



Plural Pronouns Under Uncertainty: A Multi-Method Investigation of Underspecification?

Derya Çokal

University of Cologne

12.08.2026

Today's Topics

- Annotators' (i.e., raters') disagreements on singular and plural pronouns
- Theoretical claims on plural pronouns
- Psycholinguistic evidence on underspecified or good-enough representation
- Three experiments on disagreement cases of:
 - Plural pronoun in mereological cases
 - Plural references to set/subset group references
 - ERP signatures for ambiguous plural pronouns

Corpus evidence

Annotators' / Raters' disagreement: mereology cases

Example

- 1a. The engineer hooked up the engine **to** the boxcar and sent **it** to London.
- 1b. The engineer separated the engine **from** the boxcar and sent **it** to London.

(Poesio & Reyle, 2001)

Raters' disagreement: plural pronoun

Example

2a. For a “barrier-free Bremen”, [several hundred disabled people] went into the streets yesterday — and demonstrated for “Equality, not Barriers”.

2 b. “Why always us?” [the congregated] asked on the posters.

(Versley, 2008)

Raters' disagreement: plural pronoun in set-subset

Example

1. The Fusion Music Festival line-up in Birmingham was extremely eclectic. The bands had rehearsed consistently. The guitarists had extensively checked the stage. They had prepared carefully.

Raters' disagreement: plural pronoun in non-subset

Example

2. The Fusion Music Festival line-up in Birmingham was extremely eclectic. The bands had rehearsed consistently. The journalists had extensively checked the stage. They had prepared carefully.

The same disagreement pattern can have different sources

Possible source	What it means
Annotation noise	A response resulting from misunderstanding, inattention or annotation error.
Referential ambiguity	The discourse genuinely licenses more than one referent.
Different commitments	Different people commit to different licensed interpretations.
Underspecification	An individual leaves the pronoun's reference partially unresolved.
Revision	An individual initially selects one interpretation and later updates it.

- Offline label distributions alone cannot distinguish these possibilities.
- Online measures can help identify the cognitive source of disagreement.

Theoretical explanations on plurality

Theoretical claims on plurality

- A common role = a maximal group effect is seen (Koh & Clifton, 2002).
- Maximal grouping is not allowed in all plural cases (Çokal & Sturt, 2017).
- Too difficult to individuate the two relevant individuals (Gordon et al., 1999; Moxey et al., 2004; Sanford & Moxey, 1995).

Examples:

3a. Tom, Mary **and** Harry went to a restaurant last night. **They** ate a pizza. (They = Tom + Mary + Harry)

3b. Tom recognized Mary **and** Harry in the restaurant. **They** ate a pizza. (They $\neq$ Mary and Harry)

Psycholinguistic evidence on underspecified or
good-enough representation

Two accounts of shallow processing

Underspecified representation

- Semantic distinctions may remain unspecified.
- Readers follow a partial interpretation.
- Reduced processing cost may provide evidence consistent with underspecification.

Good-enough representation

- Representation is lacking in detail, sketchy, heuristic, and incomplete — but "good enough"
(Ferreira, 2003; Ferreira, Bailey, & Ferraro, 2002)
- Misinterpretation is not revised.
- Top-down input from semantic memory provides enough equilibrium for processing
(Karimi & Ferreira, 2015)

Research Questions (RQ)

1. When people disagree about a plural pronoun's referent, what does that disagreement reveal about their underlying discourse representations?
 - We combine offline responses, eye movements and ERPs to investigate the source of disagreement.

Study 1: Plural reference to structured entities

Çokal, D., Filik, R., Sturt, P., & Poesio, M. (2023). Anaphoric reference to mereological entities. *Discourse Processes*, 60(3), 202-223

Stimuli in mereology cases

Examples

- 1a. The engineer hooked up the engine **to** the boxcar and sent **it/them** to London.
- 1b. The engineer separated the engine **from** the boxcar and sent **it/them** to London.

