Certain ungrammaticality or uncertain grammaticality: Deciding between frequent errors and infrequent grammatical structures

Maayan Keshev & Aya Meltzer-Asscher, Tel-Aviv University
Introduction
How do humans recover the intended meaning from noisy utterances?

Sentence processing models typically assume error-free input, but production/perception errors are common in language use.

Recent studies suggest that comprehends engage in (Bayesian) rational noisy-channel processing - interpretation is pulled towards more probable "near-neighbors", depending on:

- The likelihood of the alternative utterance given perceived utterance
- The prior probability of the different utterances

\[ P(\text{utterance} / \text{perceived input}) \propto P(\text{perceived input} / \text{utterance}) \cdot P(\text{utterance}) \]

Levy 2008; Levy et al., 2009; Gibson et al., 2013
Introduction
How do humans recover the intended meaning from noisy utterances?

Readers reevaluate spelling of an early word when it has a near neighbor that allows a structure with higher prior probability.

Levy and his colleagues (2009) exhibited that local coherence effects (slowdown at the ambiguous verb in A vs. B) disappear for prepositions of a sparse orthographic neighborhood (no difference in C vs. D).

A. The coach smiled at the player tossed the frisbee
B. The coach smiled at the player thrown the frisbee
C. The coach smiled toward the player tossed the frisbee
D. The coach smiled toward the player thrown the frisbee

[replace ‘at’ with: ‘as’/’and’?]
[no locally coherent string]
[no near neighbor]
[no coherent string/ neighbor]
The current study
How improbable should a structure be for readers to consider it noisy?

We test the effect of uncertainty and prior probability using Hebrew relative clauses with temporary ambiguity [object relative vs. subject with an agreement mismatch].

We show that the interpretation of agreement mismatch is modulated by the prior probabilities of the analyses.
The current study
The temporary SR/OR ambiguity in Hebrew

pagasnu et ha-studentim se-hikir
We-met acc. the-students.PL that-knew.SG
The current study
The temporary SR/OR ambiguity in Hebrew

\[ \text{pagasnu et ha-studentim se-hikir} \quad \text{ha-dikan} \]

We-met acc. the-students.PL that-knew.SG (nom.) the-Dean

OR + post-verbal subject (very rare)
The current study
The temporary SR/OR ambiguity in Hebrew

pagasnu et ha-studentim se-hikir  et ha-mazkira
We-met acc. the-students.pl that-knew.sg  acc. the-secretary

OR + post-verbal subject  (very rare)

Ungrammatical SR (verb-filler number mismatch)
The current study
The temporary SR/OR ambiguity in Hebrew

pagasnu et ha-studentim se-hikir
We-met acc. the-students.PL that-knew.SG

OR + post-verbal subject (very rare)

Ungrammatical SR (verb-filler number mismatch)

In this case of mismatch (FILLER.PL + VERB.SG), rational noisy-channel predicts that a SR will be formed. But – plural verbs allow another outlet…
The current study
The temporary SR/OR ambiguity in Hebrew (VERB.PL)

pagasnu et ha-student se-hikiru
We-met acc. the-student.SG that-knew.W.PL

OR + post-verbal subject (very rare)

Ungrammatical SR (verb-filler number mismatch)
The current study

The temporary SR/OR ambiguity in Hebrew (VERB.PL)

\[
pagasnu et \; ha\text{-}student \; se\text{-}hikiru \; oto\text{-}be\text{-}kol\; ha\text{-}xugim
\]
We-met acc. the-student.SG that-knew.W.PL him in-all departments

OR + post-verbal subject (very rare)

Ungrammatical SR (verb-filler number mismatch)

OR + impersonal null subject (common)
The current study
The temporary SR/OR ambiguity in Hebrew (VERB.PL)

pagasnu et ha-student se-hikiru oto be-kol ha-xugim
We-met acc. the-student.SG that-knewW.PL him in-all departments

OR + post-verbal subject (very rare)

Ungrammatical SR (verb-filler number mismatch)

OR + impersonal null subject (common)

Is an OR preferred in this mismatch (FILLER.SG + VERB.PL)?

Corrupted input, very probable structure

Pristine input, somewhat probable?

