

Interspeaker Variation in Copular Agreement with Disjoined Subjects: An Optimality Theoretic Account

Jiayi Lu ¹, Brandon Papineau ², Sarang Jeong ², Alexia Hernandez ¹, Emily Goodwin ², Arto Anttila ²

¹ University of Pennsylvania

² Stanford University

Background: subject-verb agreement

The students *is/are in the room.

The student is/*are in the room.

Background: S-V agreement with disjunction

Either the students or the teacher (is/are) in the room.

Either Bill or Mary (is/are) in the room.

English Peterson (1986), Foppolo and Staub (2020), German Himmelreich and Hartmann (2023)

Foppolo and Staub (2020)

- F&S examined the agreement pattern of “SG or SG” subjects in a reading time + acceptability judgment experiment.
- “SG or SG are” only triggered a minuscule acceptability degradation and RT slow down compared to other agreement errors.
- F&S concluded that there is a **grammatical lacuna**: speakers’ grammars do not prescribe a verb number following a disjointed subject.

Against a “grammatical lacuna”

- For disjunction agreement, it is clearly not “anything goes”.
- There are clear cases where only plural copula is allowed.

*Either the teachers or the students are/*is in the room.*

*Either the teacher or the students are/*is in the room.*

- Anecdotally, even for SG or SG, there seem to be some systematic inter-speaker variations.

Bresnan (2001)

is: SG

are: PL

are: Null

*PL: No plural copula/verb

*SG: No singular copula/verb

AGREE_{S-V}: A copula/verb must agree with the subject

Adapted from Bresnan (2001)

Bresnan (2001)

Mary <i>be(sg)</i> in the room	*PL	AGREE _{s-v}	*SG
→ is			*
are (plural)	*!	*	
are (null)		*!	

The students <i>be(pl)</i> in the room	*PL	AGREE _{s-v}	*SG
is		*	*!
are (plural)	*!		
→ are (null)		*	

Crucially, *PL is undominated, making “are (plural)” an impossible candidate

Proposal (inspired by Bresnan 2001)

Lexicon:

is: SG

are: Null

Constraints:

*SG: No singular copula/verb.

AGREE_{close}: The verb/copula needs to agree with the closer disjunct.

This is in fact the prescriptive rule for agreement with disjoined subjects (Garner, 2016; Bostwick 2023).

AGREE_{each}: The verb/copula needs to agree with each disjunct.

We assume *PL is always undominated, making *are*:PL an impossible candidate.

Possible Rankings

AGREE _{close}	>>	AGREE _{each}	>>	*SG
AGREE _{close}	>>	*SG	>>	AGREE _{each}
AGREE _{each}	>>	AGREE _{close}	>>	*SG
AGREE _{each}	>>	*SG	>>	AGREE _{close}
*SG	>>	AGREE _{close}	>>	AGREE _{each}
*SG	>>	AGREE _{each}	>>	AGREE _{close}

Pattern #1

AGREE_{close} >> **AGREE_{each}** >> ***SG**

AGREE_{each} >> **AGREE_{close}** >> ***SG**

AGREE_{close} >> ***SG** >> **AGREE_{each}**

Either John or Mary **is** in the room.

Either John or the students **are** in the room.

Either the students or John **is** in the room.

Either the students or the teachers **are** in the room.

Input	Output	AGREE_{close}	AGREE_{each}	*SG
either SG or SG	→ is (SG)			*
	are (null)	*!	*	
either SG or PL	is (SG)	*	*	*!
	→ are (null)	*	*	
either PL or SG	→ is (SG)		*	*
	are (null)	*!	*	
either PL or PL	is (SG)	*	*	*!
	→ are (null)	*	*	

Pattern #2

AGREE_{each} >> *SG >> AGREE_{close}

Either John or Mary **is** in the room.

Either John or the students **are** in the room

Either the students or John **are** in the room.

Either the students or the teachers **are** in the room.

Input	Output	AGREE _{each}	*SG	AGREE _{close}
either SG or SG	→ is (SG)		*	
	are (null)	*!		*
either SG or PL	is (SG)	*	*!	*
	→ are (null)	*		*
either PL or SG	is (SG)	*	*!	
	→ are (null)	*		*
either PL or PL	is (SG)	*	*!	*
	→ are (null)	*		*

Pattern #3

*SG >> AGREE_{close} >> AGREE_{each}

*SG >> AGREE_{each} >> AGREE_{close}

Either John or Mary **are** in the room.

