The Syntax of Ellipsis Resolution: Eye-tracking Evidence from $\phi$-Feature Mismatches

Helena Aparicio, Katie Franich, Ming Xiang

University of Chicago

NELS 45

October 30, 2014
In elliptical constructions some unpronounced material is being interpreted:

(1) Abbey is very tall. Ben, too.

(2) Abbey was reading a book. Ben was too.
Syntactic vs. Semantic Identity

A central question in the study of ellipsis:

- Does the identity condition for ellipsis hold over syntactic or semantic representations (or both)?
Syntactic vs. Semantic Identity

- **Syntactic Identity Accounts:** Ellipsis is licensed through identity between syntactic phrase markers

(Sag 1976; Fiengo & May 1994; Chung et al. 1995; Frazier & Clifton 2001; Baltin 2012; Merchant 2013)

Antecedent:  

```
    John
   /   \  
VP    
|     |
V     NP
|     |
played violin
```

Ellipsis:  

```
    Bill
did
   /   \  
<VP>    
|     |
V     NP
|     |
played violin
```
Semantics Identity Accounts: Ellipsis is licensed through identity between semantic representations

(Hardt 1993; Dalrymple et al. 1991; Ginzburg & Sag 2000; Merchant 2001; Culicover & Jackendoff 2005; Merchant 2010; Merchant 2014)

Antecedent:  

```
John
   VP₁
     V
     NP
      played violin
```

Ellipsis:  

```
Bill
did pro VP₂
```

(3) \[ \text{pro}_{VP}^n g = g(n) \]

(4) \[ \lambda x. x \text{played the violin}_1 = \lambda x. x \text{played the violin}_2 \]
Voice Mismatch

- Voice mismatches under ellipsis have constituted a major domain for investigation of this question (Merchant 2013).
  - The ungrammaticality of certain voice mismatches suggest that the identity condition is syntactic:

(5) *This problem was looked into by John, and Bill did too.

(6) *This information was released but Gorbachov didn’t.
However, certain mismatches are **acceptable**:

(7) This problem was to have been looked into, but nobody did.

A noted **cline in acceptability for VPE voice mismatches** suggests that the facts are more nuanced (Frazier and Clifton 2006; Arregui, et al. 2006; Kim et al. 2011):

(8) a. None of the astronomers saw the comet, but John did.
   b. ? Seeing the comet was nearly impossible, but John did.
   c. ?? The comet was nearly impossible to see, but John did.
   d. ??? The comet was nearly unseeable, but John did.

(Arregui, et al. 2006)
Possible sources for the acceptability cline:

1. Discourse coherence plays a role in the resolution of these mismatches (Kehler 2000; Kertz 2013).
Inflectional $\phi$-Feature Mismatch

- **Inflectional $\phi$-feature mismatches** constitute another potential area to investigate the nature of the identity condition
  - Inflectional $\phi$-features are not seen to be part of the discourse representation
Roadmap

1. Overview of accounts for inflectional $\phi$-feature mismatches in ellipsis

2. Experiment: Feature mismatch in Spanish in three constructions:
   - Ellipsis
   - Non-elliptical (‘Full’)
   - Deep anaphora

3. Proposal

4. Conclusions
Claims found in the literature:

- **Inflectional φ-features are not relevant in ellipsis computation.** (Bobaljik and Zocca 2011; Merchant 2014)

Greek (Merchant 2014)

(9) O Petros ine ikanos, ala i Maria dhen ine <ikanι>. the Petros is **capable.m.sg** but the Maria not is **capable.f.sg** ‘Petros is capable, but Maria isn’t.’

Brazilian Portuguese (Nunes & Zocca 2009; Masullo & Depiante 2004, for Spanish)

(10) O João é alto e a Maria também é <alta>. the João is **tall.masc.sg** and the Maria also is **tall-fem.sg** ‘João is tall and Maria is too.’

