reading, in spite of its ungrammaticality, leading to increased RTs compared to an unambiguous control.

Comparison with earlier approaches

- Bottom-up lexical frequencies:

  Many psycholinguists have noted the importance of lexical frequencies (e.g., MacDonald, Pearlmutter & Seidenberg, 1994; Trueswell, 1996). Some of these have also noted the importance of top-down syntactic expectations as well.

  But none have provided a formal testable proposal, which is testable on arbitrary new structures / lexical ambiguities.

Comparison with earlier approaches

Top-down syntactic expectations:

- Earlier psycholinguistic proposals:
  Gibson (1991): no probabilities
  Elman (1991); Tabor et al., (1997): **not analytic**
  Jurafsky (1996), Hale (2002): **top-down only**
  Corley & Crocker (2000): N-grams: **top-down only**

- Computational linguistic proposals:
  E.g., see summaries of N-grams etc. in Jurafsky & Martin (2000); Manning & Schutze (2000)
  Top-down only.
  One key insight: Need **smoothing** for low probability estimates. Don't assume that some things **never** occur: too hard to know this (**language is complicated!**). Thus smooth low probability estimates to small non-zero probabilities.

Comparison with earlier approaches

How does this fit within the storage / expectation cost component of the Dependency locality theory (Gibson, 1998; 2000)?

Syntactic expectations having near probability 1 of eventually occurring are associated with syntactic expectation cost.

Both this probabilistic disambiguation theory (the CDLB hypothesis) and the DLT are highly top-down, expectation-driven hypotheses.

The serial / parallel question

- Does the human parser retain exactly one structure for the input at each parse state **(serial),** or does it sometimes retain more than one **(parallel)**? See e.g., Gibson & Pearlmutter (2000); Lewis (2000) for discussion.

- If the parser were parallel, then there should be evidence of multiple parses during ambiguous regions.

- E.g., we should see storage costs for each structure

- But we don't: the ambiguous versions are at least as fast as the unambiguous versions (Traxler, Pickering & Clifton, 1988; van Gompel, Pickering & Traxler, 2001; Chen, Gibson & Wolf, in press).

- No evidence of the alternative structures (cf. Spivey & Tanenhaus, 1998; Pearlmutter & Mendelsohn, 1999).

Experiments to test the top-down / bottom-up hypothesis

- Post-prepositional environment:

  a. The lawyer for that skilled surgeon asked for a raise.

  b. The lawyer for those skilled surgeons asked for a raise.

- Prediction of context-dependent lexical frequency hypothesis (Tabor et al., 1997): "that" should pattern like "those" in this environment, because "that" almost never occurs as a complementizer in this environment:

Environment	Category of "that"			Category of "those"			Category of "this"		
	Comp	Det	Pron	Comp	Det	Pron	Comp	Det	Pron
Post-verb	2908	248	200	0	55	54	0	484	236
Post-preposition	9	551	373	0	190	290	0	1972	356
Sentence-initial	51	121	292	0	21	49	0	508	652

Experiments to test the top-down / bottom-up hypothesis

- Post-prepositional environment:
  a. The lawyer for that skilled surgeon asked for a raise.
  b. The lawyer for those skilled surgeons asked for a raise.

- Prediction of the top-down / bottom-up hypothesis (CDLB), including the context-independent lexical frequency hypothesis: "that" should be read slower than "those" post-prepositionally, for the same reason that it is read slower following verbs like "visited":

- Local Lexical Biases for "that" **following a preposition**:
  - CDLB (that$_{comp}$) = .775 * .01 = .0078
  - CDLB (that$_{det}$) = .111 * .398 = .0364
  - CDLB (that$_{pronoun}$) = .115 * .044 = .0078

  Therefore, **following a preposition** the relative likelihoods for pursuing:
  a complementizer = 0.136
  a determiner = 0.775
  a pronoun = 0.088