Either John or the students **are** in the room

Either the students or John **are** in the room.

Either the students or the teachers **are** in the room.

Input	Output	*SG	AGREE _{close}	AGREE _{each}
either SG or SG	is (SG)	*!		
	→ are (null)		*	*
either SG or PL	is (SG)	*!	*	*
	→ are (null)		*	*
either PL or SG	is (SG)	*!		*
	→ are (null)		*	*
either PL or PL	is (SG)	*!	*	*
	→ are (null)		*	*

Experiment 1: Methods

N = 200 (Prolific, self-reported native English speakers)

In this experiment, you will read a series of sentences and will be asked to indicate the word which best completes it.

Please type the word 'is' or 'are' in the blank.

If you're unsure about your choice, follow your first instinct.

When you're ready to begin, press the space bar.

Either those nurses or Bill attending to the patients.

Submit

Experiment 1: Stimuli

4 conditions:

Either SG or SG: *Either John or Mary ___ interviewing the movie star.*

Either SG or PL: *Either John or those boys ___ interviewing the movie star.*

Either PL or SG: *Either those boys or John ___ interviewing the movie star.*

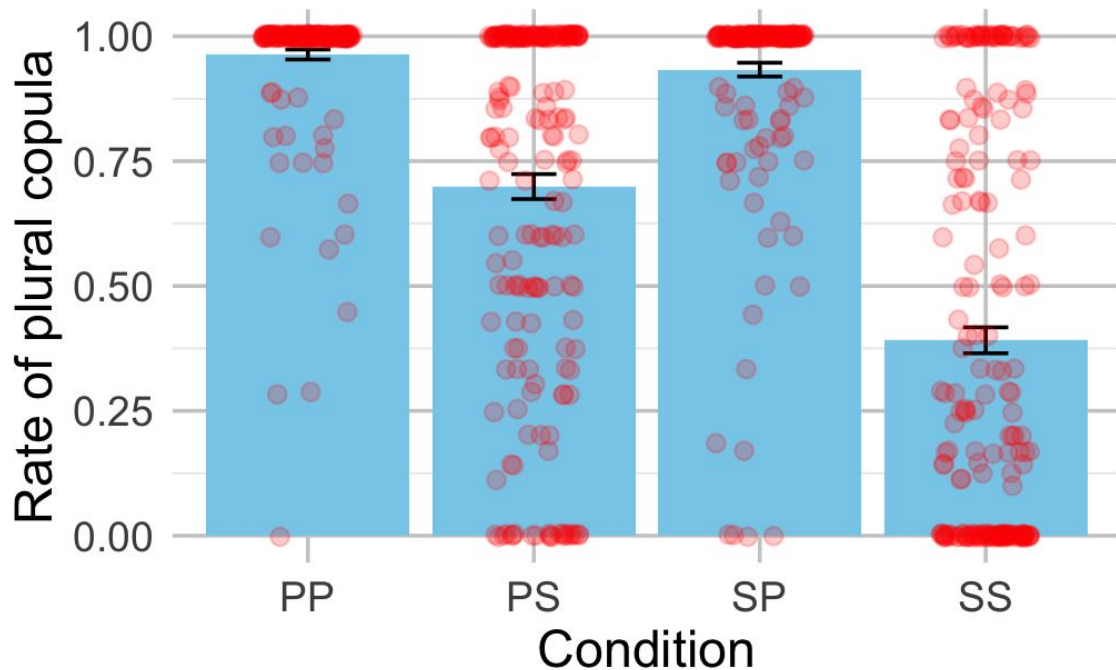
Either PL or PL: *Either these girls or those boys ___ interviewing the movie star.*