Design

1. **Experiment 1 (forced-choice task)** (i.e., ask participants what it/they refer to)
2. **Experiment 2 (sentence completion)** (i.e., give the initial part of the sentences and complete the rest and then underline what it and they refer to)
3. **Experiment 3 (paraphrase)** (i.e., paraphrase given sentences)

Brief results

Results

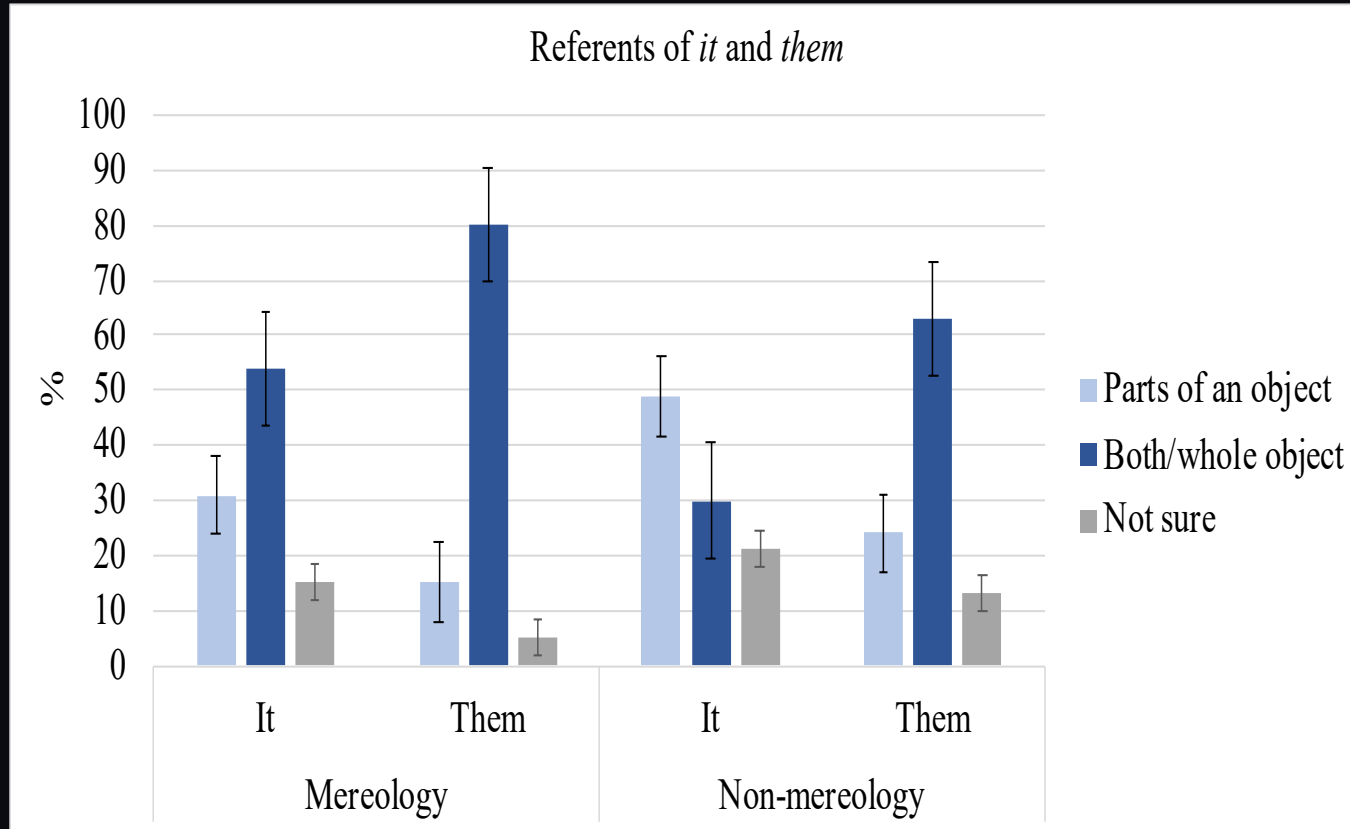


Figure 1. Forced-choice experiment. Percentage of referents of *it* and *them* in the mereology and non-mereology conditions

- Parts of an object:
 - engine or goods wagon
- Both objects:
 - engine + goods wagon
- Whole object (i.e., structured plural entity):
 - engine + goods wagon = train

Interim conclusions

- A mereological relation can create a structured whole-object referent in addition to the explicitly mentioned parts.
- Variation in antecedent choices may reflect different levels of discourse representation: an individual part, a sum of parts, or a constructed whole.