  Therefore, there is some likelihood in following the complementizer reading, in spite of its ungrammaticality, leading to increased RTs compared to an unambiguous control.
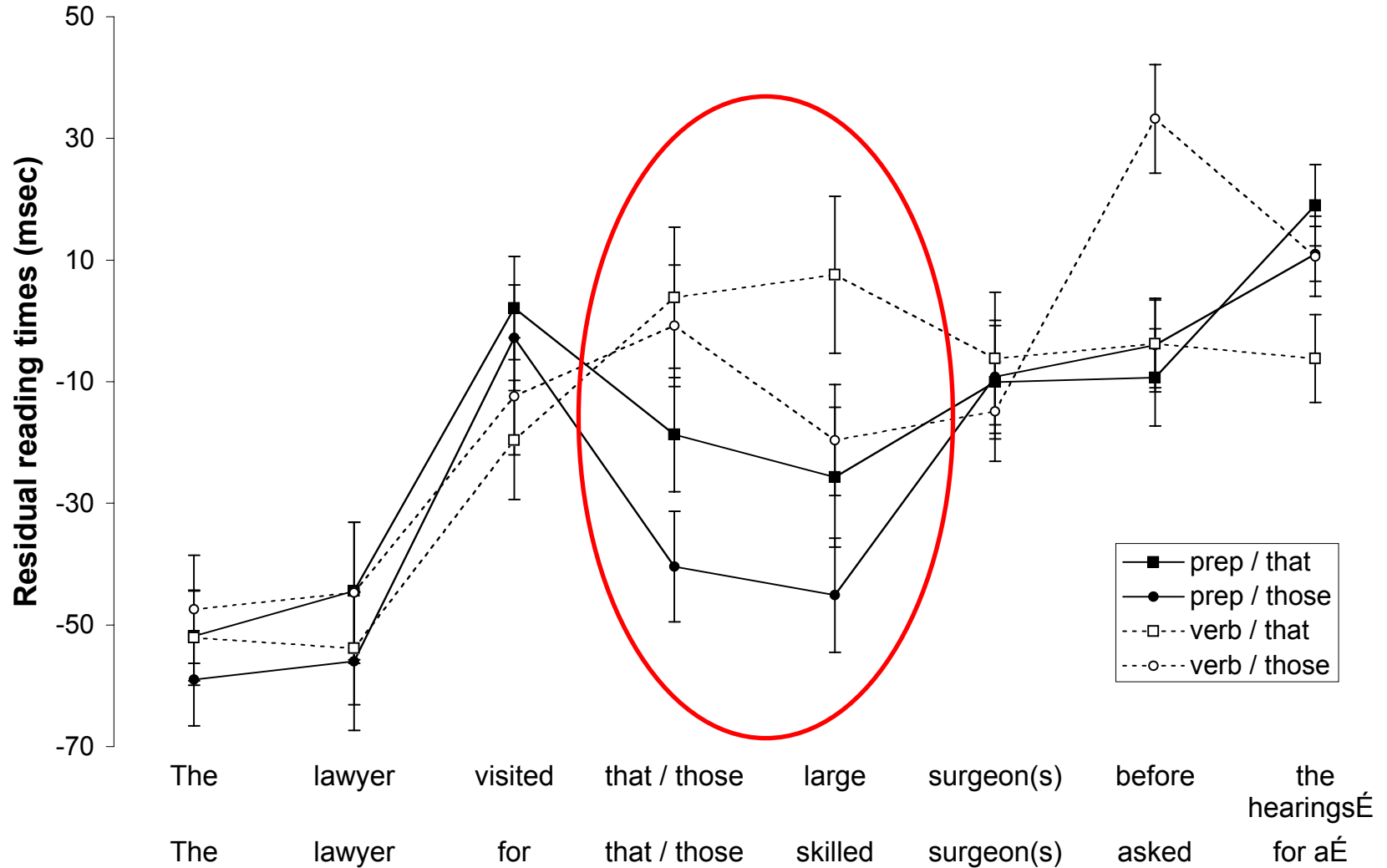
Experiment 1

Materials:

a. Prep, that: The lawyer for that skilled surgeon asked for a raise.

b. Prep, those: The lawyer for those skilled surgeons asked for a raise.

c. Verb, that: The lawyer visited that skilled surgeon before the hearings began.

d. Verb, those: The lawyer visited those skilled surgeons before the hearings began.

60 participants, 20 items, 74 fillers, Latin Square design.

Experiment 1: Results

Experiment 1: Results

Experiment 1 results: Same slowdown for "that" following a preposition as for a verb.

