

Grant Goodall

University of California, San Diego

Experimenting with *wh*-movement in Spanish

Two goals

An analysis

word order in
Spanish *wh*-
questions

A technique

experimental
syntax

What is experimental syntax?

- non-linguist subjects
- a clearly defined task, with training and/or practice
- factorial design for the construction of sentences

Example of factorial design

extraction	+ that	-that
Subject	Who do you think that saw Mary?	Who do you think saw Mary?
Object	Who do you think that Mary saw?	Who do you think Mary saw?
None	Do you think that John saw Mary?	Do you think John saw Mary?

What is experimental syntax?

- non-linguist subjects
- a clearly defined task, with training and/or practice
- factorial design for the construction of sentences
- a counterbalanced and randomized sentence list

Example of counterbalanced list

	Set A	Set B	Set C	Set D	Set E	Set F
Type 1	A1	B1	C1	D1	E1	F1
Type 2	A2	B2	C2	D2	E2	F2
Type 3	A3	B3	C3	D3	E3	F3
Type 4	A4	B4	C4	D4	E4	F4
Type 5	A5	B5	C5	D5	E5	F5
Type 6	A6	B6	C6	D6	E6	F6

Example of counterbalanced list

	Set A	Set B	Set C	Set D	Set E	Set F
Type 1	A1	B1	C1	D1	E1	F1
Type 2	F2	A2	B2	C2	D2	E2
Type 3	E3	F3	A3	B3	C3	D3
Type 4	D4	E4	F4	A4	B4	C4
Type 5	C5	D5	E5	F5	A5	B5
Type 6	B6	C6	D6	E6	F6	A6

Example of counterbalanced list

	Set A	Set B	Set C	Set D	Set E	Set F
Type 1	A1	B1	C1	D1	E1	F1
Type 2	F2	A2	B2	C2	D2	E2
Type 3	E3	F3	A3	B3	C3	D3
Type 4	D4	E4	F4	A4	B4	C4
Type 5	C5	D5	E5	F5	A5	B5
Type 6	B6	C6	D6	E6	F6	A6

Subject #1

What is experimental syntax?

- non-linguist subjects
- a clearly defined task, with training and/or practice
- factorial design for the construction of sentences
- a counterbalanced and randomized sentence list
- quantitative results
- statistical analysis of the results
- Note: This is not an “all or nothing” list!

Is it worth the trouble?

- more certainty about data
- more precision with subtle contrasts
- lingua franca of cognitive science

Particularly good for Romance syntax

- less of a tradition of careful field work
 - compared to Austronesian, Athabaskan, etc.
- mix of native and non-native linguists
 - data not always subject to same degree of scrutiny (compared to Chinese, etc.)
- significant dialect differences, often unacknowledged or not treated carefully

My focus here

- Experimental techniques allow us to capture **gradience** in judgments in a precise and reliable way.
- I will show:
 - such gradience exists (in previously unrecognized ways)
 - attempting to account for it ends up changing the way we do syntax.

Area: Inversion in *wh*-questions

(1)a. *Qué **Juan** leyó en la biblioteca?

b. Qué leyó **Juan** en la biblioteca?

‘What did Juan read in the library?’

- Well-known phenomenon in generative syntax since at least Torrego (1984).

Similar in many other Romance languages

Che cosa ha detto Maria? [Italian]

‘What did Maria say?’

Qu’a dit Jean? [French]

‘What did Jean say?’

Onde foi a Maria? [European Port.]

‘Where did Maria go?’

Què farà en Joan? [Catalan]

‘What will Joan do?’

Adónde fue María? [Spanish]

‘Where did María go?’

Unacceptability of (1a) co-varies with nature of intervening subject

(2)a. ?*Qué **tú** leíste en la biblioteca?

‘What did **you** read in the library?’