*For nominal φ-feature mismatches see Bobaljik & Zocca (2011) and Merchant (2014).*
Inflectional $\phi$-Feature Identity Under Ellipsis

- Inflectional $\phi$-features are not typically analyzed as contributing to the semantic representation.
- The availability of inflectional $\phi$-feature mismatches can be accounted for if we assume a semantic identity condition for ellipsis:

\[(11) \text{ An } XP_E \text{ can be elided under identity with an antecedent } YP_A \text{ only if } [XP] = [YP].\]
The acceptability of inflectional $\phi$-feature mismatches can also be reconciled with syntactic identity if we assume an analysis in which predicate adjectives enter the derivation unspecified for $\phi$-features (Nunes & Zocca 2009):

(12) O João é alto e a Maria também é
tall.masc.sg the João is tall and the Maria also is
<alta>.
tall-fem.sg
‘João is tall and Maria is too.’

(13) a. [[O João] é [AgrP Agr masc.sg [AP alt-]]]
b. [[a Maria] também é [AgrP Agr fem.sg [AP alt-]]]
Our Claims

1. Contra previous claims, **ellipsis is sensitive to inflectional $\phi$-feature mismatches**, as evidenced by decreased acceptability and increased processing difficulty for mismatches between the ellipsis site and the antecedent.

2. Ellipsis resolution is sensitive to the syntactic structure of the antecedent. A purely semantic resolution process for ellipsis cannot account for this data (Xiang et al. 2014).

3. There exists a different time-course effect for deep anaphora and ellipsis resolution suggestive of syntactic identity.
Three structures tested:

1. Ellipsis
2. Non-elliptical (‘Full’)

(14) [ Hankamer attempts to stuff a 9-inch ball through a 6-inch hoop ]

a. Sag: It’s not clear that you’ll be able to do it.
   (Deep anaphora)

b. #Sag: It’s not clear that you’ll be able to.
   (Ellipsis/Surface anaphora)

(Hankamer & Sag 1976: 392)
Lo as Deep-anaphor

Lo does not require a linguistic antecedent. It can be pragmatically controlled:

Context: The tallest players on the basketball team are getting picked to play in the national team. Juan does not get picked and he yells angrily at the coach:

(15) a. Por qué no he sido seleccionado? Yo lo soy
for what not have been selected? I it am
también!
too!
‘Why haven’t I been selected? I am <tall> too!’

b. Por qué no he sido seleccionado? #Yo también!
for what not have been selected? I too!
‘Why haven’t I been selected? I am <tall> too!’
Two features examined:

1. Gender (masculine vs. feminine)
2. Number (singular vs. plural)

2 X 2 Design:

1. Match (feature match versus mismatch with the antecedent)
2. Feature on the Subject of the Second Clause (SSC)
Gender Subexperiment: Ellipsis

(16) a. El ciclista es alto y el futbolista también. The.m cyclist is tall.m and the.m football-player too ‘The cyclist is tall and the football player is too.’
[Ellipsis, Match, Masc. Subject]

b. La ciclista es alta y la futbolista también. The.f cyclist is tall.f and the.f football-player too ‘The cyclist is tall and the football player is too.’
[Ellipsis, Match, Fem. Subject]

c. La ciclista es alta y el futbolista también. The.f cyclist is tall.f and the.m football-player too ‘The cyclist is tall and the football player is too.’
[Ellipsis, Mismatch, Masc. Subject]

d. El ciclista es alto y la futbolista también. The.m cyclist is tall.m and the.f football-player too ‘The cyclist is tall and the football player is too.’
[Ellipsis, Mismatch, Fem. Subject]
(17)  a. El ciclista es alto y **el futbolista** es alto también.
The.m cyclist is tall.m and the.m football-player is tall.m too
'The cyclist is tall and the football player is tall too.'
[Full, Match, Masc. Subject]

b. La ciclista es alta y **la futbolista** es alta también.
The.f cyclist is tall.f and the.f football-player is tall.f too
'The cyclist is tall and the football player is tall is tall too.'
[Full, Match, Fem. Subject]

c. La ciclista es alta y **el futbolista** es alto también.
The.f cyclist is tall.f and the.m football-player is tall.m too
'The cyclist is tall and the football player is tall too.'
[Full, Mismatch, Masc. Subject]

d. El ciclista es alto y **la futbolista** es alta también.
The.m cyclist is tall.m and the.f football-player is tall.f too
'The cyclist is tall and the football player is tall is tall too.'
[Full, Mismatch, Fem. Subject]
Gender Subexperiment: Deep-anaphor ‘Lo’

