Anti-local contexts improve the overall speed of dependency completion

Matthew Wagers
Department of Linguistics
University of California, Santa Cruz
Goal

• What is the relation between predictive forces in language comprehension and the concept of focal attention?
Focal attention

• … is extremely limited

For sequentially-presented information, the capacity of focal attention appears limited to the last “unit” processed (Wickelgren et al., 1980; Garavan, 1998; Cowan, 2001; McElree, 2006; Jonides et al., 2008).

Building structured representations for sequentially-presented input will often require shunting information between memory and focal attention.
Longer == easier

- Anti-local contexts

Longer can be easier.

E.g., Jaeger, Fedorenko & Gibson, submitted

The understudy that the agent telephoned about the job in Los Angeles shared the story …
Question

• Today’s investigation

• What is the nature of anti-local facilitation?
  
  • RT facilitation is fed by many factors: strength of underlying encoding, speed of processing, etc. etc.

  • Which one of these changes?

  • Measure directly with S.A.T. response-signal method.
Locality in comprehension

which driver... fainted
Locality in comprehension

which driver fainted

which driver did Sue say fainted
Local is easier

**ACCEPTABILITY RATINGS**
Sprouse et al. ‘10

lower rating
longer dependency
Local is easier

**ACCEPTABILITY RATINGS**
Sprouse et al. ‘10

![Graph showing z-score rating over time for non-island and island conditions.](image)

**ERP TIMECOURSE**
Phillips et al. ‘05

![ERP timecourse showing activity over time.](image)

longer is later
Local is easier

**ACCEPTABILITY RATINGS**
Sprouse et al. ‘10

**ERP TIMECOURSE**
Phillips et al. ‘05

**SAT ASYMPTOTIC ACCURACY**
McElree et al. ‘03

less likely to succeed
**Local is easier**

Gibson (2000)
Dependency Locality Theory

Fodor (1978)
Gap-finding is hard

Frazier (1987)

Wanner & Maratsos (1978)
Storing incomplete dependencies is hard
Longer ≠ harder

Konieczny (2000)
German RC preverbal intraposition

Vasishth & Lewis (2006)
Hindi center embedding & object relative clauses

RC-modified subjects
Causes of facilitation

INTRODUCTION

Vasishth & Lewis (2006)

Mechanism of facilitation

RT ~ X

Why?
The officer was informed that the driver fainted/*drained.

+Adverb
... the driver abruptly fainted

+PP
... the driver of the ambulance fainted

+Subject Relative Clause
... the driver who wrecked the ambulance fainted

+Object Relative Clause
... the driver who the ambulance hit fainted
Speed-accuracy tradeoff

INTRODUCTION

Sentence (250 ms/word)

Fixation Point (500 ms)

Response Cue

Variable Processing Time

<<tone>>

admired

had

soldier

the

believed

general

the

that

report

the

was

It

+
Theoretical outcomes

Accuracy difference

Rate difference
Wagers & McElree (2009) actual data
Wagers & McElree (2009) actual data
Wagers & McElree (2009) summary

- Facilitation only observed in the +ADVERB conditions
- However, RCs were simple
- Give anti-locality a better chance by extending the RCs
- Follow the Jaeger et al. materials design
Exp. 1: Materials and Methods

- **ADJACENT**
  Wounded by the rebel in the trenches near the border, *the soldier twitched/snagged.*

- **+Object Relative Clause/NoPP**
  In the trenches near the border, *the soldier that the rebel wounded twitched/snagged.*

- **+Object Relative Clause/+1PP**
  In the trenches, *the soldier that the rebel wounded near the border twitched/snagged.*

- **+Object Relative Clause/+2PP**
  *The soldier that the rebel wounded in the trenches near the border twitched/snagged.*
Materials and Methods

• Acceptability × Length

  ADJACENT
  +OBJECT RC (ORC.noPP)
  +OBJECT RC/1PP (ORC.2PP)
  +OBJECT RC/2PP (ORC.1PP)

• 36 item sets

• MR-SAT
  • n = 10, compensated
  • Fillers with sentence-medial errors
  • Three sessions + with a practice session
  • Liu et al. (2009): Competitive model analysis

\[ d' = \lambda \cdot \left(1 - e^{-\beta \cdot (t-\delta)}\right) \]
Results: Object Relative Clauses

[Graph showing the relationship between lag latency (ms) and discriminability (d-prime) for different conditions: ORC.noPP, ORC.PP1, ORC.PP2.]