Study 2: Plural pronouns in set and subset contexts

Çokal, D., & Poesio, M. (re-submitted). Processing plural pronouns with overlapping antecedents: Underspecified or immediate commitment. *Discourse Processes*.

Design

Part 1 (Offline)

1. Sentence completion
2. Paraphrase

Part 2 (Online)

1. Online eye-tracking reading

Sample stimuli & predictions

A set-subset condition with an ambiguous verb :

The Fusion Music Festival line-up in Birmingham was extremely eclectic. **The bands** had rehearsed consistently. **The guitarists** had extensively checked the stage. **They** had prepared carefully.

- *They* = common role (i.e., musicians) = a maximal group reference
(a group [bands + guitarists])  shorter reading times

Sample stimuli & predictions

Non-subset condition with an ambiguous verb:

The Fusion Music Festival line-up in Birmingham was extremely eclectic. **The bands** had rehearsed consistently. **The builders** had extensively checked the stage. **They** had prepared carefully...

- *They* = no common role = no maximal group reference

(a group $\neq$ bands + builders) $\longrightarrow$ longer reading times

Sample stimuli & predictions

A set-subset condition with an unambiguous verb:

The Fusion Music Festival line-up in Birmingham was extremely eclectic. **The bands** had rehearsed consistently. **The guitarists** had extensively checked the stage. **They** had **strummed** carefully.

- Initial interpretation = maximal group but one group
interpretation 'strummed' is primed → longer reading times

Sample stimuli & predictions

Non-subset condition with an unambiguous verb:

The Fusion Music Festival line-up in Birmingham was extremely eclectic. **The bands** had rehearsed consistently. **The builders** had extensively checked the stage. **They** had repaired carefully.

- No overlapping features = one group is primed  shorter reading times

Brief results

Results

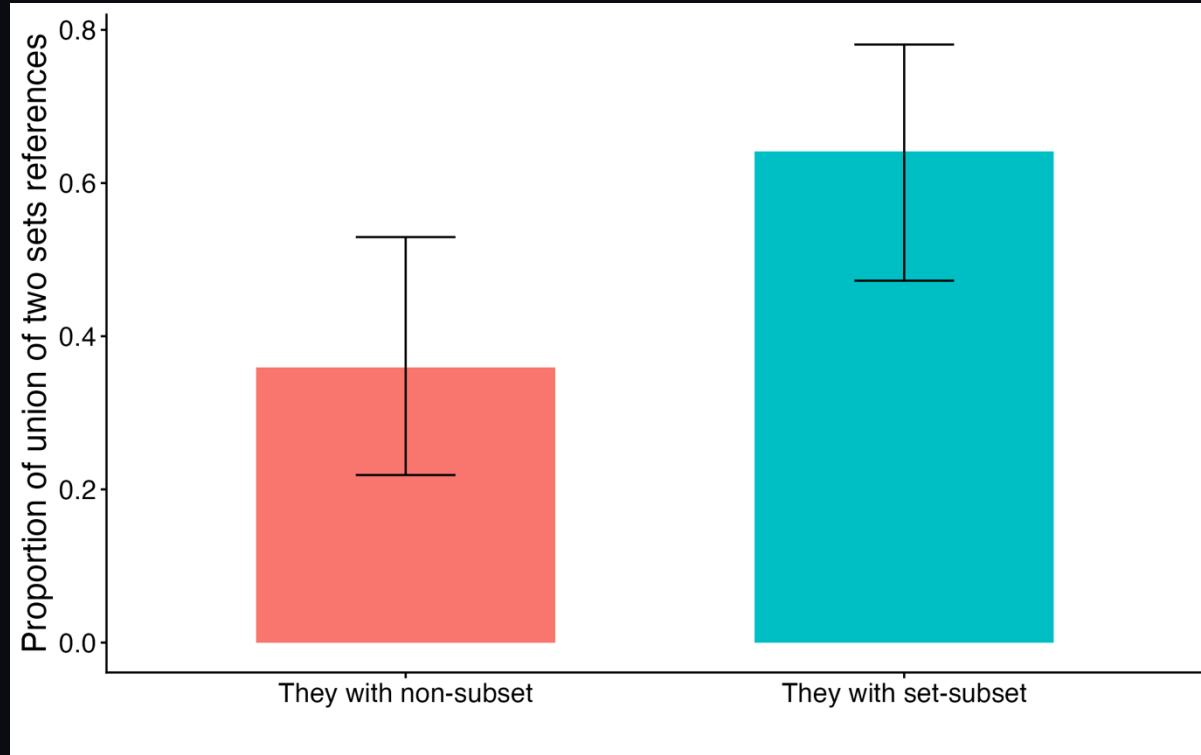


Figure 2. Sentence completion experiment. References of *they* to the set-subset and non-subset groups.