This is consistent with the context-independent lexical frequency approach (CDLB), but not predicted by the context-dependent lexical frequency approach.
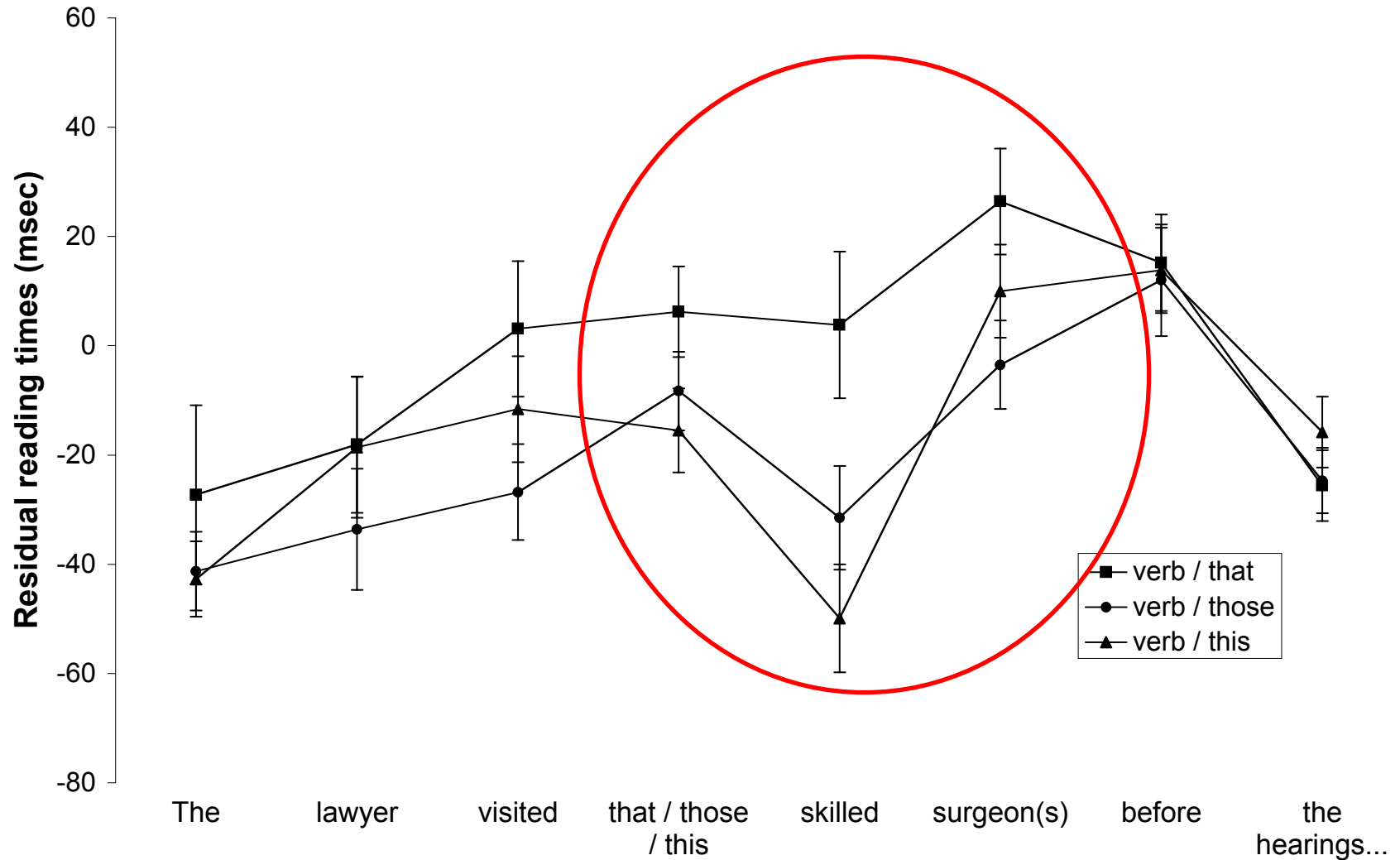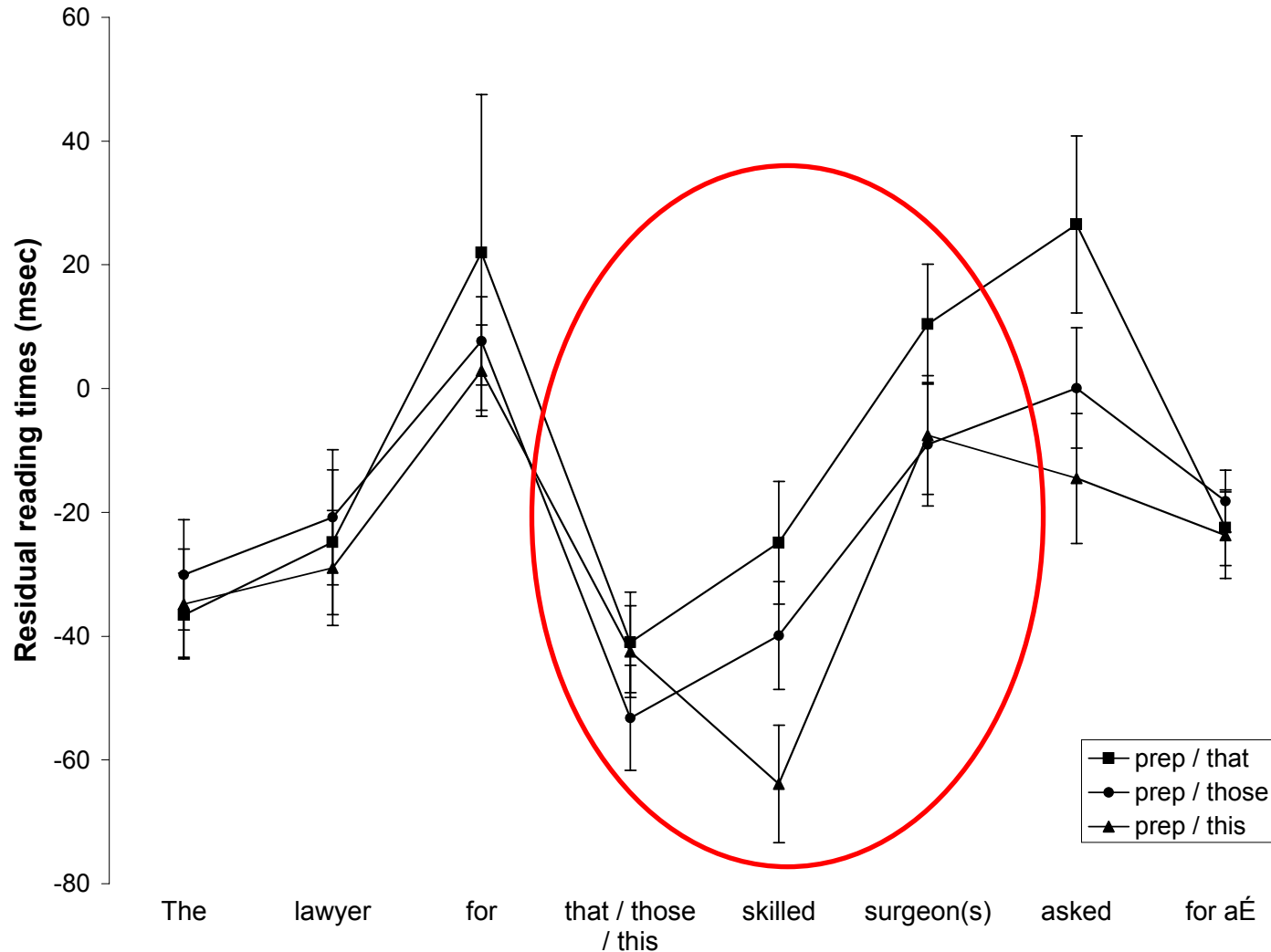
Experiment 2