24 items, 72 trials in total; 48 fillers

Experiment 1: Results

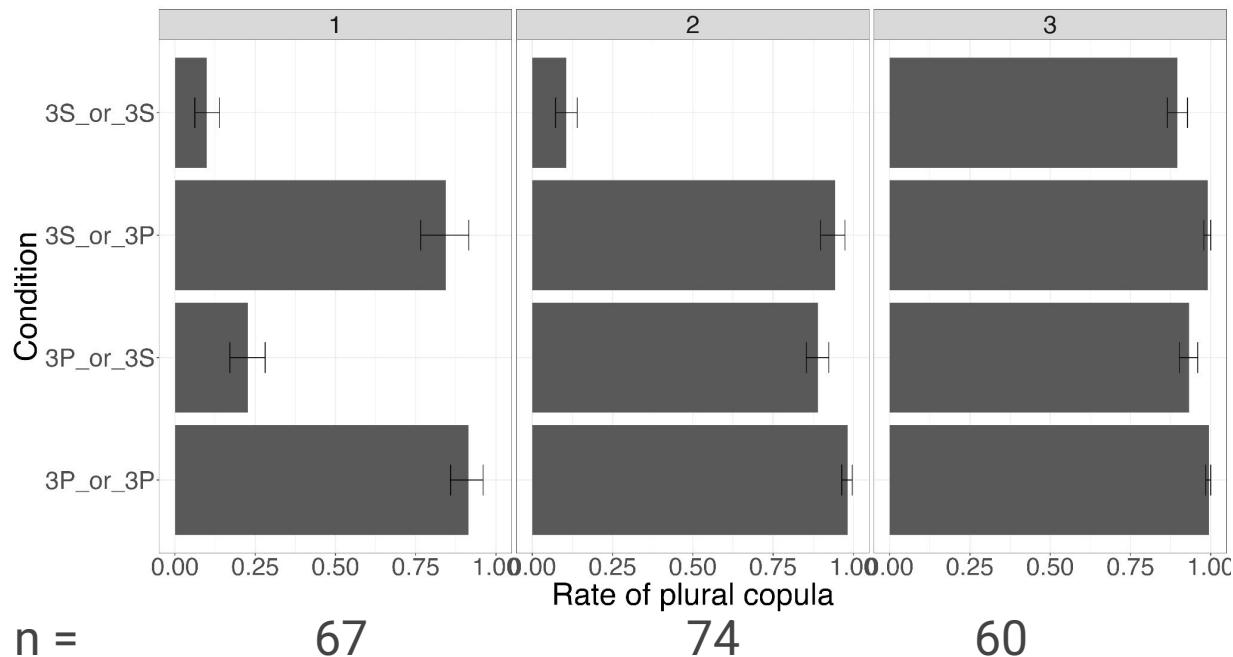
The responses show clear bi-modality for the “PL or SG” and “SG or SG” conditions.

But for “PL or PL” and “SG or PL”, participants overwhelmingly favored plural agreement.



Experiment 1: Results

We did a k-means clustering analysis. Three distinct clusters emerged.*



Cluster #1: close-disjunct agreement

Cluster #2: plural, except for SG or SG

Cluster #3: always plural

n = 67 74 60
*: the optimal number of clusters is determined by the elbow method and the silhouette method

Prediction verified

1

Either John or Mary **is** in the room.

Either John or the students **are** in the room

Either the students or John **is** in the room.

Either the students or the teachers **are** in the room.

2

Either John or Mary **is** in the room.

Either John or the students **are** in the room.

Either the students or John **are** in the room.

Either the students or the teachers **are** in the room.

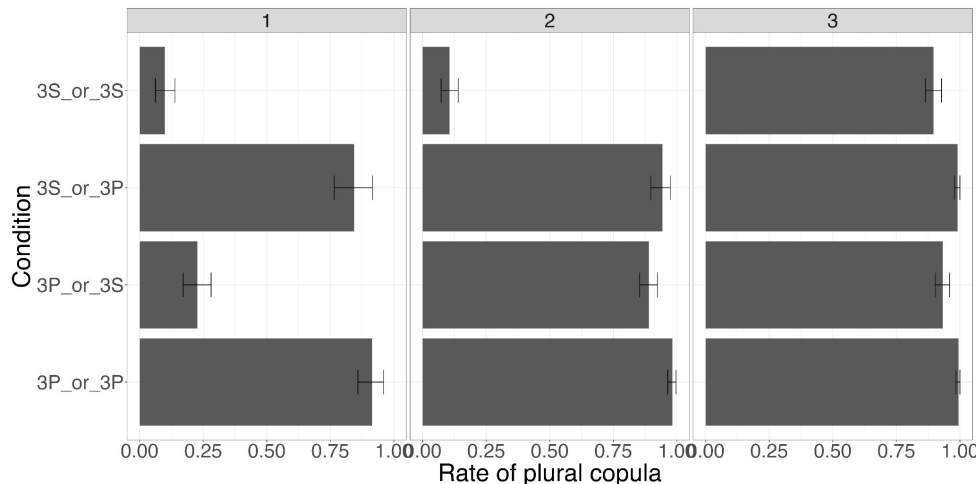
3

Either John or Mary **are** in the room.