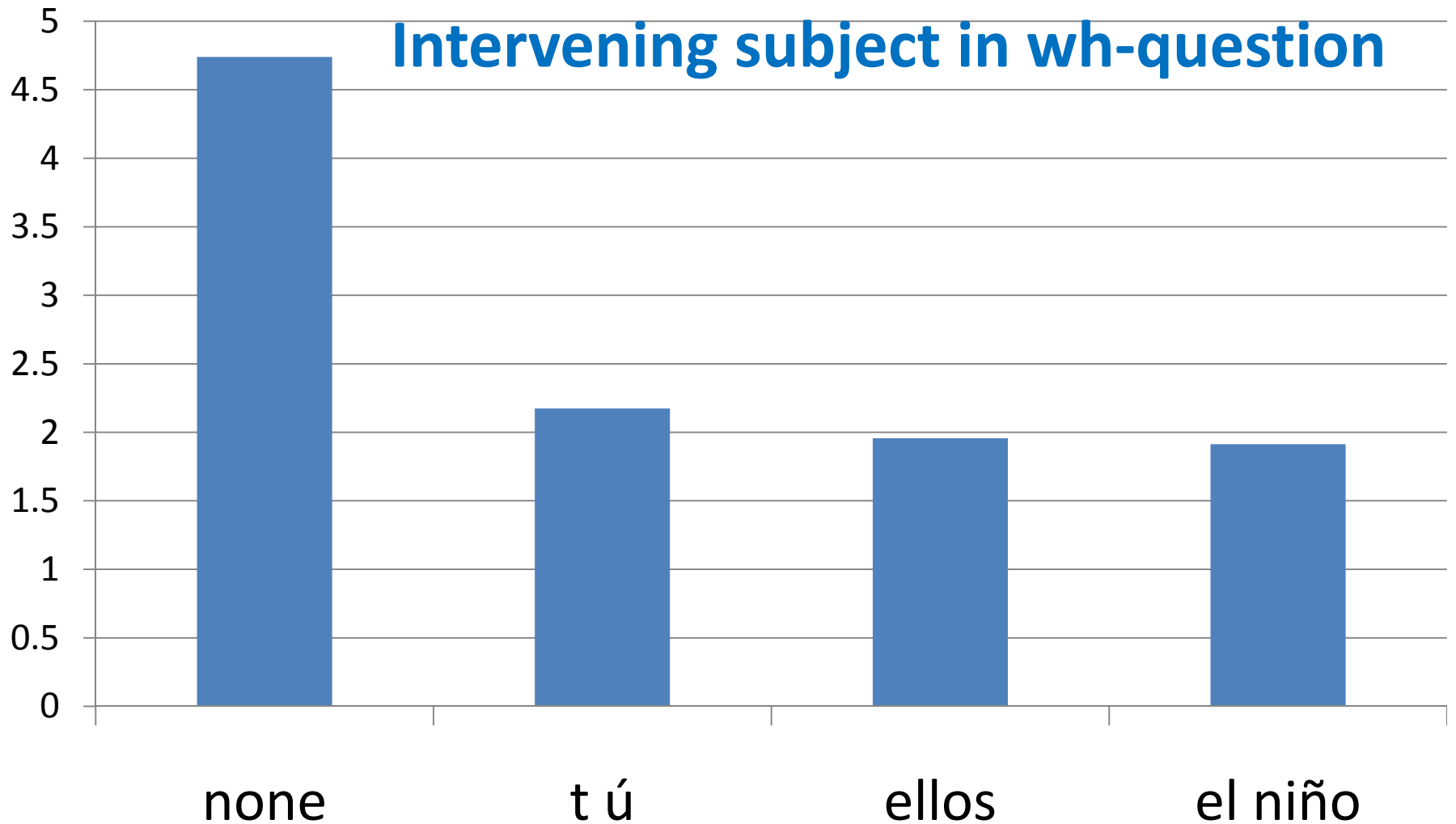
b. *Qué **ellos** leyeron en la biblioteca?

‘What did **they** read in the library?’