Materials:

a. Prep, that: The lawyer for that skilled surgeon asked for a raise.

b. Prep, those: The lawyer for those skilled surgeons asked for a raise.

c. Prep, this: The lawyer for this skilled surgeon asked for a raise.

d. Verb, that: The lawyer visited that skilled surgeon before the hearings began.

e. Verb, those: The lawyer visited those skilled surgeons before the hearings began.

f. Verb, this: The lawyer visited this skilled surgeon before the hearings began.

97 participants, 36 items, 94 fillers, Latin Square design.

Experiment 2 results: Verb items

Experiment 2 results: Preposition items

Experiment 2: Results

Experiment 2 results: Same slowdown for "that" relative to "this" or "those" following a preposition as for a verb.

This is consistent with the context-independent lexical frequency approach (CDLB), but not predicted by the context-dependent lexical frequency approach.

Experiment 3: Test of a prediction of the CDLB hypothesis: Sentence-initial contexts

- Top-down syntactic expectations and bottom-up lexical frequencies
- Local Lexical Biases for "that" **sentence-initially**:
  - ➢ CDLB (that_{comp}) = .775 * .0104 = .0081
  - ➢ CDLB (that_{det}) = .111 * .215 = .0239
  - ➢ CDLB ($\text{that}_{pronoun}$) = .115 * .188 = .0216

  Therefore, **sentence-initially** the relative likelihoods for pursuing:
  a complementizer = 0.151
  a determiner = 0.446
  a pronoun = 0.404

 Therefore there is a preference for the deictic determiner / pronoun reading.

