Structuring Expectation

Matt Wagers
University of California, Santa Cruz

WCCFL 33
Simon Fraser University, Vancouver
March 27-29, 2015
Modes of explanation

• How do we negotiate the incremental uncertainty of language?

• Explanations for what is easy and what is difficult in language processing come in three (overlapping) flavors

  • properties of the representation/representation-building
  • what’s likely
  • what’s useful in context
What is likely?

• Language processing is probabilistic in at least the sense that, for some unfolding expression, we have an ordering of possible analyses based on likelihood or confidence.

• Difficulties occur in language processing when incoming information dramatically shifts our pre-existing allocation of confidence or likelihood (Entropy Reduction, Hale, 2001; Surprisal, Levy, 2008).

  • The horse raced past the barn … \( S \to NP \ VP \)

  • The horse raced past the barn fell. \( S_1 \to NP \ VP; NP \to NP \ S_2 \)

• Very general, approximate and useful formalization

  • conditional probability: \( P(w_{n+1}|w_0w_1\ldots w_n) \)
Three case studies

- Animacy and English relative clauses
- Agreement inside English DPs
- Wh-Agreement and Person in Chamorro
Three case studies

- Animacy and English relative clauses
- Agreement inside English DPs
- Wh-Agreement and Person in Chamorro
QUESTION

Why are object relative clauses easier to understand when the relativized argument is inanimate?
Relative clauses—sources of difficulty

- Generally speaking, subject relative clauses are easier to process and understand than object relative clauses.

- but … the SRC > ORC advantage can be neutralized under a variety of conditions.
  - for example, if the RC subject is a pronoun (Bever, 1974, Gordon et al. 2001)
  - or, if the relativized argument is inanimate (Mak et al. 2002, Traxler et al. 2002, Gennari & MacDonald, 2005, i.a.)
Relative clauses — role of animacy

- E.g. Traxler, Morris & Seely (2002):
  - **ANIMATE**
    The director [SRC that ___ watched the movie ]
    The director [ORC that the movie pleased ___ ]
  - **INANIMATE**
    The movie [SRC that ___ pleased the director ]
    The movie [ORC that the director watched ___ ]
HYPOTHESIS

_Animate_ RC heads are predictively linked to a subject gap.

_Inanimate_ heads are not.

(cf. Active Filler Strategy).
Relative clause expectations: filled gap design

The kindergarten teacher pointed out …

the friendly child who the young girl has played with __ incessantly.

the friendly child with whom the young girl has played __ incessantly.

The kindergarten teacher pointed out …

**the friendly child** who, as of yesterday, **the young girl** has played with incessantly.

**the friendly child** with whom, as of yesterday, **the young girl** has played incessantly.

Relative clause expectations: filled gap design

The kindergarten teacher pointed out …

**ANIMATE**

the friendly child who, as of yesterday, the young girl has played with incessantly.

the friendly child with whom, as of yesterday, the young girl has played incessantly.

**INANIMATE**

the colorful toy which, as of yesterday, the young girl has played with incessantly.

the colorful toy with which, as of yesterday, the young girl has played incessantly.

Animacy and the expectation for a subject gap

**ANIMATE**

**FILLED GAP EFFECT**

**INANIMATE**

Pendleton & Wagers (2014)
Animacy and the expectation for a subject gap

Replication attempt

Pendleton & Wagers (2014)
Relative clauses—role of animacy

• Filled-gap design results:

  • **ANIMATE** relative clause heads generate an expectation for a subject gap

  • **INANIMATE** relative clause heads do not

**HYPOTHESIS**

*Animate* RC heads are predictively linked to a subject gap.

*Inanimate* heads are not.
... how does animacy exert this influence?

- What expectations should individuals hold based on their language experience?

Roland, Dick & Elman (2007)

% subject gap

<table>
<thead>
<tr>
<th></th>
<th>Brown</th>
<th>Switchboard</th>
</tr>
</thead>
<tbody>
<tr>
<td>Animate</td>
<td>75%</td>
<td>91%</td>
</tr>
<tr>
<td>Inanimate</td>
<td>47%</td>
<td>31%</td>
</tr>
</tbody>
</table>
... how does animacy exert this influence?

- Replicated in COCA and Gigaword (parsed NYT subsection)
- ... but, what about that RC-initial adjunct?