(18) a. El ciclista es alto y el futbolista lo es también.
    The.m cyclist is tall.m and the.m football-player LO is too
    ‘The cyclist is tall and the football player is tall too.’
    [Lo, Match, Masc. Subject]

b. La ciclista es alta y la futbolista lo es también.
    The.f cyclist is tall.f and the.f football-player LO is too
    ‘The cyclist is tall and the football player is tall too.’
    [Lo, Match, Fem. Subject]

c. La ciclista es alta y el futbolista lo es también.
    The.f cyclist is tall.f and the.m football-player LO is too
    ‘The cyclist is tall and the football player is tall too.’
    [Lo, Mismatch, Masc. Subject]

d. El ciclista es alto y la futbolista lo es también.
    The.m cyclist is tall.m and the.f football-player LO is too
    ‘The cyclist is tall and the football player is tall too.’
    [Lo, Mismatch, Fem. Subject]
Number Subexperiment: Ellipsis

(19) a. El profesor es severo-∅ y el decano también. The.sg professor is strict.sg and the.sg dean too ‘The professor is strict and the dean is too.’
[Ellipsis, Match, Sing. Subj]
b. Los profesores son severos y los decanos también. The.pl professors are strict.pl and the.pl deans too ‘The professors are strict and the deans is too.’
[Ellipsis, Match, Pl. Subject]
c. Los profesores son severos y el decano también. The.pl professors are strict.pl and the.sg dean too ‘The professors are strict and the dean is too.’
[Ellipsis, Mismatch, Sing. Subject]
d. El profesor es severo-∅ y los decanos también. The.sg professor is strict.sg and the.pl deans too ‘The professor is strict and the deans too.’
[Ellipsis, Mismatch, Pl. Subject]
(20) a. El profesor es severo-∅ y el decano es severo-∅ también. The.sg professor is strict.sg and the.sg dean is strict.sg too
‘The professor is strict and the dean is strict too.’ [Full, Match, Sg. Subject]

b. Los profesores son severos y los decanos son severos The.pl professors are strict.pl and the.pl deans are strict.pl
también. too
‘The professors are strict and the deans are strict too.’ [Full, Match, Pl. Subject]

c. Los profesores son severos y el decano es severo-∅ también. The.pl professors are strict.pl and the.sg dean is strict.sg too
‘The professors are strict and the dean is strict too.’ [Full, Mismatch, Sg. Subject]

d. El profesor es severo-∅ y los decanos son severos también. The.sg professor is strict.sg and the.pl deans are strict.pl too
‘The professor is strict and the deans are strict too.’ [Full, Mismatch, Pl. Subject]
Number Subexperiment: Deep-anaphor ‘Lo’

(21)  a. El profesor es severo-∅ y el decano lo es también.
   ‘The professor is strict and the dean is strict too.’
   [Lo, Match, Sg. Subject]

   b. Los profesores son severos y los decanos lo son también.
   ‘The professors are strict and the deans are strict too.’
   [Lo, Match, Pl. Subject]

   c. Los profesores son severos y el decano lo es también.
   ‘The professors are strict and the dean is strict too.’
   [Lo, Mismatch, Sg. Subject]

   d. El profesor es severo-∅ y los decanos lo son también.
   ‘The professor is strict and the deans are strict too.’
   [Lo, Mismatch, Pl. Subject]
Hypothetical Results

Syntactic Identity:

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Ellipsis</th>
<th>Full</th>
<th>Lo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading Times</td>
<td><img src="chart1.png" alt="" /></td>
<td><img src="chart1.png" alt="" /></td>
<td><img src="chart1.png" alt="" /></td>
</tr>
</tbody>
</table>