  Note also: The probability of following the determiner / pronoun reading is not 100%. Therefore there should be an difficulty effect compared to an unambiguous control, such as "those".
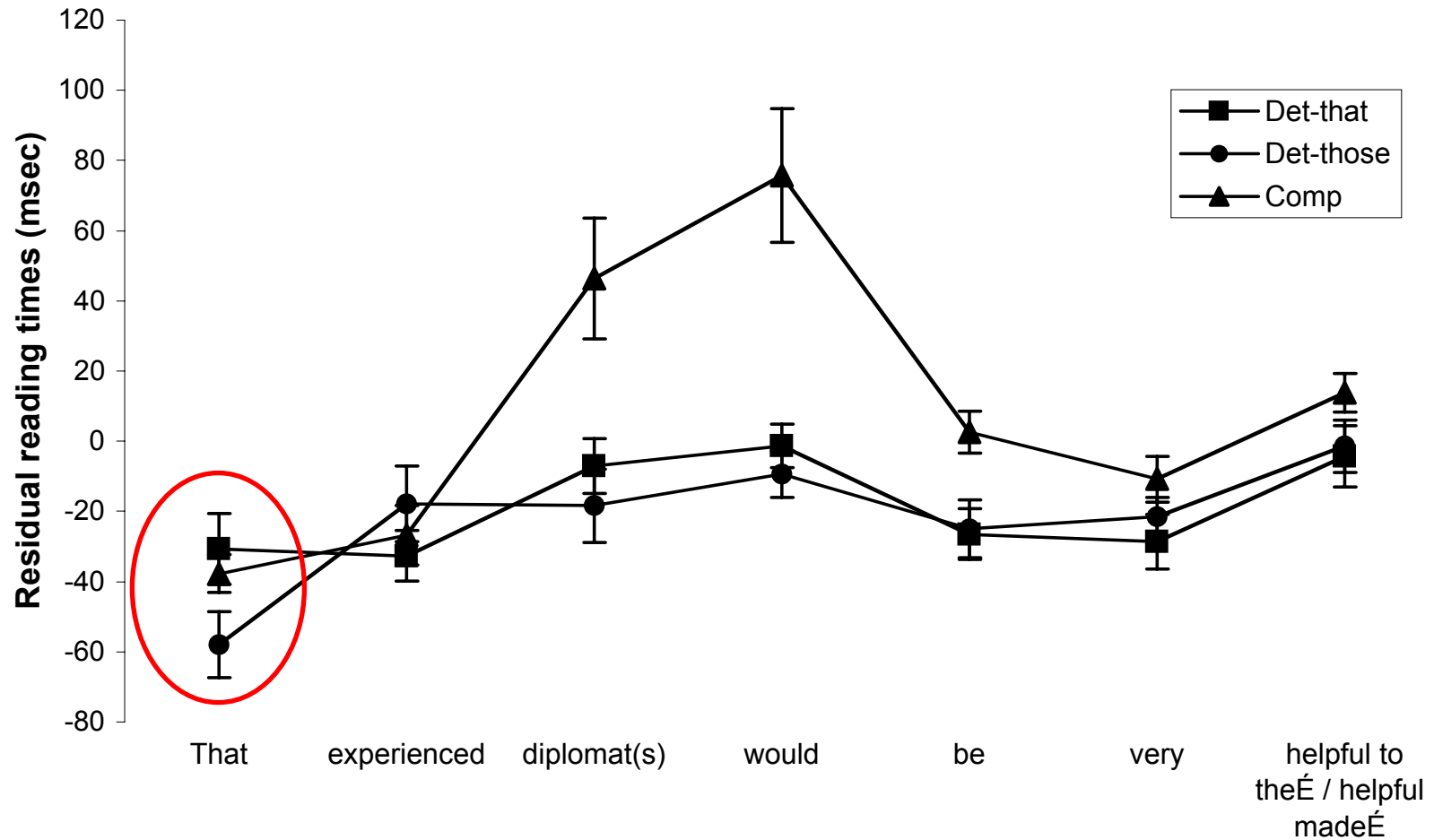
Experiment 3

- **Materials:**

a. Determiner "that": That experienced diplomat would be very helpful to the lawyer.

b. Determiner, "those": Those experienced diplomats would be very helpful to the lawyer.

c. Complementizer "that": That experienced diplomats would be very helpful made the lawyer confident.