Roland, Dick & Elman (2007)

<table>
<thead>
<tr>
<th></th>
<th>Brown</th>
<th>Switchboard</th>
</tr>
</thead>
<tbody>
<tr>
<td>Animate</td>
<td>75%</td>
<td>91%</td>
</tr>
<tr>
<td>Inanimate</td>
<td>47%</td>
<td>31%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Gigaword</th>
<th>NYT subsection</th>
</tr>
</thead>
</table>
| Animate        | 99%  
    n = 195 |               |
| Inanimate      | 94%  
    n = 449 |               |
... how does animacy exert this influence?

- Cloze task (Amazon Mechanical Turk) using actual experimental materials (n=400)

% subject gap completions

<table>
<thead>
<tr>
<th></th>
<th>no RC-initial adjunct</th>
<th>adjunct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Animate</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>Inanimate</td>
<td>88%</td>
<td></td>
</tr>
</tbody>
</table>
... how does animacy exert this influence?

- **Mediated view**
  - \( P(w_{n+1} = \text{“the”} | w_0w_1 \ldots w_n) \)
  - Pre-RC adjunct \( \Rightarrow \) ORC analysis very unlikely

- **Direct view, I**
  - \( P(\text{gap:SUBJ} | \text{hd:}+\text{ANIM}) \gg P(\text{gap:OBJ} | \text{hd:}+\text{ANIM}) \)
  - \( P(\text{gap:OBJ} | \text{hd:}–\text{ANIM}) \gg P(\text{gap:SUBJ} | \text{hd:}–\text{ANIM}) \)
... how does animacy exert this influence?

• **Direct view, I**
  
  • $P(\text{gap:SUBJ} \mid \text{hd:}+\text{ANIM}) \gg P(\text{gap:OBJ} \mid \text{hd:}+\text{ANIM})$
  
  • $P(\text{gap:OBJ} \mid \text{hd:-ANIM}) \gg P(\text{gap:SUBJ} \mid \text{hd:-ANIM})$

• **Direct view, II**
  
  • $U(\text{gap:SUBJ} \mid \text{hd:}+\text{Anim}) \gg U(\text{gap:OBJ} \mid \text{hd:}+\text{Anim})$
  
  • $U(\text{gap:OBJ} \mid \text{hd:-Anim}) \gg U(\text{gap:SUBJ} \mid \text{hd:-Anim})$
... how does animacy exert this influence?

- **Mediated view:**
  predictions reflect any (known) contingencies

- **Direct view, I:**
  predictions reflect contingencies on grammatically-active features (or some distinguished set)

- **Direct view, II:**
  predictions optimize well-formedness
Three case studies

- Animacy and English relative clauses
- Agreement inside English DPs
- Wh-Agreement and Person in Chamorro
QUESTION

Why are we susceptible to (erroneously) disagree with singular DPs but not with plural DPs?
Agreement attraction

- The **path** to the **monuments** is/?are littered with bottles.
- The **paths** to the **monument** are/*is littered …

Based on Eberhard, Cutting & Bock (2005)
**HYPOTHESIS**

*Plural* features are maintained in working memory more durably than *singular* features (e.g., because they are marked).
Bi-partite model of memory

• Two basic memory states:
  • **active/focal** - stringent capacity limitations - fast processing
  • **passive** - virtually unlimited - requires retrieval/slower

**Direct evidence**
Broadbent 1958
Wickelgren et al., 1980
Garavan, 1998; Cowan, 2001
McElree, 2006
Verhaegen & Basak, 2007
Jonides et al., 2008

**Architecture**
ACT-R: Lewis & Vasishth, 2005
Full/reduced representations Hinton, 1990

**Similarity-based retrieval interference effects**
Gordon et al., 2001, et seq
Drenhaus et al., 2008
Badecker & Kuminiak, 2007
Wagers, Lau & Phillips, 2009
HYPOTHESIS v.2

*Plural* features are more likely than *singular* features to be maintained in the active (fast) state of memory.
Testing for the maintenance of [PL]

PL demonstrative DP

those monkeys (*monkey)
those face-making monkeys (*monkey)
those mischievous, face-making monkeys (*monkey)

SG demonstrative DP

that monkey (*monkeys)
that face-making monkey (*monkeys)
that mischievous, face-making monkey (*monkeys)
Testing for the maintenance of [PL]

Method

Speed-accuracy tradeoff
(multiple-response)

• response-signal technique
  Wickelgren et al. 1980, McElree et al. 2003

• sensitive measure of retrieval speed
  McElree, 2006
Fig. 3 presents hypothetical SAT functions illustrating how different SAT timecourse patterns can discriminate between alternative retrieval processes. Consider first the expected result that interpolating more material between the filler and gap position decreases the accuracy of responding. Recall that this could be because there is a lower probability that a representation of the filler is available when the verb is processed and/or because there is a higher probability of misanalyzing material up to and including the final verb. If additional material decreases only the overall accuracy of responding, the corresponding functions will differ in asymptotic level alone. Panel A depicts two hypothetical conditions that differ in this manner.