Semantic Identity:

<table>
<thead>
<tr>
<th>Condition</th>
<th>Ellipsis</th>
<th>Full</th>
<th>Lo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading Times</td>
<td><img src="chart2.png" alt="" /></td>
<td><img src="chart2.png" alt="" /></td>
<td><img src="chart2.png" alt="" /></td>
</tr>
</tbody>
</table>
Procedure

- Participants were 49 native (Iberian) Spanish speakers (23 female; age M=32 years)
- Eye-tracking experiment to examine online processing of \( \phi \)-feature mismatch
- Offline grammaticality judgments also elicited for each item during experiment
- 80 items administered through a Latin Square Design
- 50 ungrammatical fillers
- Number and Gender features tested over two sub-experiments (40 items each)
Critical Regions:

a. Ellipsis
| El ciclista | es alto | y | el futbolista | también. |
| The cyclist | is tall.m | and | the.m football.player | too. |

b. Full
| El ciclista | es alto | y | el futbolista | es alto | también. |
| The cyclist | is tall.m | and | the.m football.player | is tall.m | too. |

c. Lo
| El ciclista | es alto | y | el futbolista | lo es | también. |
| The cyclist | is tall.m | and | the.m football.player | it is | too. |
Eye-Movements

- **Regression Path (RP):** Total durations of the fixations in all the regions up to and including the region of interest, before the region is exited to the right (early processing measure)

- **Total Time (TT):** Total duration for all fixations in a given region (late processing measure)

- Two additional measures **First Fixation** and **First Pass** also collected, but no significant results to report
Grammaticality Judgments
Grammaticality Judgment Results

- All sentences judged to be highly acceptable (> 75% ‘yes’ resp.)
Grammaticality Judgment Results

- All sentences judged to be highly acceptable (> 75% ‘grammatical’ resp.)
- Signif. mismatch penalty in Ellipsis and Lo conditions; not in Full conditions
Grammaticality Judgment Results

- All sentences judged to be highly acceptable ( > 75% ‘grammatical’ resp.)
- Signif. mismatch penalty in Ellipsis and Lo conditions; not in Full conditions
Grammaticality Judgment Results

- All sentences judged to be highly acceptable (> 75% ‘grammatical’ resp.)
- Signif. mismatch penalty in Ellipsis and Lo conditions; not in Full conditions
Grammaticality Judgments

All sentences judged to be highly acceptable (> 75% ‘grammatical’ resp.)

Signif. mismatch penalty in Ellipsis and Lo conditions; not in Full conditions

Same effect found for Number items
Grammaticality Judgments

- All example sentences judged to be highly acceptable (> 75% ‘grammatical’ resp.)
- Significant mismatch penalty in Ellipsis and Lo conditions; not in Full conditions
- Same effects found for Number items
Eyetracking Data
Eyetracking Data

- Fixation types used:
  1. Regression Path (RP): Earlier stage processing measure
  2. Total Time (TT): Late stage processing measure
Eyetracking Data

Critical Region: Regression Path Fixation Times

CR Results

- Significant mismatch penalty in Elliptical sentences, but not Full or Lo sentences
Eyetracking Data

Critical Region: Regression Path Fixation Times

Significant mismatch penalty in Elliptical sentences, but not Full or Lo sentences
Eyetracking Data

Critical Region: Regression Path Fixation Times

![Bar chart showing log-transformed reading times for different SSC and condition combinations.](chart.png)

**CR Results**
- Significant mismatch penalty in Elliptical sentences, but not Full or Lo sentences.
- Same effect found for Gender and Number items.
Eyetracking Data

Critical Region: Regression Path Fixation Times

CR Results

- Significant mismatch penalty in Elliptical sentences, but not Full or Lo sentences
- Same effect found for Gender and Number items
- Same effects found for TT
Eyetracking Data

Critical Region:

- Significant mismatch penalty for RP for Elliptical sentences; no penalty for Full or Lo structures
- Same effect found for Gender and Number items
- Same effects found for TT
Eyetracking Data