Experiment 3: Results
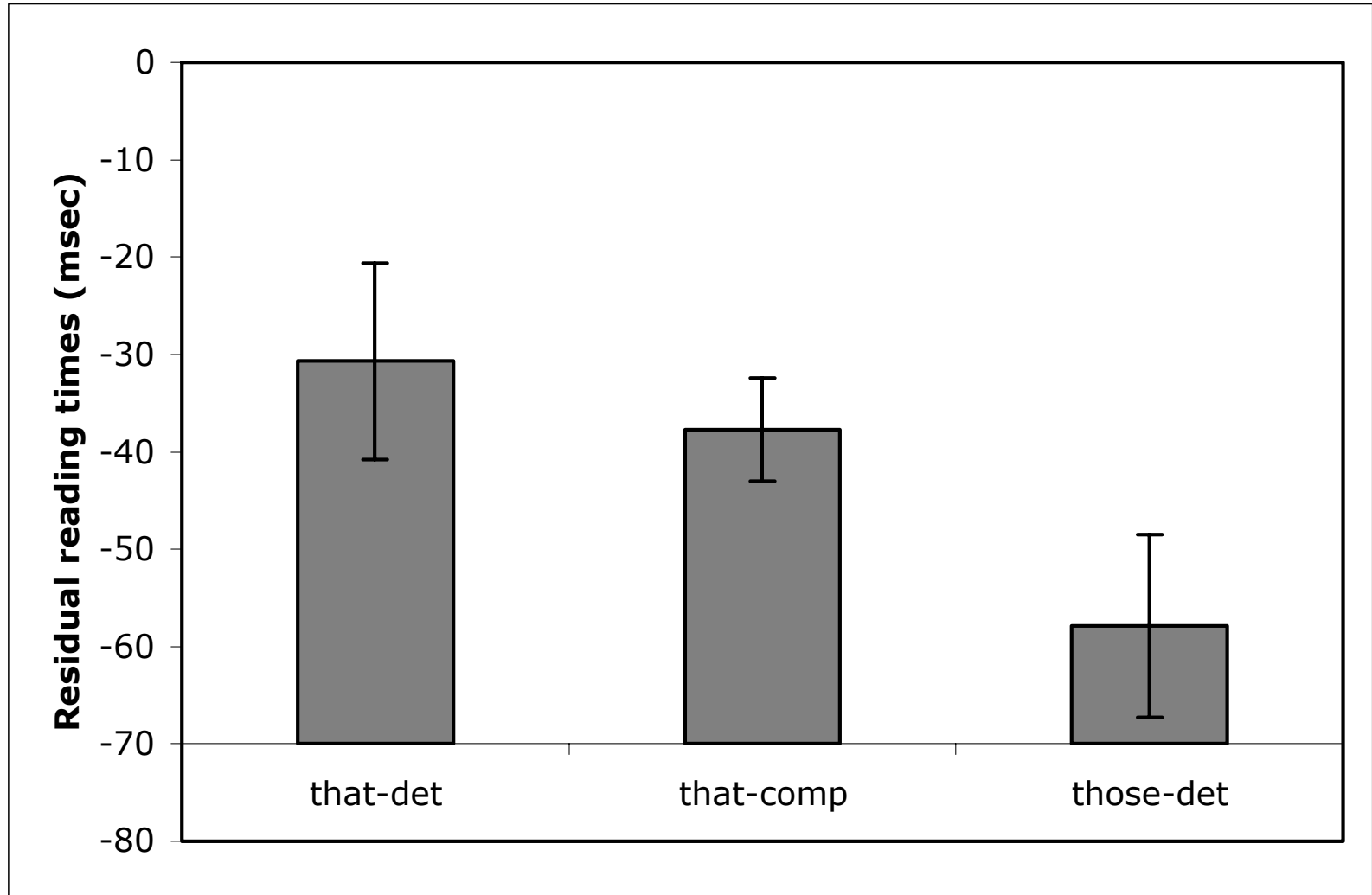
Experiment 3: Results

Experiment 3: Results

Experiment 3 results: People are slightly slower reading "that" sentence-initially than "those", as predicted by the top-down / bottom-up CDLB approach.

New data: Tabor, Galantucci & Richardson (2004)

Old result: The MV/RR ambiguity can result in difficulty:

The player (who was) tossed the frisbee smiled at the coach

New experiment: MV/RR following an object NP, where the MV should be ungrammatical.