A sample stimuli:

Set-subset:

The bands had rehearsed consistently.
The guitarists had extensively checked the stage.

They.....

Non-subset:

The bands had rehearsed consistently.
The builders had extensively checked the stage.

They.....

Total reading times in adverbial region

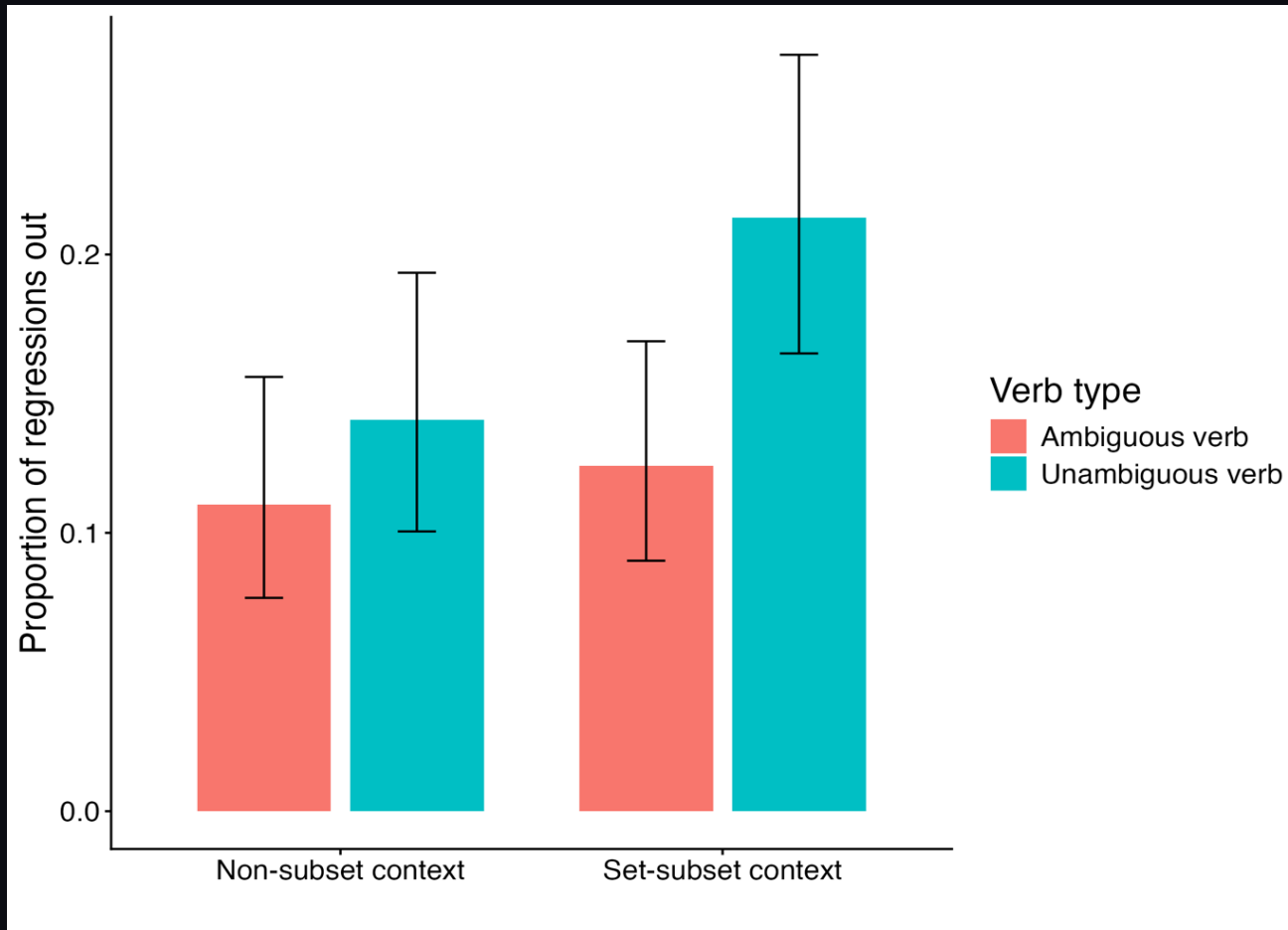


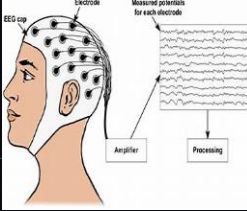
Figure 3. Eye-tracking experiment. Total reading times in the adverbial region