The pre-asymptotic portion of the SAT function measures processing speed or dynamics, jointly specified by the intercept of the function (when accuracy departs from chance, \(d_0 = 0\)) and the rate at which accuracy grows from intercept to asymptote. The intercept measures the minimum time needed to form an interpretation that would serve to discriminate acceptable from unacceptable forms. The rate of the SAT function reflects either the rate of continuous information accrual if processing is continuous or the distribution of finishing times if processing is discrete or quantal (Dosher, 1976, 1979, 1981, 1984; Meyer, Irwin, Osman, & Kounois, 1988). In either case, differences in intercept or rate implicate underlying differences in the speed of processing. This situation is depicted in Panel B of Fig. 3, where the functions are associated with different intercepts and rates of rise to a common asymptote.

If access to the filler's representation requires a search process when the matrix verb is encountered, then the SAT intercept and/or rate of will systematically slow as more material is interpolated between the filler and gap. Rate or intercept differences can arise from factors other than retrieval speed; for example, they might arise from differences in the inherent complexity of computing
Testing for the maintenance of [PL]

Analysis

Asymptotic difference
Reflects the likelihood of completing a parse/process.

Rate difference
Reflects speed of processing how quickly information accumulates continuously, or the differences in an underlying discrete finishing time distribution.
Testing for the maintenance of [PL]

Results

Computing non-adjacent agreement versus adjacent agreement

Singular: +92 ms

Plural: no longer

**HYPOTHESIS**

*Plural* features are more likely than *singular* features to be maintained in the active (fast) state of memory.

Wagers & McElree (2012)
... how does number exert its influence?

- What expectations should we hold based on language experience?

Occurrence of N head as the next element in DP is always less surprising for singular DPs

Wagers & McElree (2012)
... how does number exert its influence?

- **Mediated**: if predictions reflect (known) contingencies for DPs, the number-bearing N head is always expected sooner when $\text{Dem}=\text{SG}$

- **Direct**: if predictions stem from properties of feature structure, the marked PL feature is able to persist in working memory (= facilitating longer DPs?)
Three case studies

Animacy and English relative clauses

Agreement inside English DPs

Wh-Agreement and Person in Chamorro
QUESTION

How does obligatoriness and optionality in agreement paradigms affect interpretation of cross-indexed dependencies?
Chamorro: Wh-Agreement

(1) Ha fåhan si Vicente i gima’ Antonio.
   “Vicente bought Antonio’s house.”

(2) Håyi fumåhan ___ i gima’?
   “Who bought the house?”
(3) **Håfa ha fåhan** si Maria ___ gi tenda?
what? AGR buy NM Maria LOC store

“What did Maria buy at the store?”

(4) **Håfa finåhan-ña** si Maria ___ gi tenda?
what? WH[OBJ].buy-AGR NM Maria LOC store
HYPOTHESIS

Object Wh-Agreement will trigger faster interpretation of an object gap dependency compared to ordinary S-V agreement.
Chamorro: Wh-Agreement

Kuåntu na chinina  prinensåm-mu nigap …
how many L shirt iron [WH:OBJ] - AGR yesterday
“How many shirts did you iron yesterday…?”

Kuåntu na patgun låhi  prinensåm-mu nigap …
how many boys iron [WH:OBJ] - AGR yesterday
“How many boys did you iron yesterday…?”

Kuåntu na chinina/patgun låhi un prensa nigap …
AGR iron
Transitive clauses SELF-PACED LISTENING

**ANOMALY EFFECT** $p < .005$

"How many X did you iron yesterday afternoon?"

"How many X’s sleeves did you iron yesterday?"

**INTERACTION** $p = .05$

Wagers, Borja & Chung (2015)
HYPOTHESIS

Object Wh-Agreement will trigger faster interpretation of an object gap dependency compared to ordinary S-V agreement.
... how does Wh-Agreement exert its influence?

- Why is the anomaly contrast evident earlier for Wh-Agreement?