Results: Object Relative Clauses

![Graph showing results of Object Relative Clauses. The graph plots discriminability (d-prime) against lag latency (ms). The x-axis represents lag latency in milliseconds, ranging from 0 to 6000. The y-axis represents discriminability, ranging from 0 to 3. The graph includes lines and data points for different conditions: ORC.noPP, ORC.PP1, and ORC.PP2.](image-url)
Results: Object Relative Clauses
## Results: Best-fit parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>noPP</th>
<th>PP+1</th>
<th>PP+2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asymptote $\lambda$</td>
<td>3.37</td>
<td>3.49</td>
<td>3.54</td>
</tr>
<tr>
<td>Rate $\beta$ (sec$^{-1}$)</td>
<td></td>
<td>0.746</td>
<td></td>
</tr>
<tr>
<td>Intercept $\delta$ (sec)</td>
<td></td>
<td>0.724</td>
<td></td>
</tr>
</tbody>
</table>

2064 ms

$R^2 = 0.98$
## Results: Best-fit parameters

<table>
<thead>
<tr>
<th></th>
<th>noPP</th>
<th>PP+1</th>
<th>PP+2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Asymptote</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda$</td>
<td>3.37</td>
<td>3.49</td>
<td>3.54</td>
</tr>
<tr>
<td><strong>Rate</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta$ (sec$^{-1}$)</td>
<td>0.746</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Intercept</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\delta$ (sec)</td>
<td>0.724</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2064 ms

$R^2 = 0.98$
Results: Adjacent v. ORC

![Graph showing discriminability (d-prime) vs. lag latency (ms) for Adjacent and ORC.noPP conditions.](image)
Results: Adjacent v. ORC
Summary

• For RC-modified subjects, there was no rate effect of adding more PPs

• There was an increase in asymptotic accuracy as more PPs were added:
  \[
  3.37 \text{ d'} < 3.49 \text{ d'} < 3.54 \text{ d'}
  \]
  consistent across subjects

• **Concern**: overall processing was slow and adjacent subject-verb dependencies were slowest of all
Experiment 2

• The preposed XPs that controlled for ordinal position may have significantly dampened/swamped the subject-verb relevant processing

  • XPs were attachable to either subject or verb

  • … and sometimes ambiguously

• Experiment 2 uses local environments that are identical to Experiment 1, but with an unambiguous embedding context to control for ordinal position
Materials and Methods

• **ADJACENT**
The medic who was tending wounds in the trenches near the border observed that

the soldier twitched/*snagged.

• **+Object Relative Clause/NoPP**
The medic in the trenches near the border observed that

the soldier that the rebel wounded twitched/*snagged.

• **+Object Relative Clause/+2PP**
The medic observed that

the soldier that the rebel wounded in the trenches near the border twitched/*snagged.

• **+ADVERB**
The medic who tended wounds in the trenches near the border observed that

the soldier slightly twitched/*snagged.
Materials and Methods

- Acceptability × Length

  **ADJACENT**

  + OBJECT RC (ORC.noPP)

  + OBJECT RC/2PP (ORC.2PP)

  + ADVERB

- 36 item sets
- Fillers identical to experiment 1
- MR-SAT
  - n = 10, course credit for a Semantics course
  - Five sessions + 1 practice session
Results: all data

[Graph showing discriminability (d-prime) vs. lag latency (ms) with different conditions represented by distinct markers and colors.]
Results: Adjacent, +Adverb, +ORC

![Graph showing the relationship between discriminability (d-prime) and lag latency (ms). The graph includes data points for Adjacent, +Adv, and +ORC.noPP conditions.]
Results: Best-fit curve
**Results: Best-fit curve**