Adverbial region

- The builders had extensively checked the stage. They had repaired **carefully** and...
- The guitarists had extensively checked the stage. They had strummed **carefully** and...

Interim conclusions

- Overlapping group features (i.e., set-subset) lead to a union-of-groups interpretation more strongly than a relation between non-subset group context.
- In the online reading experiment, there is an ambiguity advantage.
- The distribution of antecedent choices is conditioned by the semantic relationship between the candidate groups.



Study 3: ERPs on ambiguous plural pronouns in German

Çokal, D., Phillip, M., Konuk, N.G., Villalba, M., von Heusinger, K., Poesio, M. (accepted). ERP Signatures of Prediction Error in the Comprehension of Ambiguous Plural Pronouns in German.
Language, Cognition, & Neuroscience.

Our predictions

- Readers would detect ambiguity and frontal sustained negative shift around 300ms (Nref) would be elicited (Van Berkum et al., 2007).
- Readers do not commit to a particular referent and thus no Nref effect would be observed.
- **Prediction error:** readers initially select one interpretation and update it when the verb identifies a different referent (see Bornkessel-Schlesewsky & Schlesewsky, 2019 for a review on prediction; Nour Eddine et al., 2024 for a predictive coding model of the N400; Friston, 2010; Rao & Ballard, 1999).
- Prediction error = N400
- Discourse updating = P600 (Delogu et al., 2019, 2021; Schumacher, 2009)

Plural pronouns in GERMAN!

- Ambiguous verb: The modern artists were present. The sculptors/journalists were there. It was a nice ambience. *They* browsed
 - Es war in einem Ausstellungssaal. Die modernen Künstler waren anwesend. Die Bildhauer/Die Journalisten waren da. Es war ein nettes Ambiente. Sie überflogen| ...
- Unambiguous verb: The modern artists were present. The sculptors/ journalists were there. It was a nice ambience. *They* carved/or wrote
 - Es war in einem Ausstellungssaal. Die modernen Künstler waren anwesend. Die Bildhauer/Die Journalisten waren da. Es war ein nettes Ambiente. Sie schnitzten/verfassten