- **Presence** of Wh-Agreement *promotes* dependency construction because it must be licensed

- **Absence** of Wh-Agreement *restrains* dependency construction because it is compatible with other continuations
Informativity of the bare form

- **Absence** of Wh-Agreement *restrains* dependency construction because it is compatible with other continuations

  - **Possessor extraction** → no Wh-Agreement

    Hāyi un lâksi chininå-ña?
    who? 2SG sew shirt-AGR
    “Whose shirt did you sew?”
Informativity of the bare form

- Odds in favor of possessor gap:
  \[
  \frac{P(\text{Poss} | \text{Verb})}{P(\text{Obj} | \text{Verb})} = \frac{P(\text{Verb} | \text{Poss}) \times P(\text{Poss})}{P(\text{Verb} | \text{Obj}) \times P(\text{Obj})}
  \]

- Preference survey (n=13)
  - Object extraction: bare v. Wh-Agreeing 72%
  - Possessor: bare v. periphrastic form 74%

- Odds in favor of possessor extraction given a bare verb? ~ Prior odds of a possessor extraction

Wagers, Borja & Chung (2015)
Prior odds of a possessor extraction

- Probably pretty low? especially for transitive objects

- In a recent production study we elicited (at least) 691 gap-containing relative clauses:

<table>
<thead>
<tr>
<th>Argument</th>
<th>Transitive Subject</th>
<th>Object</th>
<th>Intransitive Subject</th>
<th>Locative</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>210</td>
<td>10</td>
<td>308</td>
<td>106</td>
</tr>
<tr>
<td>Possessor</td>
<td>0</td>
<td>1</td>
<td>16</td>
<td>40</td>
</tr>
</tbody>
</table>

For RCs and objects, prior odds are 1:10 (but, small n)
Intransitive clauses SELF-PACED LISTENING

SMALL ANOMALY EFFECT  \( p < .05 \)

NO INTERACTION

Wagers, Borja & Chung (2015)
Informativity of the bare form

- Absence of Wh-Agreement restrains dependency construction because it is compatible with other continuations

- Possessor extraction → no Wh-Agreement

Håyi un lâksi chininå-ña?
who? 2SG sew shirt-AGR
“Whose shirt did you sew?”
Informativity of the bare form

- Absence of Wh-Agreement restrains dependency construction because it is compatible with other continuations
  
  - Possessor extraction $\rightarrow$ no Wh-Agreement
    
    Håyi un låksi chininå-ña?
    
    who? 2SG sew shirt-AGR
    
    “Whose shirt did you sew?”

  - No possessor extraction over 3.PERS DP
    *Håyi ha låksi si Bedu’ chininå-ña?
    
    who? 3SG sew Bedu’ shirt-AGR
    
    (“Whose shirt did Bedu’ sew?”)
Alignment hierarchies and prediction

- *Subj=3.pers > Obj=2
  *Ha li’i hao si Dolores nigap
  3SG see 2SG UNM Dolores yesterday
  ‘Dolores saw you yesterday’

- *Subj=DP > Obj=3.pers anim. pron.
  *Ha li’i’ gui’ si Maria
  3SG see 3SG UNM Maria
  ‘Maria saw him’

- The Chamorro Person-Animacy Hierarchy
  2.pers > 3.pers anim. pron. > anim. non-pron. > inanimate

cf. Aissen, 1997;
Christianson & Ferreira, 2005, Christianson & Cho, 2009
Compare effect of \texttt{+WH.AGR} to \texttt{3.PERS}

\textbf{3PERS, \texttt{-WH.AGR}}

\textit{Bula katpinteru [ ha apåşi si Bedu’ \_ gi lanchu gi ma’pus na mes ].}

There were a lot of carpenters who Bedu’ paid at the ranch last month.

\textbf{3PERS, \texttt{+WH.AGR}}

\textit{Bula katpinteru [ inapasi-ñả si Bedu’ \_ gi lanchu gi ma’pus na mes ].}

There were a lot of carpenters who Bedu’ paid at the ranch last month.
Compare effect of +WH.AGR to 3.PERS

PLAUSIBLE EXTRACTIONS

3PERS, -WH.AGR

Bula katpinteru ha apåsi si Bedu’ gi lanchu gi ma’pus na mes.
Bedu’ paid a lot of carpenters at the ranch last month.

3PERS, +WH.AGR

Bula katpinteru inapasi-ña si Bedu’ gi lanchu gi ma’pus na mes.
Bedu’ paid a lot of carpenters at the ranch last month.

2PERS, -WH.AGR

Bula katpinteru un apåsi gi lanchu gi ma’pus na mes.
You paid a lot of carpenters at the ranch last month.
Compare effect of +WH.AGR to 3.PERS

**IMPLAUSIBLE EXTRACTIONS**

3PERS, -WH.AGR

*Bula katpinteru ha dingding si Bedu’* gi lanchu gi ma’pus na mes.
Bedu’ rang a lot of carpenters at the ranch last month.