- $+\text{Adv} > \text{Adjacent} >> +\text{RC}$

Wagers & McElree (2009)
## Results: Best-fit parameters

<table>
<thead>
<tr>
<th></th>
<th>Adjacent</th>
<th>+ADV</th>
<th>+ORC noPP</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Asymptote</strong> $\lambda$</td>
<td>2.75</td>
<td>2.80</td>
<td>2.76</td>
</tr>
<tr>
<td><strong>Rate</strong> $\beta$ (sec$^{-1}$)</td>
<td>1.69</td>
<td>1.74</td>
<td>1.48</td>
</tr>
<tr>
<td><strong>Intercept</strong> $\delta$ (sec)</td>
<td></td>
<td>0.466</td>
<td></td>
</tr>
</tbody>
</table>
## Results: Best-fit parameters

<table>
<thead>
<tr>
<th></th>
<th>Adjacent</th>
<th>+ADV</th>
<th>+ORC noPP</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Asymptote</strong></td>
<td>2.75</td>
<td>2.80</td>
<td>2.76</td>
</tr>
<tr>
<td>$\lambda$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Rate</strong></td>
<td>1.69</td>
<td>1.74</td>
<td>1.48</td>
</tr>
<tr>
<td>$\beta$ (sec$^{-1}$)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Intercept</strong></td>
<td></td>
<td>0.466</td>
<td></td>
</tr>
<tr>
<td>$\delta$ (sec)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Average ‘speed’: $1/\beta + \delta$
### Results: Best-fit parameters

<table>
<thead>
<tr>
<th></th>
<th>Adjacent</th>
<th>+ADV</th>
<th>+ORC noPP</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Asymptote (\lambda)</strong></td>
<td>2.75</td>
<td>2.80</td>
<td>2.76</td>
</tr>
<tr>
<td><strong>Rate (\beta) (sec(^{-1}))</strong></td>
<td>1.69</td>
<td>1.74</td>
<td>1.48</td>
</tr>
<tr>
<td><strong>Intercept (\delta) (sec)</strong></td>
<td></td>
<td>0.466</td>
<td></td>
</tr>
</tbody>
</table>

1056 ms 1040 ms 1142 ms

\(+Adv > Adjacent >> +RC\)

Wagers & McElree (2009)
Results: Adjacent v. ORC.PP2

[Graph showing discriminability (d-prime) vs. lag latency (ms) with data points for Adjacent and ORC.PP2 conditions.]
Results: Adjacent v. ORC.PP2

The graph shows the discriminability (d-prime) as a function of lag latency (ms) for two conditions: Adjacent and +ORC.PP2. The +ORC.PP2 condition is indicated by green circles, while the Adjacent condition is indicated by black circles. The graph indicates that the +ORC.PP2 condition is faster than the Adjacent condition.
Results: Adjacent v. ORC.PP2

The graph shows the discriminability (d-prime) as a function of lag latency (ms) for two conditions: Adjacent and +ORC.PP2. The +ORC.PP2 condition is indicated to be slower than the Adjacent condition.
## Results: Best-fit parameters

<table>
<thead>
<tr>
<th></th>
<th>Adjacent</th>
<th>+ADV</th>
<th>+ORC noPP</th>
<th>+ORC PP2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Asymptote ( \lambda )</strong></td>
<td>2.75</td>
<td>2.80</td>
<td>2.76</td>
<td>2.58</td>
</tr>
<tr>
<td><strong>Rate ( \beta ) (sec(^{-1}))</strong></td>
<td>1.69</td>
<td>1.74</td>
<td>1.48</td>
<td>3.75</td>
</tr>
<tr>
<td><strong>Intercept ( \delta ) (sec)</strong></td>
<td>0.466</td>
<td></td>
<td>0.673</td>
<td></td>
</tr>
<tr>
<td><strong>Intercept</strong></td>
<td>1056 ms</td>
<td>1040 ms</td>
<td>1142 ms</td>
<td>939 ms</td>
</tr>
</tbody>
</table>