Brief results

Pronoun region

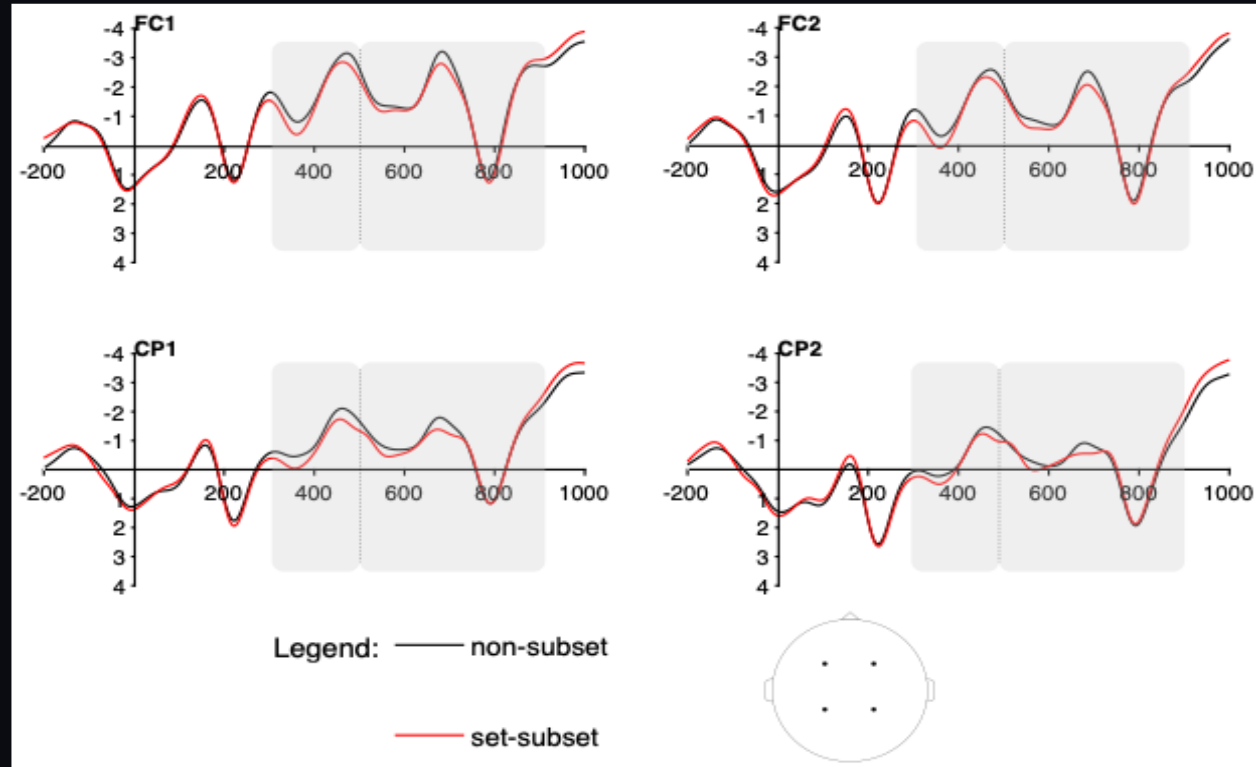


Figure 4. Grand average ERPs relative to the (ambiguous) plural personal pronoun for four selected electrodes. Negativity is plotted upwards

Ambiguous verb: The sculptors/journalists ...They| **browsed**|

Unambiguous verb: The sculptors/ journalists... They| **carved**/or **wrote**

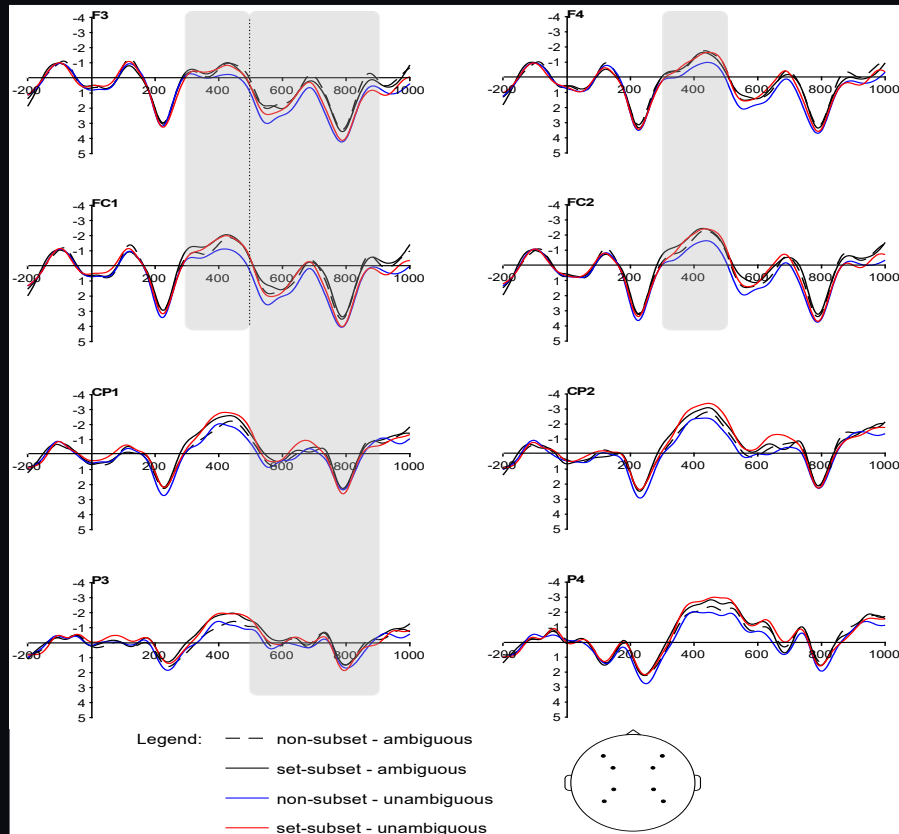


Figure 5. Grand average ERPs (N400) relative to the verb for eight selected electrodes.