a. Ambiguous MV/p-part, post-object-NP context:
The coach smiled at the player **tossed the frisbee**.
b. Ambiguous MV/p-part, "who was" context:
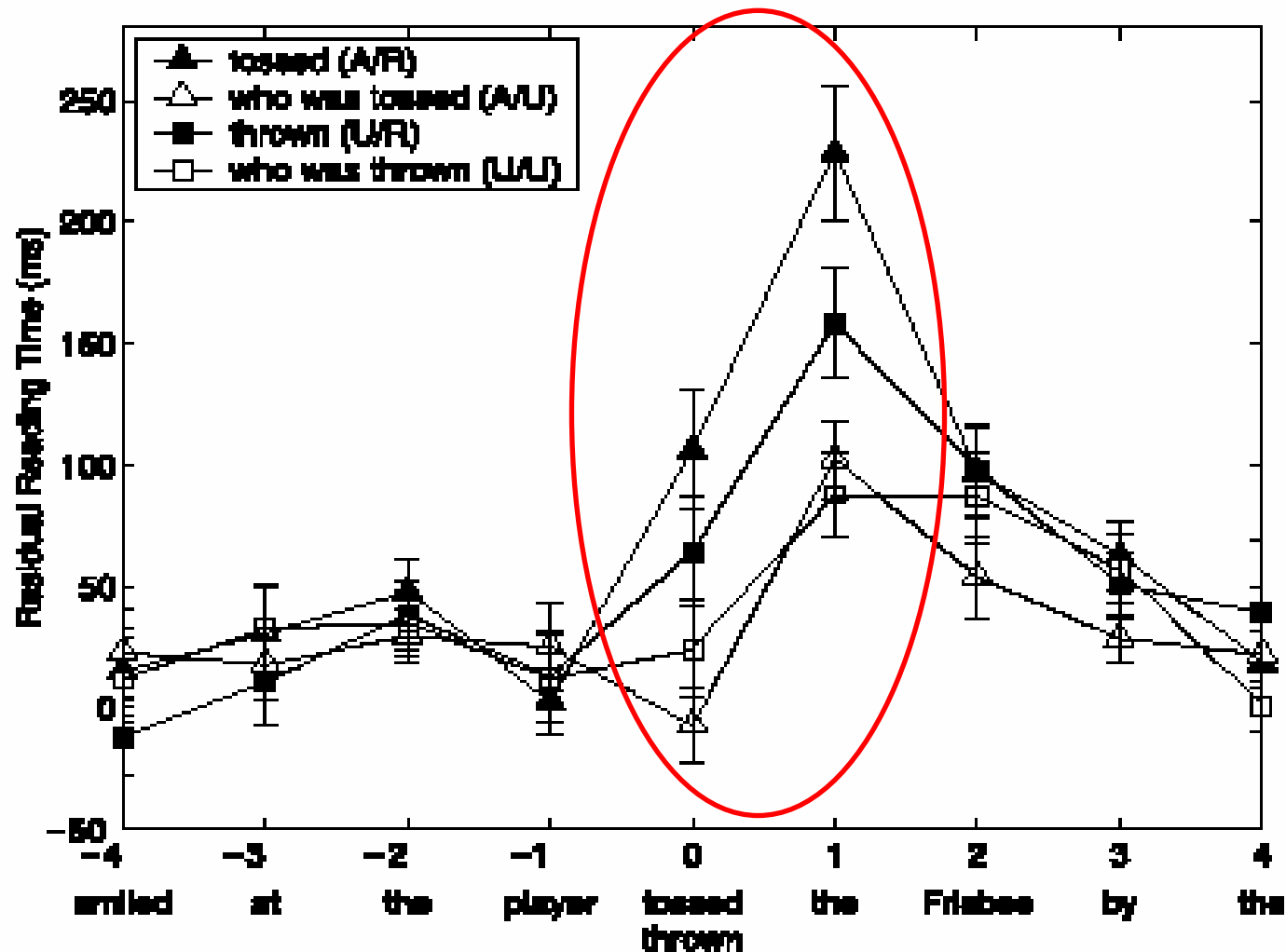The coach smiled at the player who was **tossed the frisbee**.
c. Unambiguous p-part, post-object-NP context:
The coach smiled at the player **thrown the frisbee**.
d. Unambiguous p-part, "who was" context:
The coach smiled at the player who was **thrown the frisbee**.

Tabor et al. (2004): Results

Tabor et al. (2004): Results

- People have difficulty with the ambiguous RR, even in a syntactic context in which the MV is disallowed.

- Tabor et al.'s (2004) proposed model: two parts to sentence processing, one of which is a moving window of local syntactic coherence: "the player tossed the frisbee" is grammatical as an SVO structure, so build that. Like N-grams, going back 2 or 3 words (Corley & Crocker, 2000).

- Note: This is very different from Tabor et al.'s (1997) model: Local syntactic coherence explicitly **did not** matter: "The lawyer visited that cheap hotel…"
  "that" as a complementizer is syntactically incompatible with the local input.

The CDLB account of
Tabor et al.'s (2004) data

- Lexical frequencies for ambiguous MV/RR verbs in Tabor et al.'s materials:
  MV: 45.7%; past-part: 54.3%

- Syntactic expectations following a sequence like "The coach smiled at the player":
  1% passive-participle; 41% prepositions; 19% conjunctions; etc.

- CDLB ($tossed_{MV}$) = .457 * .01 = .00457

- CDLB ($tossed_{past-part}$) = .543 * .02 = .01086

- P(MV) = 0.296

- P(passive-part) = 0.704

Therefore, people follow the MV interpretation about 30% of the time, resulting in an immediate reanalysis effect, because this lexical entry is not compatible with the current structure.

The CDLB account of
Tabor et al.'s (2004) data

- Lexical frequencies for ambiguous MV/RR verbs in Tabor et al.'s materials:

  MV: 45.7%; past-part: 54.3%

- Syntactic expectations following a sequence like "The coach smiled at the player who was" (the unambiguous control):

  41% passive-participle; 0% main verbs.

- CDLB ($tossed_{MV}$) = .457 * .01 = .00457

- CDLB ($tossed_{past-part}$) = .543 * .42 = .2281

- P(MV) = 0.02

- P(passive-part) = 0.98

Therefore, people follow the past-part interpretation most of the time in the unambiguous control, as desired.

Similar results for the unambiguous past-part verbs "thrown": no MV reading, so no tendency to follow an MV interpretation.