Either John or the students **are** in the room.

Either the students or John **are** in the room.

Either the students or the teachers **are** in the room.



Experiment 1: Interim Discussion

- We showed experimental evidence for 3 distinct response patterns among the population for S-V agreement with disjointed subject.
- The 3 response clusters can be nicely captured by an OT-based account of disjunction agreement based on Bresnan (2001):

Lexicon: *is* (SG), *are* (Null)

Constraints: *SG, AGREE_{close}, AGREE_{each}

Person agreement

Exp.1 examined number (SG vs. PL) agreement.

Does the pattern extend to person (1st, 2nd, 3rd person) agreement?

Lexicon

am: 1SG

is: 3SG

are: Null

Constraints

*SG: No singular copula/verb

AGREE_{close}: Verb/copula needs to agree with closer disjunct **in number & person.**

AGREE_{each}: Verb/copula needs to agree with each disjunct **in number & person.**

Scan for tableau:



Person agreement: predictions

Cluster #1:

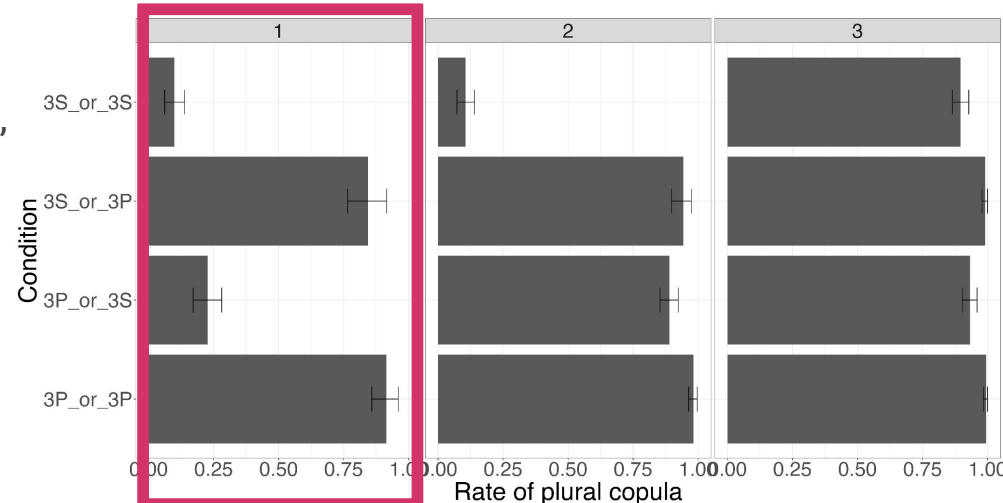
$\text{AGREE}_{\text{close}} \gg *SG$

Prediction:

If the second disjunct is 1SG, the copula is “am”

If the second disjunct is 3SG, the copula is “is”

Otherwise, the copula is “are”



Scan for tableau:



Person agreement: predictions

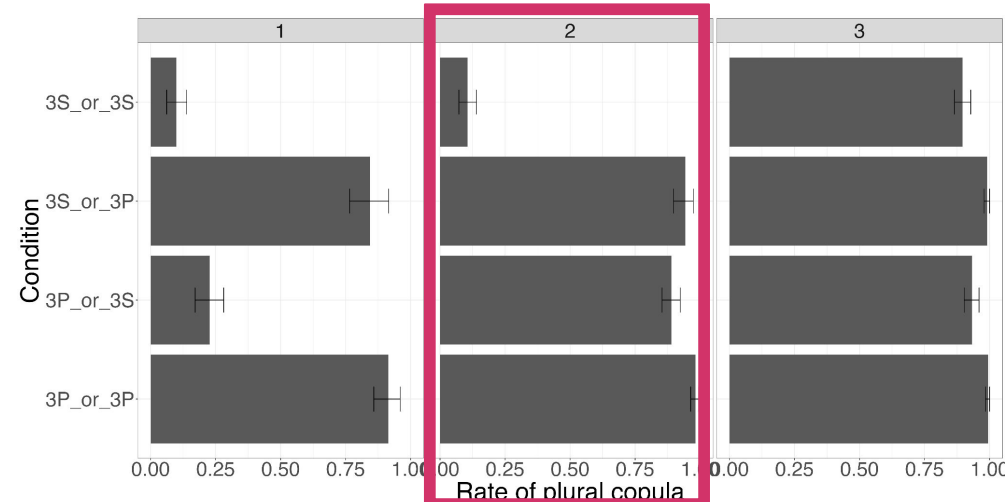
Cluster #2:

$\text{AGREE}_{\text{each}} \gg *SG \gg \text{AGREE}_{\text{close}}$

Prediction:

For “3SG or 3SG”, the copula is “is”

Otherwise, the copula is “are”



Scan for tableau:



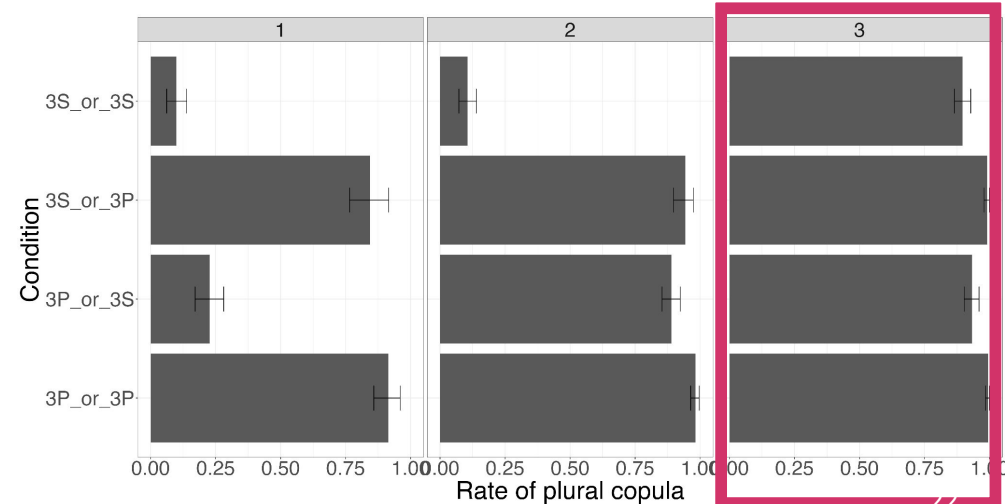
Person agreement: predictions

Cluster #3

*SG >> {AGREE_{close}, AGREE_{each}}

Prediction:

For all conditions, the favored copula is “are”.



Experiment 2: Methods and Stimuli

N = 298 (prolific, self-reported native English speakers). Similar interface as Exp.1

14 conditions:

Either {3SG, 3PL, 1SG, 1PL} or {3SG, 3PL, 1SG, 1PL} ($4 \times 4 = 16$ conditions)