3PERS, +WH.AGR

*Bula katpinteru diningdeng-ña si Bedu’* gi lanchu gi ma’pus na mes.
Bedu’ rang a lot of carpenters at the ranch last month.

2PERS, -WH.AGR

*Bula katpinteru un dingding* gi lanchu gi ma’pus na mes.
You rang a lot of carpenters at the ranch last month.
Compare effect of +WH.AGR to 3.PERS

- **2. PERS/-WH**:
  - Listening time (-duration) (ms)
  - un apåsi (you paid)
  - un dingding (you rang)

- **3. PERS/+WH**:
  - inapasi-ña si Bedu’ (Bedu’ paid)
  - diningdeng-ña si Bedu’ (Bedu’ rang)

- **3. PERS/-WH**:
  - ha apåsi si Bedu’ (Bedu’ paid)
  - ha dingding si Bedu’ (Bedu’ rang)

- **ANOMALY EFFECT**
  - 3.PERS and +WH.AGR act similarly!

- **n.s.** (not significant)

Wagers, Borja & Chung (*in prep*)
On-line anomaly effect

<table>
<thead>
<tr>
<th></th>
<th>2.PERS</th>
<th>3.PERS</th>
</tr>
</thead>
<tbody>
<tr>
<td>+WH.AGR</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>-WH.AGR</td>
<td>✗ ✗</td>
<td>✓</td>
</tr>
</tbody>
</table>

Wagers, Borja & Chung (in prep)
... how does Wh-Agreement exert its influence?

- **Mediated**: if predictions reflect (known) contingencies, the chance of an object gap is essentially maximal when Wh-Agreement is present - just like when a 3.PERS DP subject is present.

- **Direct**: if predictions stem from grammatical licensing requirements
  - Object Wh-Agreement predictively extends the representation to include an object gap, triggering interpretation.
  - 3.PERS DP subjects require “justification” from PERS-ANIM constraints → gap must be a VP-internal argument.
<table>
<thead>
<tr>
<th></th>
<th>Finding</th>
<th>Direct Mechanism</th>
</tr>
</thead>
<tbody>
<tr>
<td>RC</td>
<td>Animate RC heads are prospectively linked to subject gap, but inanimate heads are not.</td>
<td>Licensing of animacy?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Online “harmonic alignment”?</td>
</tr>
<tr>
<td>Agr</td>
<td>Faster processing of complex PL demonstrative DPs.</td>
<td>PL is more durable in memory.</td>
</tr>
<tr>
<td>Wh</td>
<td>Wh-Agreement triggers early interpretation of movement dependencies.</td>
<td>Wh-Agreement prospectively agrees with object gap.</td>
</tr>
</tbody>
</table>
# Taking stock

<table>
<thead>
<tr>
<th>Finding</th>
<th>Mediated Mechanism</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>RC</strong></td>
<td>Animate RC heads are prospectively linked to subject gap, but inanimate heads are not.</td>
</tr>
<tr>
<td><strong>Agr</strong></td>
<td>Faster processing of complex PL demonstrative DPs.</td>
</tr>
<tr>
<td><strong>Wh</strong></td>
<td>Wh-Agreement triggers early interpretation of movement dependencies.</td>
</tr>
</tbody>
</table>
Managing our expectations

- Comprehenders exhibit misalignments between predicted dependency elements - as demonstrated in online experiments - and the most probable continuation - as estimated from a corpus, or a Cloze task

- Independent mechanisms affect ordering of predictions -
  - the *value* of one representation over another: which satisfies more constraints? or is incrementally more grammatical? (Pritchett, 1992, Chater, Crocker, Pickering, 1998, Borja, Wagers, & Chung 2015)
  - workspace constraints: how features/constituents are maintained in short-term memory (Wagers & McElree, 2013)
Ignorance is bliss

- Hard to tell whether comprehenders are not using contingencies, or are or ignorant of them.

- Ignorance is not a bad thing.
  - Syntactic probabilities are not universal or absolute - e.g., genre/context dependence. More abstract generalizations may be more stable than “construction-specific” ones (Roland et al. 2007)

- What’s stored is what drove learning and generalization?
  - input/intake distinction of Gagliardi & Lidz (2014)
Thank you

Emily Pendleton
UCSC B.A. ‘13

Manuel F. Borja
Inetnun Åmut yan
Kutturan Natibu, CNMI

Brian McElree
NYU Psychology

Sandra Chung
UCSC Linguistics

National Science Foundation BCS #1251429 (MW, SC)