Post-Critical Region (CR+1): Regression Path Fixation Times

CR+1 Results

- Significant mismatch penalty for RP in Lo sentences, but not Full sentences
Eyetracking Data

**Post-Critical Region (CR+1): Regression Path Fixation Times**

- **SSC = Feminine**
  - Full: 5.0
  - Lo: 5.5

- **SSC = Masculine**
  - Full: 6.0
  - Lo: 6.5

- **SSC = Plural**
  - Full: 7.0
  - Lo: 7.5

- **SSC = Singular**
  - Full: 5.0
  - Lo: 5.5

**CR+1 Results**

- Significant mismatch penalty for RP in *Lo* sentences, but not Full sentences.
Eyetracking Data

Post-Critical Region (CR+1): Regression Path Fixation Times

CR+1 Results
- Significant mismatch penalty for RP in \( Lo \) sentences, but not Full sentences
- Same effect found for Gender and Number items
Eyetracking Data

Post-Critical Region (CR+1): Regression Path Fixation Times

CR+1 Results

- Significant mismatch penalty for RP in Lo sentences, but not Full sentences
- Same effect found for Gender and Number items
- No significant results for TT
Eyetracking Data

Post-Critical Region (CR+1):

- Significant mismatch penalty for RP in Lo sentences, but not Full sentences
- Same effect found for Gender and Number
- No significant results for TT
Contra previous claims, ellipsis is sensitive to inflectional $\phi$-feature mismatches, as evidenced by decreased acceptability and increased processing difficulty for mismatches between the ellipsis site and the antecedent.
Contra previous claims, ellipsis is sensitive to inflectional \( \phi \)-feature mismatches, as evidenced by decreased acceptability and increased processing difficulty for mismatches between the ellipsis site and the antecedent.

This provides evidence that ellipsis resolution is sensitive to the syntactic structure of the antecedent. A purely semantic resolution process for ellipsis cannot account for this data (Xiang, et al. 2014).
What about mismatch effect with *Lo* sentences?
Results Summary

What about mismatch effect with *Lo* sentences?

- Clear time-course difference in mismatch effect between Ellipsis and *Lo* sentences provide evidence for differences in the nature of this effect.
What about mismatch effect with *Lo* sentences?

- Clear time-course difference in mismatch effect between Ellipsis and *Lo* sentences provide evidence for differences in the nature of this effect

1. Ellipsis mismatches processed earlier in time (critical region)
2. *Lo* sentences processed later in time (post-critical region)
Sensitivity to feature mismatch for deep anaphora has been found in previous literature

Time-course differences between deep and surface anaphora have also been found
Explaining the time-course difference

**Surface Anaphora:** early effects of mismatch in Ellipsis provide evidence for calculation of syntactic identity

**Deep Anaphora:** later effects of mismatch on \( L_0 \) suggest that mismatch is not relevant at the syntactic level; rather, discourse parallelism between antecedent and anaphor factors into increased processing times
Mismatch Sensitivity in Surface Anaphora

Our Proposal:

1. Mismatch is grammatical

2. Added processing cost of mismatch has to do with the particular strategy used by the parser in locating a suitable (matching) antecedent (Kim, et al. 2011)
Mismatch Sensitivity in Surface Anaphora

- Assuming that structure building in sentence parsing occurs incrementally, a variety of parsing heuristics may be used to allow for maximal efficiency
Mismatch Sensitivity in Surface Anaphora

- Assuming that structure building in sentence parsing occurs incrementally, a variety of parsing heuristics may be used to allow for maximal efficiency.

- For ellipsis, such a heuristic has been proposed:
Assuming that structure building in sentence parsing occurs incrementally, a variety of parsing heuristics may be used to allow for maximal efficiency.