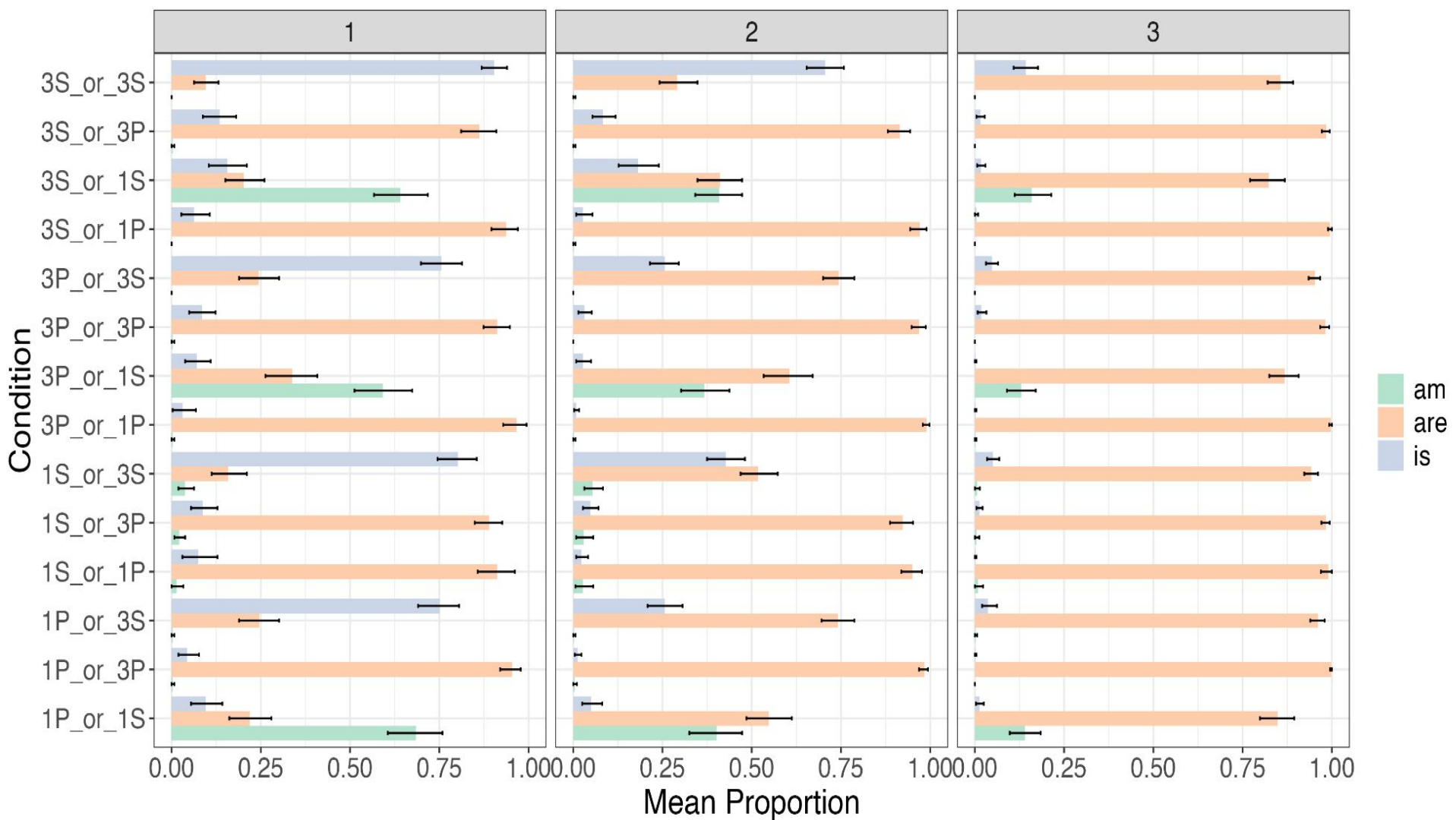
Minus 2 conditions: Either 1SG or 1SG (either I or I); Either 1PL or 1PL (either we or we)

70 items, 130 trials in total; 60 fillers

Experiment 2: Results

The 3rd person conditions replicated the 3 clusters identified in Exp.1.

We classified the speakers into 3 clusters based on their 3rd person responses, and examined their responses to the rest of the conditions.



Experiment 2: Interim Discussion

Experiment 2 provides further corroboration for our account of disjunction agreement:

- For all three clusters, the most preferred copula choices for all conditions exactly match our prediction.
- However, response preferences are less clear-cut than in Exp.1, potentially due to fatigue, since Exp.2 is much longer.
- It is also possible that some cluster 1 speakers are misclassified as cluster 2 by our algorithm, giving rise to the less clear-cut preference in cluster 2.

Discussion

Empirical contribution:

- We documented systematic inter-speaker variation in copular agreement with disjointed subjects
- There appears to be 3 distinct response patterns among the population.

Theoretical contribution:

- Disjunction agreement is **not a grammatical lacuna** where “anything goes” (Foppolo & Staub 2020)
- The 3 response patterns are nicely captured by our optimality-theoretic account
- This study adds to the list of OT-based accounts of optimal agreement selection (Bresnan 2001; Ortmann 2002; Badecker 2007; Himmelreich & Hartmann 2023, *inter alia*)
- This study showcases how OT can be a powerful explanatory tool for morphosyntactic microvariation

The Puzzle of Acquisition

How do speakers manage to acquire one of the three grammars?

- No correlation between response and demographic information.
- Corpus results do not mirror the experimental findings.

It seems like the input is so impoverished, that the speakers mostly just acquire a random ranking of constraints.