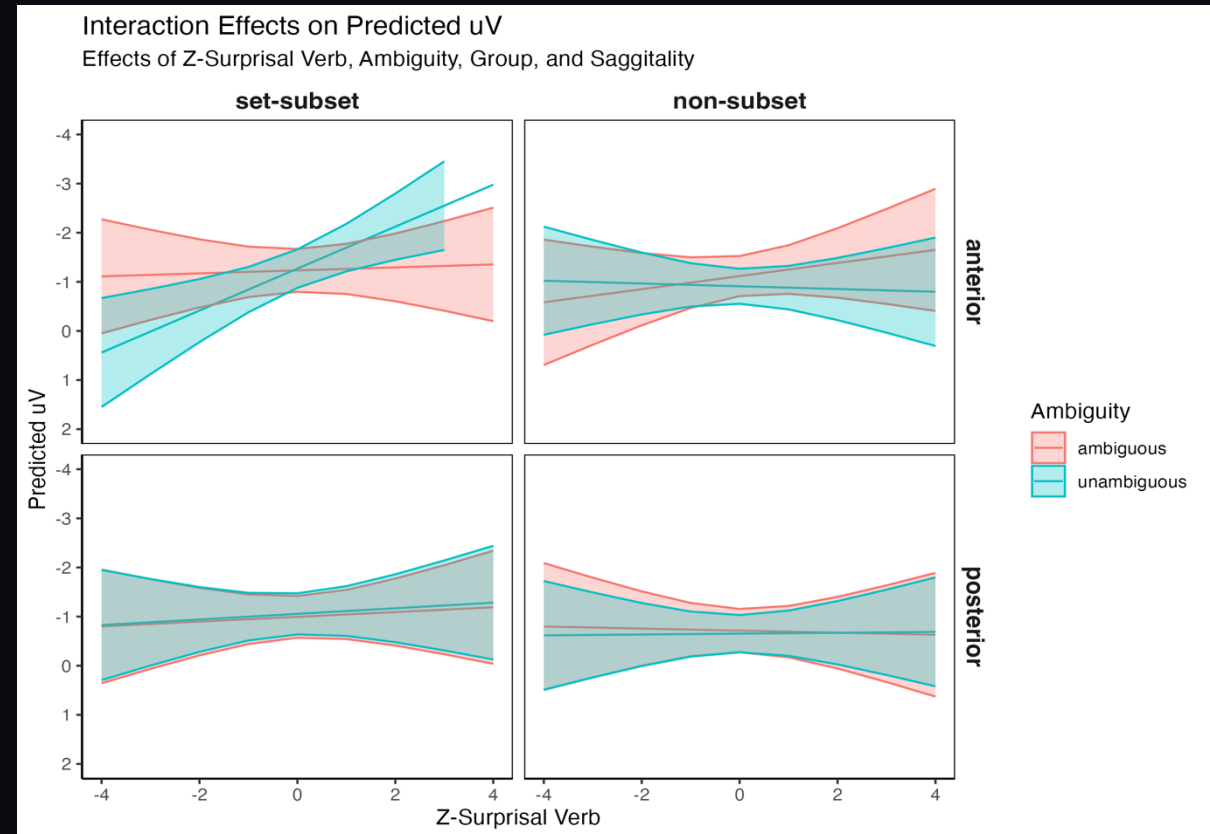
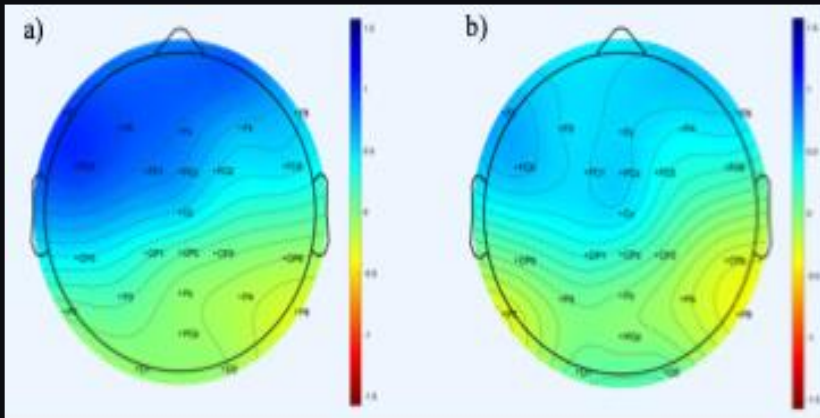