Tabor et al.'s (2004) proposed model

- Tabor et al.'s (2004) proposed model:
  - ➢ (a) global expectations of what's likely to come next (context-dependent expectations);
  - ➢ (b) a moving window of local syntactic coherence: "the player tossed the frisbee" is grammatical as an SVO structure, so build that. Like N-grams, going back 2 or 3 words (Corley & Crocker, 2000).

- So why is there difficulty at "tossed the frisbee" in the temporarily ambiguous version?

  Because there is **competition** between the ungrammatical MV SVO structure and the correct RR structure.

But:

1. This is a one of the **first** examples of "competition" between multiple readings for the same input. As we have seen earlier, people tend to process temporarily ambiguous material **faster** than the disambiguated versions.
2. No theory is provided about the grammar that would allow such a structure.

Tabor et al.'s (2004) proposed model

Conceptual issue: Tabor et al.'s (2004) proposal assumes that people keep track of **both** (a) **global expectations** of what's likely to come next (context-dependent expectations; and (b) **local expectations**, derived from a local window, like n-grams.

The CDLB approach assumes just the first component.

Work in progress:
Tests of the CDLB hypothesis: Test 1

- Tabor et al.'s (2004) materials:

  The lawyer met the defendant (who was) examined (the previous day) by the district attorney.

- Observation 1: Tabor et al. (2004) always had an adverbial NP following the ambiguous MV/p-part ("the previous day")

- Test 1: The CDLB predicts that this NP is not needed to generate the ambiguity effect.

Work in progress:
Tests of the CDLB hypothesis: Test 2

- Tabor et al.'s (2004) materials:

  The lawyer met the defendant (who was) examined the previous day by the district attorney.

- Observation 2: The matrix verb does not subcategorize for a p-part.

- Test 2: Change the matrix verb so that it subcategorizes for a p-part. The CDLB predicts a reduction in the ambiguity effect:

The lawyer (met / had / got) the defendant (who was) examined by the district attorney.

Work in progress:
Tests of the CDLB hypothesis: Test 3

- Tabor et al.'s (2004) materials:

  The lawyer met the defendant / evidence (who was) examined the previous day by the district attorney.

- Observation 3 (due to Tabor et al. 2004): The plausibility of the NV defendant / evidence examined as a MV/RR affects the ambiguity effect.

- This is predicted by the CDLB approach if the MV reading is being followed a substantial fraction of the time: confusion with the MV.

- Test 3: Vary the matrix verb so that it subcategorizes for a p-part. The CDLB predicts a reduction in the ambiguity effect:

The lawyer (met / got) the (defendant / evidence) (who was) examined by the district attorney.

The CDLB predicts a reduction in the ambiguity effect for the "defendant" cases when following a matrix verb like "got", because the MV will be considered less in this case.

Work in progress:
Tests of the CDLB hypothesis: Test 4

- Look at lexical ambiguity more generally: E.g., N/V ambiguity in English. Whenever there is a word that is highly biased toward one interpretation, there should be parsing evidence of that interpretation, even in unambiguous syntactic environments.

- **Biased N examples**:

  a. Ambiguous, noun context: Helen wants a **bag** made of leather for her birthday.

  b. Ambiguous, verb context: Helen wants to **bag** the dried fruit for the winter.

  c. Unambiguous, noun context: Helen wants a **jacket** made of leather for her birthday.

  d. Unambiguous, verb context: Helen wants to **add** the dried fruit to the pie she is making.

Work in progress:
Tests of the CDLB hypothesis: Test 4

- Look at lexical ambiguity more generally: E.g., N/V ambiguity in English. Whenever there is a word that is highly biased toward one interpretation, there should be parsing evidence of that interpretation, even in unambiguous syntactic environments.

- **Biased V examples**:

  a. Ambiguous, verb context: John has to **build** a table for the outside terrace by late May.

  b. Ambiguous, noun context: John has a **build** of a man who exercises on a regular basis.

  c. Unambiguous, verb context: John has to **carry** a table to his neighbor's place for the barbecue.

  d. Unambiguous, noun context: John has a **personality** of a man who has overcome many difficulties in life.

Thus the CDLB applies very generally, in normal language processing.

Big open question

- Where do the syntactic / semantic expectation probabilities come from? What is their form?
- This is probabilistic knowledge, probably learned.

- Very complex: Context-dependent phrase-structure information.

Sentence processing: Ambiguity Resolution

- Multiple factors are involved in sentence processing / ambiguity resolution:
  - ➤ Syntactic structure:
    - Keep dependencies close
    - Syntax: keeping track of probabilities of upcoming syntactic heads
    - Syntactic storage: minimize syntactic expectations
  - ➤ Lexical frequencies: important in disambiguation

  - ➤ Plausibility of the resultant structures
  - ➤ Discourse context
  - ➤ Intonation