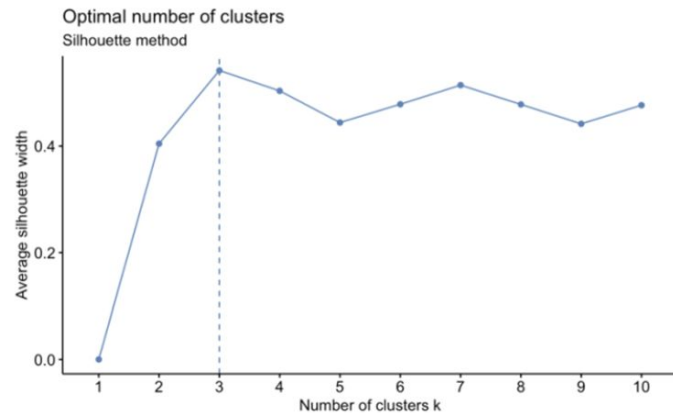
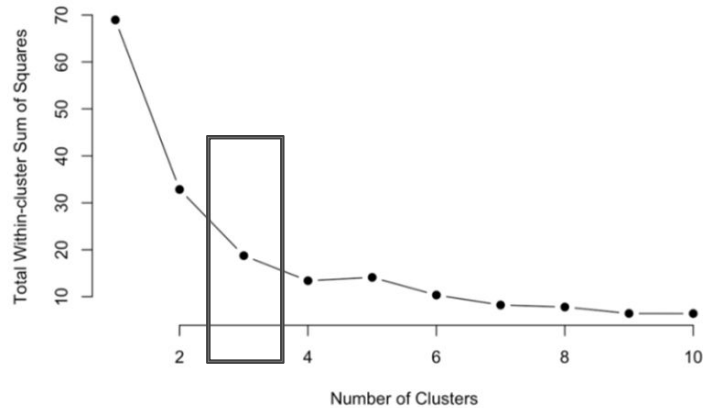
See Han, Lidz, and Musolino (2007) for a similar case in Korean V-raising.



Thank you!

Appendix: Determining the optimal number of clusters

We did a k-means clustering analysis. There seem to be three distinct clusters.



The Silhouette Method

$$a(i) = \frac{1}{|C_I| - 1} \sum_{j \in C_I, i \neq j} d(i, j)$$

$$b(i) = \min_{J \neq I} \frac{1}{|C_J|} \sum_{j \in C_J} d(i, j)$$

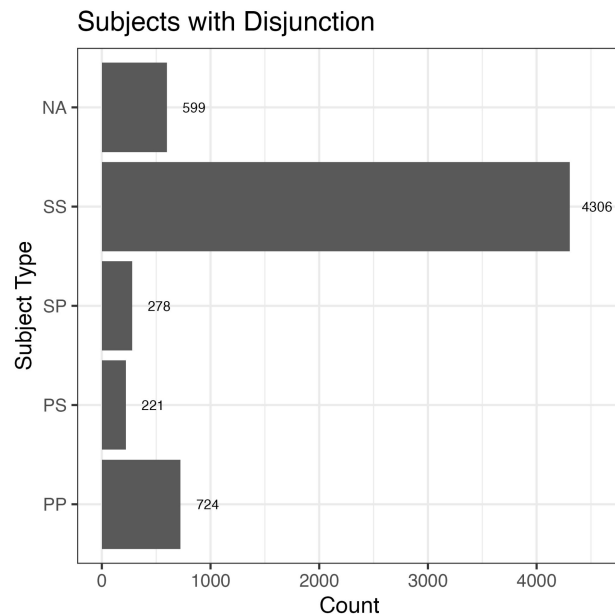
$$s(i) = \begin{cases} 1 - a(i)/b(i), & \text{if } a(i) < b(i) \\ 0, & \text{if } a(i) = b(i) \\ b(i)/a(i) - 1, & \text{if } a(i) > b(i) \end{cases}$$

Preliminary corpus analysis

- Scraped 384 million words of the common crawl (archive of internet text from 2017) and extracted subjects with two disjuncts
 - Number, verb agreement tags taken from parser
 - Exclude examples with Existential There, or modals blocking agreement
- Found 23,374 examples, with 6,128 using copular “be”

Preliminary corpus analysis

- 9.77% of examples: parser cannot assign number to subject
 - Of these, 25.87% have “you” as subject
- SS subjects most common (70.26%)



Preliminary corpus analysis

- SS subjects appear infrequently with plural verb marking
 - Contrast with participants in group 3 from empirical data