RC.PP2 >> +Adv > Adjacent >> +RC.noPP
## Results: Best-fit parameters

<table>
<thead>
<tr>
<th></th>
<th>Adjacent</th>
<th>+ADV</th>
<th>+ORC noPP</th>
<th>+ORC PP2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asymptote $\lambda$</td>
<td>2.75</td>
<td>2.80</td>
<td>2.76</td>
<td>2.58</td>
</tr>
<tr>
<td>Rate $\beta$ (sec$^{-1}$)</td>
<td>1.69</td>
<td>1.74</td>
<td>1.48</td>
<td>&gt; 3.75</td>
</tr>
<tr>
<td>Intercept $\delta$ (sec)</td>
<td>0.466</td>
<td></td>
<td></td>
<td>&lt; 0.673</td>
</tr>
<tr>
<td>Time (ms)</td>
<td>1056</td>
<td>1040</td>
<td>1142</td>
<td>939</td>
</tr>
</tbody>
</table>

RC.PP2 >> +Adv > Adjacent >> +RC.noPP
Results: Model comparison

Single intercept model

Dual intercept model

lag latency (ms)

discriminability (d-prime)
Results: Graphical model comparison

- Single intercept model
- Dual intercept model

![Graph showing comparison between single and dual intercept models](attachment:image.png)
### Results: Model comparison quantified

<table>
<thead>
<tr>
<th>Model Type</th>
<th>Adjusted R-squared</th>
<th>Deviance</th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>4-4-1 Single intercept</td>
<td>0.9917</td>
<td>-147.41</td>
<td>-129.41</td>
<td>-109.44</td>
</tr>
<tr>
<td>4-4-2 Dual intercept</td>
<td>0.9938</td>
<td>-156.47</td>
<td>-136.47</td>
<td>-114.28</td>
</tr>
<tr>
<td>4-4-4 Saturated intercept</td>
<td>0.9938</td>
<td>-157.30</td>
<td>-133.31</td>
<td>-106.68</td>
</tr>
</tbody>
</table>

\[ G^2(1) = 9.1, p < .005 \]
\[ G^2(2) = 0.8, n.s. \]

Consistent parameter ranking across participants (p < .05)
Results: Across participants

ORC_PP2- Adj

Intercept 1/Rate
δ (sec) 1/β (sec)
Results: Summary

• When the V-dependencies are strictly local, we observe two distinct effects on speed for the strong anti-local context:

  • An intercept shift  
    Discriminating information is available much later  

  • A rate increase  
    Information is accrued much faster

• Overall:  
  a facilitation in speed
Relation to reaction times

![Graph showing the relationship between discriminability (d-prime) and lag latency (ms) for Adjacent and ORC_PP2 conditions.](image)
Conclusion

- Strongly anti-local S-V relationship formation is associated with **faster dynamics**
  - Consistent with predictive accounts (sloughing, working ahead)
  - Less consistent with memory-strength accounts

- However, it associated with a **cost**: discriminative information is **available later**

- **Facilitation** obtains on balance for modest-to-high accuracy processing
What is the cost?

- **Focus of attention costs**
  - 85 ms / McElree et al. (2003)
  - 87 ms, 74 ms / Wagers & McElree (2009)
  - 83 ms / this study [ORC.noPP-Adjacent]

- **Intercept cost:**
  - 207 ms ( +44% )
What is the cost?

- Previous studies:
  - ... $[\text{VP}]_{S} [\text{NP}]_{NP} \rightarrow [\text{NP V}]_{S}$

- Current study
  - ... $[\text{NP}]_{PP-2} [\text{PP-1}]_{VP} [\text{VP}]_{S} [\text{NP}]_{NP} \rightarrow [\text{NP V}]_{S}$
  - ... $[\text{NP}]_{PP-2} [\text{PP-1}] \rightarrow \ldots [\text{PP-1}]_{VP} [\text{VP}]_{S} [\text{NP}]_{NP} \rightarrow [\text{NP V}]_{S}$