Pristine input, very improbable structure
The current study
Outline & Predictions

FILLER.PL+VERB.SG
[e.g. We met the students that knows]
OR + post-verbal subject (rare)
Ungrammatical SR
→ Form a SR

FILLER.SG+VERB.PL
[e.g. We met the student that know]
OR + post-verbal subject (rare)
Ungrammatical SR
OR + impersonal null subject
→ Form an OR?
The current study
Outline & Predictions

FILLER.PL+VERB.SG
[e.g. We met the students that knows]

OR + post-verbal subject (rare)
Ungrammatical SR
→ Form a SR

Experiments 1A&B: Does a post-verbal subject cause processing difficulty?

FILLER.SG+VERB.PL
[e.g. We met the student that know]

OR + post-verbal subject (rare)
Ungrammatical SR

OR + impersonal null subject
→ Form an OR?

Experiments 2A&B: Does a (post-verbal) object cause processing difficulty (filled-gap effect)?
The current study
Outline & Predictions

<table>
<thead>
<tr>
<th>FILLER.PL+VERB.SG</th>
<th>FILLER.SG+VERB.PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>[e.g. We met the students that knows]</td>
<td>[e.g. We met the student that know]</td>
</tr>
<tr>
<td><strong>OR + post-verbal subject (rare)</strong></td>
<td><strong>OR + post-verbal subject (rare)</strong></td>
</tr>
<tr>
<td><strong>Ungrammatical SR</strong></td>
<td><strong>Ungrammatical SR</strong></td>
</tr>
<tr>
<td>→ Form a SR</td>
<td>→ Form an OR?</td>
</tr>
</tbody>
</table>

Experiment 3: Sentence completion study
Experiment 1A

FILLER.PL + VERB.SG → Form a SR
Processing disruption at the subject?
SPR, 36 participants, 30 sets + 45 fillers (grammatical)
Experiment 1A

Materials

MATCH

We met the { student\textsubscript{SG} / students\textsubscript{PL} } that by the end of term decided\textsubscript{SG} eventually the principal to expel due to poor grades.

'B We met the students that, by the end of term, the principal decided to expel due to poor grades'.

MISMATCH

We met the { student\textsubscript{SG} / students\textsubscript{PL} } that by the end of term decided\textsubscript{SG} eventually the principal to expel due to poor grades.

BASELINE: We met because, by the end of term, decided\textsubscript{SG} eventually the principal to expel the student due to poor grades.

FILLER.PL + VERB.SG → Form a SR
Processing disruption at the subject?
SPR, 36 participants, 30 sets + 45 fillers (grammatical)
Experiment 1A

Results

Increased RTs at the post-verbal subject, in the **match** condition relative to the unambiguous baseline.

→ Readers constructed a SR
Experiment 1A

Results

The post-verbal subject reveals increased RTs also in the mismatch condition (relative to baseline).

→ Readers form a SR even when the verb mismatches the filler.
Experiment 1B
Materials

MATCH 1-MISMATCH 2-MISMATCH
We met the { student.SG-M / students.PL-M / students.PL-F } that by the end of term decided.SG-M eventually the principal to expel due to poor grades.

'We met the students that, by the end of term, the principal decided to expel due to poor grades'.

BASELINE: We met because, by the end of term, decided.SG eventually the principal to expel the student due to poor grades.
Experiment 1B

Results

Replication of the findings of Experiment 1a.

FILLER.PL + VERB.SG → Form a SR
Processing disruption at the subject?
SPR, 48 participants, 28 sets + 45 fillers (grammatical)
Experiment 1B
Results

Replication of the findings of Experiment 1a.

Extension to 2-feature mismatch.

FILLER.PL + VERB.SG $\rightarrow$ Form a SR
Processing disruption at the subject?
SPR, 48 participants, 28 sets + 45 fillers (grammatical)
Outline & Predictions

**FILLER.PL+VERB.SG**
[e.g. We met the students that knows]

OR + post-verbal subject (rare)

Ungrammatical SR

→ Form a SR

**FILLER.SG+VERB.PL**
[e.g. We met the student that know]

OR + post-verbal subject (rare)

Ungrammatical SR

OR + impersonal null subject

→ Form an OR?