Figure 6. Interaction of ambiguity, context (group), sagittality, and z-surprisal verb at the verb

P600 at the Verb: Unambiguous Verbs Trigger Discourse Updating

- Unambiguous verbs (carved/wrote) elicited a larger P600 than ambiguous verbs (browsed) in both contexts.
- Effect is left-lateralized and anterior-weighted.
- The larger P600 reflects the cost of updating the discourse model.
- The process of discourse model updating is not impacted by surprisal (i.e., word probability).



Note. Topographical distribution of grand average mean amplitudes (500-900 ms) relative to the verb with (a) unambiguous verb types and (b) ambiguous verb types.

Figure 7. P600 at the verb

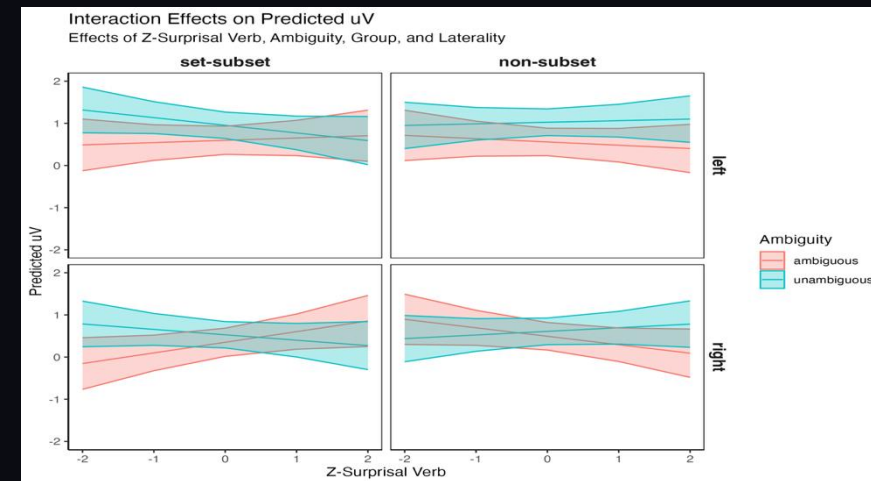


Figure 8. Interaction of context (group), ambiguity, laterality and z-surprisal verb at the verb

Interim conclusions

- Set–subset contexts lead to an initial union-of-two-sets interpretation.
- Therefore, comprehenders appear to have a partial interpretation and revise when necessary.
- Unambiguous verbs trigger discourse updating.

Let's remember our questions

RQ: What does rater's disagreement reveal about underlying discourse representations of disagreement ?

- Offline tasks identify contexts that allow multiple interpretations.
- Plural pronouns in mereology and set-subset cases.
- However, in online task, less processing costs is observed.
 - = ambiguity advantage & shallow processing.
- Prediction error is observed for some cases but not all cases.
- Readers update their discourse representation.

Implications for complex ground truth

- Preserve the distribution of antecedent choices rather than imposing a single gold label.
- Represent relations between candidate referents: part–whole, set–subset, and non-subset.
- Distinguish annotator variation from within-annotator uncertainty.
- Combine offline and online evidence to separate licensed interpretations from annotation noise.
- Complex ground truth should capture not only which antecedents are selected, but also why multiple interpretations are available.

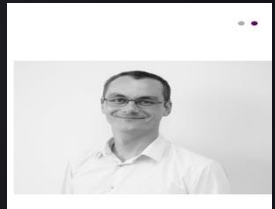
Take-home message

- Disagreement is structured, but its cognitive source must be investigated.
- A label distribution is the beginning of the explanation, not the end.



Thanks to my collaborators and
funding agencies!

And thank you for your attention.



European Research Council
Established by the European Commission



SFB 1252
**PROMINENCE
IN LANGUAGE**
Gefördert durch
DFG Deutsche
Forschungsgemeinschaft