For ellipsis, such a heuristic has been proposed:

MaxElide: Ellipsis targets configurationally higher nodes over lower nodes (Merchant 2008; Kim, et al. 2011)
Mismatch Sensitivity in Surface Anaphora

Given the structure below for the predicate (Nunes & Zocca 2005; 2009):

(22)

\[
\text{AgrP} \quad \phi: [\text{gender}] \\
\quad \text{[number]} \quad \text{AP} \\
\quad \text{Adj}
\]
Mismatch Sensitivity in Surface Anaphora

- The parser proceeds iteratively looking for the right antecedent:

  Antecedent:

  John
  is
  AgrP
  Agr
  ϕ: [+ masc]
  [+ sing]
  AP
  tall

  Ellipsis:

  Mary
  is
  AgrP
  Agr
  ϕ: [+ fem]
  [+ sing]
  AP
  tall

  Cues used by the parser: [+fem] [+sing] [AgrP]
Mismatch Sensitivity in Surface Anaphora

The parser proceeds iteratively looking for the right antecedent:

Antecedent:

John
  is
  AgrP
    Agr
      φ: [+ masc] [+ sing]
    AP
      tall

Ellipsis:

Mary
  is
  AgrP
    Agr
      φ: [+ fem] [+ sing]
    AP
      tall

Cues used by the parser: [+fem] [+sing] [AgrP]
The parser proceeds iteratively looking for the right antecedent:

- Antecedent:
  - John
  - is AgrP
  - Agr
    - $\phi$: [+ masc] [+ sing]
  - AP:
    - tall

- Ellipsis:
  - Mary
  - is AgrP
  - Agr
    - $\phi$: [+ fem] [+ sing]
  - AP:
    - tall

Cues used by the parser: [AP]
The fact that we observed high acceptability despite longer processing times in the elliptical conditions results from the extra steps the parser needs to perform in order to retrieve the correct antecedent.
Deep Anaphora
Our Proposal:

1. Similar to previous studies, sensitivity to mismatch in *Lo* sentences can be attributed to parallelism effects which may take into account factors beyond the material in the antecedent (Carlson 2001)
Conclusion

- $\phi$-features are relevant for ellipsis computation

- Early processing costs are associated with ellipsis resolution but not deep anaphora resolution

- Taken together, this constitutes evidence for syntactic identity in ellipsis resolution

- Mismatches are grammatical; parsing heuristics (i.e. MaxElide) which drive antecedent retrieval are responsible for the mismatch penalty

- Deep-anaphora resolution is sensitive to feature mismatch, suggesting that a parallelism heuristic is active for anaphor resolution
Thank You!

Acknowledgements: Jason Merchant, Greg Kobele, Chris Kennedy, Karlos Arregi, Audience of the Workshop on Language, Cognition and Computation at the University of Chicago, Audience of the Texas Linguistic Society Conference, the University of Chicago Language Processing Lab, and all the participants in the study.
References


Merchant, J. (To appear: *Lingua*). Gender mismatches under nominal ellipsis.
Explaining the time-course difference

Preview Effect on Subject of the Second Clause (CR-1)

a. **Ellipsis**

| El ciclista | es alto | y | el futbolista | **también.** |
| The cyclist | is tall.m | and | the.m football.player | **too.** |

d. **Full**

| El ciclista | es alto | y | el futbolista | **es alto** | **también.** |
| The cyclist | is tall.m | and | the.m football.player | **is tall.m** | **too.** |

c. **Lo**

| El ciclista | es alto | y | el futbolista | **lo es** | **también.** |
| The cyclist | is tall.m | and | the.m football.player | **CL is** | **too.** |
Explaining the time-course difference

Preview Effect on Subject of the Second Clause (CR-1)

a. Ellipsis
   | El ciclista | es alto | y | el futbolista | también. |
   | The cyclist | is tall.m | and | the.m football.player | too. |

b. Full
   | El ciclista | es alto | y | el futbolista | es alto | también. |
   | The cyclist | is tall.m | and | the.m football.player | is tall.m | too. |

c. Lo
   | El ciclista | es alto | y | el futbolista | lo es | también. |
   | The cyclist | is tall.m | and | the.m football.player | CL is | too. |
Explaining the time-course difference

TT: Subject of the Second Clause

SSC Results

- Significant mismatch penalty for TT in Ellipsis sentences, but not Lo or Full sentences