**Experiments 1A&B:** Does a post-verbal subject cause a processing difficulty?

**Experiments 2A&B:** Does a (post-verbal) object cause a processing difficulty (filled-gap effect)?
# Outline & Predictions

<table>
<thead>
<tr>
<th>FILLER.PL+VERB.SG</th>
<th>FILLER.SG+VERB.PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>[e.g. We met the students that knows]</td>
<td>[e.g. We met the student that know]</td>
</tr>
<tr>
<td>OR + post-verbal subject (rare)</td>
<td>OR + post-verbal subject (rare)</td>
</tr>
<tr>
<td><strong>Ungrammatical SR</strong></td>
<td>Ungrammatical SR</td>
</tr>
<tr>
<td>→ Form a SR</td>
<td>OR + impersonal null subject</td>
</tr>
<tr>
<td></td>
<td>→ Form an OR?</td>
</tr>
</tbody>
</table>

**Experiments 1A&B:** Does a post-verbal subject cause a processing difficulty? Yes!

**Experiments 2A&B:** Does a (post-verbal) object cause a processing difficulty (filled-gap effect)?
Experiment 2A

Materials

CLASSIC FGE: The teacher looked for the student that we found. The event occurred according to the bag of him at the school’s playground.

BASELINE: The teacher looked for the student after we found. The event occurred according to the bag of him at the school’s playground.

FILLER.SG + VERB.PL → Form an OR
Processing disruption at the object?
SPR, 48 participants, 24 sets + 51 fillers (grammatical)
Experiment 2A
Materials

The teacher looked for the \{ \textit{students.pl} / \textit{student.sg} \} that \textit{found.pl} eventually \textit{acc. the bag} of them / him at the school’s playground.

‘The teacher looked for the student whose bag was eventually found at the playground’.

CLASSIC FGE: The teacher looked for the student that we found .\textit{pl} eventually \textit{acc. the bag} of him at the school’s playground.

BASELINE: The teacher looked for the student after we found .\textit{pl} eventually \textit{acc. the bag} of him at the school’s playground.

FILLER.SG + VERB.PL $\rightarrow$ Form an OR
Processing disruption at the object?
SPR, 48 participants, 24 sets + 51 fillers (grammatical)
Experiment 2A

Results

In the **classic FGE** condition, increased RTs at the direct object, relative to baseline.

→ Readers constructed an OR

In the **match** condition, no slowdown at the object.

→ Readers constructed a SR

FILLER.SG + VERB.PL → Form an OR
Processing disruption at the object?

SPR, 48 participants, 24 sets + 51 fillers (grammatical)
Experiment 2A

Results

In the **mismatch** condition, increased RTs at the direct object, relative to baseline. (In alignment with that observed in the classic FGE condition).

→ Readers constructed an **OR**, when the impersonal subject reading is available.
Experiment 2B

Materials

The teacher looked for the students / student by the end of the break eventually acc. the bag of them / him at the school’s playground.

‘The teacher looked for the student whose bag was eventually found at the playground by the end of the break’.

BASELINE: The teacher looked for the student after by the end of the break we found eventually acc. the bag of him at the school’s playground.
Experiment 2B

Results

Replication with extended filler-verb distance.

FILLER.SG + VERB.PL $\rightarrow$ Form an OR
Processing disruption at the object?
SPR, 36 participants, 24 sets + 51 fillers (grammatical)
Intermediate summary

FILLER.PL+VERB.SG (Experiments 1A&B, SPR): Readers prefer constructing an ungrammatical SR, over an OR with a post-verbal subject.

FILLER.SG+VERB.PL (Experiments 2A&B, SPR): When an impersonal subject can be licensed, the OR analysis (with a null subject) is adopted.

Experiment 3 aims to replicate both findings within one experiment, and to tap readers’ preferences more directly (in a sentence completion task).
Intermediate summary

FILLER.PL+VERB.SG (Experiments 1A&B, SPR): Readers prefer constructing an ungrammatical SR, over an OR with a post-verbal subject.

FILLER.SG+VERB.PL (Experiments 2A&B, SPR): When an impersonal subject can be licensed, the OR analysis (with a null subject) is adopted.