- Relating the scope of focal attention with the chunking of syntactic category
A diversity of timing measures

- RTs masked two underlying effects
  - Speed and accuracy tradeoffs are not predictable

- Dillon et al., Thurs, Binding ziji
  faster rate was associated with lower accuracy

- Staub, Fri, frequency & predictability in fixation times
  RT distribution modeling
Collaborators and acknowledgments

- Sarah Napoli  
  (UCSC Linguistics)

- Shayne Sloggett, Pranav Anand and  
  LING116 members

- CUNY reviewers

- Office of the Dean of Humanities, UCSC,  
  and UCSC Academic Senate Committee on Research

Thank you.
Appendices
### Results: Best-fit parameters

<table>
<thead>
<tr>
<th></th>
<th>Adjacent</th>
<th>+ADV</th>
<th>+ORC noPP</th>
<th>+ORC PP2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asymptote $\lambda$</td>
<td>2.75</td>
<td>2.80</td>
<td>2.76</td>
<td>2.58</td>
</tr>
<tr>
<td>Rate $\beta$ (sec$^{-1}$)</td>
<td>1.69</td>
<td>1.74</td>
<td>1.48</td>
<td>3.75</td>
</tr>
<tr>
<td>Intercept $\delta$ (sec)</td>
<td>0.466</td>
<td></td>
<td>0.673</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1056 ms</td>
<td>1040 ms</td>
<td>1142 ms</td>
<td>940 ms</td>
</tr>
</tbody>
</table>

`RC.PP2 >> +Adv > Adjacent >> +RC.noPP`
## Results: 4-4-4 parameters

<table>
<thead>
<tr>
<th></th>
<th>Adjacent</th>
<th>+ADV</th>
<th>+ORC noPP</th>
<th>+ORC PP2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asymptote $\lambda$</td>
<td>2.75</td>
<td>2.80</td>
<td>2.76</td>
<td>2.57</td>
</tr>
<tr>
<td>Rate $\beta$ (sec$^{-1}$)</td>
<td>1.63</td>
<td>1.77</td>
<td>1.51</td>
<td>3.76</td>
</tr>
<tr>
<td>Intercept $\delta$ (sec)</td>
<td>0.467</td>
<td>0.476</td>
<td>0.476</td>
<td>0.673</td>
</tr>
<tr>
<td></td>
<td>1057 ms</td>
<td>1039 ms</td>
<td>1140 ms</td>
<td>939 ms</td>
</tr>
</tbody>
</table>

RC.PP2 >> +Adv > Adjacent >> +RC.noPP
## Results: 4-4-1 parameters

<table>
<thead>
<tr>
<th></th>
<th>Adjacent</th>
<th>+ADV</th>
<th>+ORC noPP</th>
<th>+ORC PP2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Asymptote ( \lambda )</strong></td>
<td>2.75</td>
<td>2.80</td>
<td>2.76</td>
<td>2.58</td>
</tr>
<tr>
<td><strong>Rate ( \beta (\text{sec}^{-1}) )</strong></td>
<td>1.68</td>
<td>1.72</td>
<td>1.46</td>
<td>1.77</td>
</tr>
<tr>
<td><strong>Intercept ( \delta (\text{sec}) )</strong></td>
<td>0.460</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Timing</strong></td>
<td>1056 ms</td>
<td>1042 ms</td>
<td>1143 ms</td>
<td>1023 ms</td>
</tr>
</tbody>
</table>

RC.PP2 >> +Adv > Adjacent >> +RC.noPP
Results: Across participants

Consistent parameter ranking across participants (p < .05)
Spill-over? ... lag latency plots

![Bar charts showing RT (ms) across different lags](image)
Results: Graphical model comparison

- Single intercept model
- Dual intercept model

![Graph showing comparison between Single intercept model and Dual intercept model]
Wagers & McElree (2009) Adverb specificity

**Abruptly**

**Evidently**
Exponential equation

\[ d' = \lambda \cdot \left(1 - e^{-\beta \cdot (t-\delta)}\right) \]
Accuracy ~ session