Experiment 3 aims to replicate both findings within one experiment, and to tap readers’ preferences more directly (in a sentence completion task).

- Sentences truncated after the verb (RSVP of the preamble)
- Crossing number agreement on the verb & the filler (four conditions)
- Testing for SR/OR completions
Experiment 3

Results

In filler-verb match conditions -

• Strictly SR completions, when an impersonal subject isn’t licensed.

• Mostly SR completions, when an impersonal null subject can be licensed.
Experiment 3

Results

In mismatch conditions -

• Mostly SR completions (ignoring subject-verb agreement), when only a post-verbal can restore grammaticality.

• Mostly OR completions, when an impersonal subject can be licensed.
Discussion

Our study demonstrates rational noisy channel comprehension in Hebrew, showing that:

• Readers apply elaborate probabilistic knowledge regarding the distribution of structures in the language.

• They are willing to compromise subject-verb agreement to allow for a more probable structure.

Namely, the interpretation of an agreement mismatch is modulated by the prior probability of the alternative analyses.
Discussion

Self-organized sentence processing challenges the assumption that a grammar supervises parsing by defining all possible structures. In this framework, noisy-channel effects are attributed to bottom-up activation of competing constructions. This can pull the interpretation towards a locally coherent (but globally ungrammatical) parse.

However, in our case, the adopted ungrammatical analysis is not locally coherent.

Tabor & Hutchins, 2004; Tabor, Galantucci, & Richardson, 2004
Directions for future research

When is it “rational” to assume a corrupted input?

- Quantitating the (im)probability of the structures.
- Testing whether readers assume post verbal subjects in other cases of filler mismatch (e.g. with short prepositions/case marking).
- Investigating the interaction between semantic plausibility and the structure’s frequency.
Thank you

Contact: maayankeshev@mail.tau.ac.il

Jesse Harris
Radan Nasrallah
CUNY reviewers
The likelihood of the different errors in the current study

Deletions are more likely (Gibson et al, 2013). Does it affect our finding?

**EXP1-2:**

**Impersonal subject – Post-verbal subject**

[VERB.pl] חיפש – חיפשו [VERB.sg]
[VERB] אנחנו – המלצר [NOUN.sg] [NOUN.pl] המלזר

We cannot determine if the deletion/addition contrast is a confound since we don’t know for certain whether readers amend the representation of the verb or a that of the filler.
The likelihood of the different errors in the current study

Deletions are more likely (Gibson et al, 2013). Does it affect our finding?

**EXP3:**

**Impersonal subject – Post-verbal subject**

[VERB.pl] ניסיון – חיפוש [VERB.sg]

[NOUN.sg] המלצרית – המשלצרות [NOUN.pl]

In experiment 3 this problem was removed by using the verbs and fillers in their feminine form. In this case, amending the verb requires substitution in both conditions. Amending the filler is more likely in the impersonal subject condition (in contrast to our prediction, and thus does not confound our findings).
Additional figures and data
Experiment 1A

<table>
<thead>
<tr>
<th>We passed</th>
<th>near</th>
<th>the inspector</th>
<th>while</th>
<th>that during</th>
<th>the traffic jam</th>
<th>continued</th>
<th>allegedly</th>
<th>the driver</th>
<th>the violent</th>
<th>to attack</th>
<th>acc.</th>
<th>the inspector</th>
<th>without</th>
<th>any</th>
<th>shame</th>
</tr>
</thead>
</table>
Experiment 1B

<table>
<thead>
<tr>
<th>We passed</th>
<th>near</th>
<th>the inspector</th>
<th>while</th>
<th>that during</th>
<th>the traffic jam</th>
<th>continued</th>
<th>allegedly</th>
<th>the driver</th>
<th>the violent</th>
<th>to attack</th>
<th>acc.</th>
<th>the inspector</th>
<th>without</th>
<th>any</th>
<th>shame</th>
</tr>
</thead>
</table>
The hairdresser talked with the designers because that they chose mostly acc. the models the main of them in the moment the last
Experiment 2B
The experiment also included manipulation of presentation duration [Regular: 300ms + 100ms of blank; Speeded: 100ms + 300ms of blank].