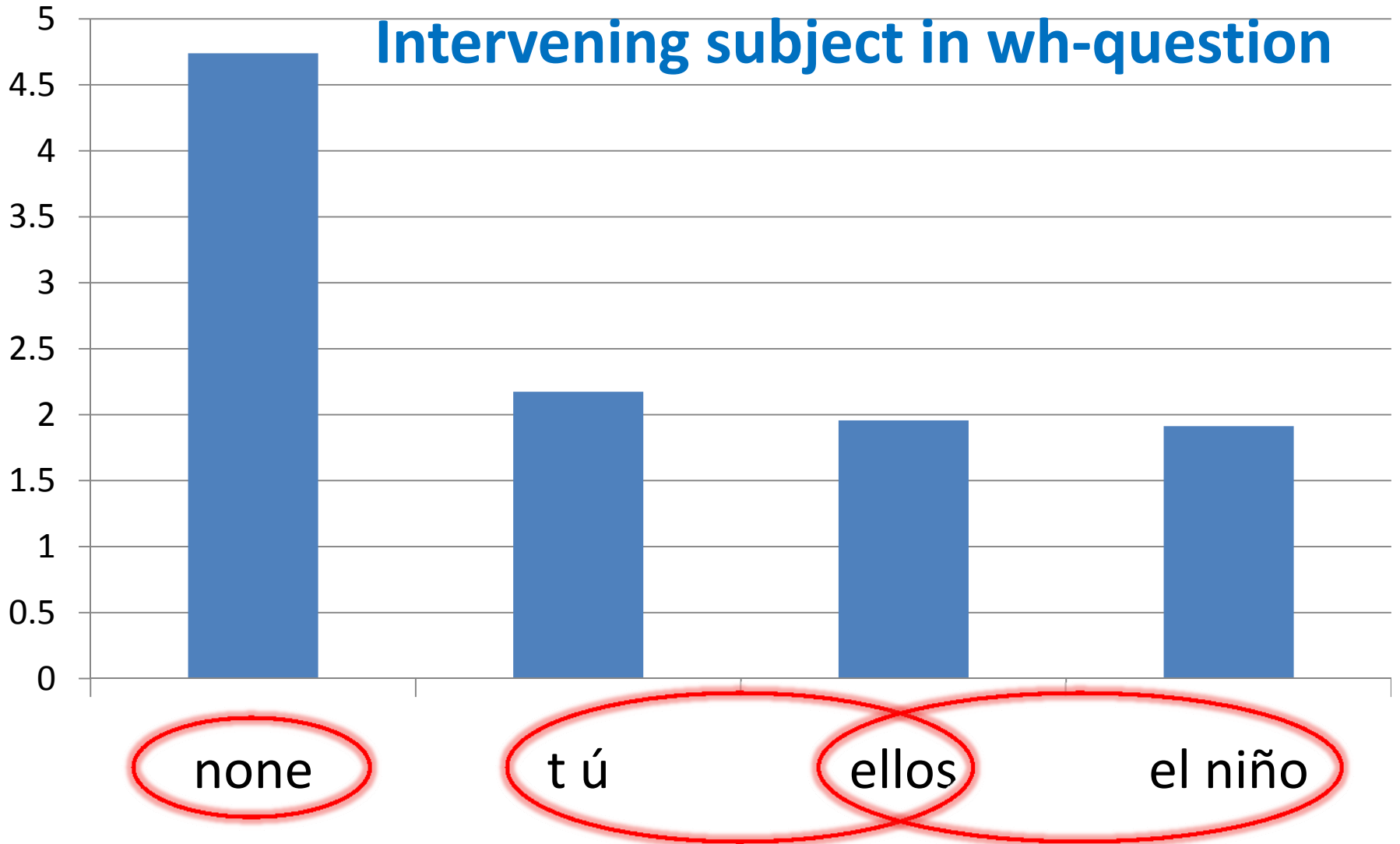
c. **Qué **el niño** leyó en la biblioteca?

‘What did **the child** read in the library?’

Intervening subject in wh-question



Intervening subject in wh-question



N=23

Generalization:

(3) 2p pronoun > 3p pronoun > lexical

- Seems unlikely to be due to syntax proper.
But what else could account for this?

Processing of filler-gap dependencies

- A'-dependencies place special demands on processor:
 - filler itself must be processed
 - filler must be stored in working memory while intervening material is processed
 - subcategorizing verb must be processed, gap posited, linked to filler.

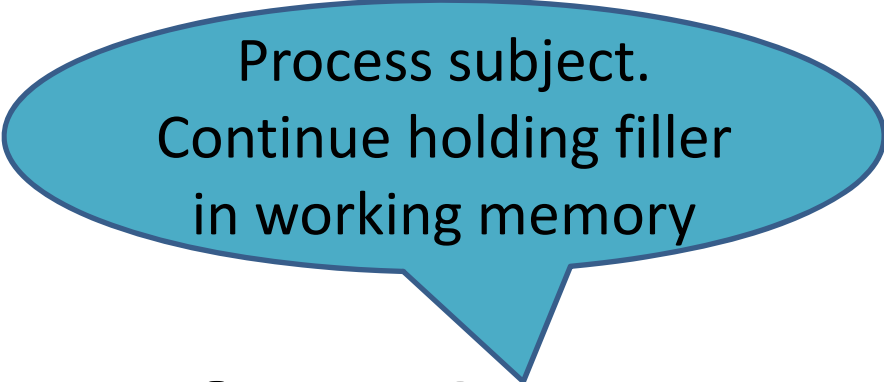
Qué



Process filler
Start search for gap

Qué

Qué el niño



Process subject.
Continue holding filler
in working memory

Qué el niño

Qué el niño leyó

Process verb and subject-verb
relation.

Retrieve filler from memory and
posit gap.

Qué el niño leyó

Processing of filler-gap dependencies

- Processing resources are limited
 - ability to keep filler in working memory is diminished by processing intervening material
 - so harder intervening material = harder gap resolution
 - lexical DP harder than pronoun
 - 3rd-person pronoun harder than 2nd-person pronoun
 - Processing reference and processing filler-gap:
same limited resources

This now predicts:

(4) wh V >

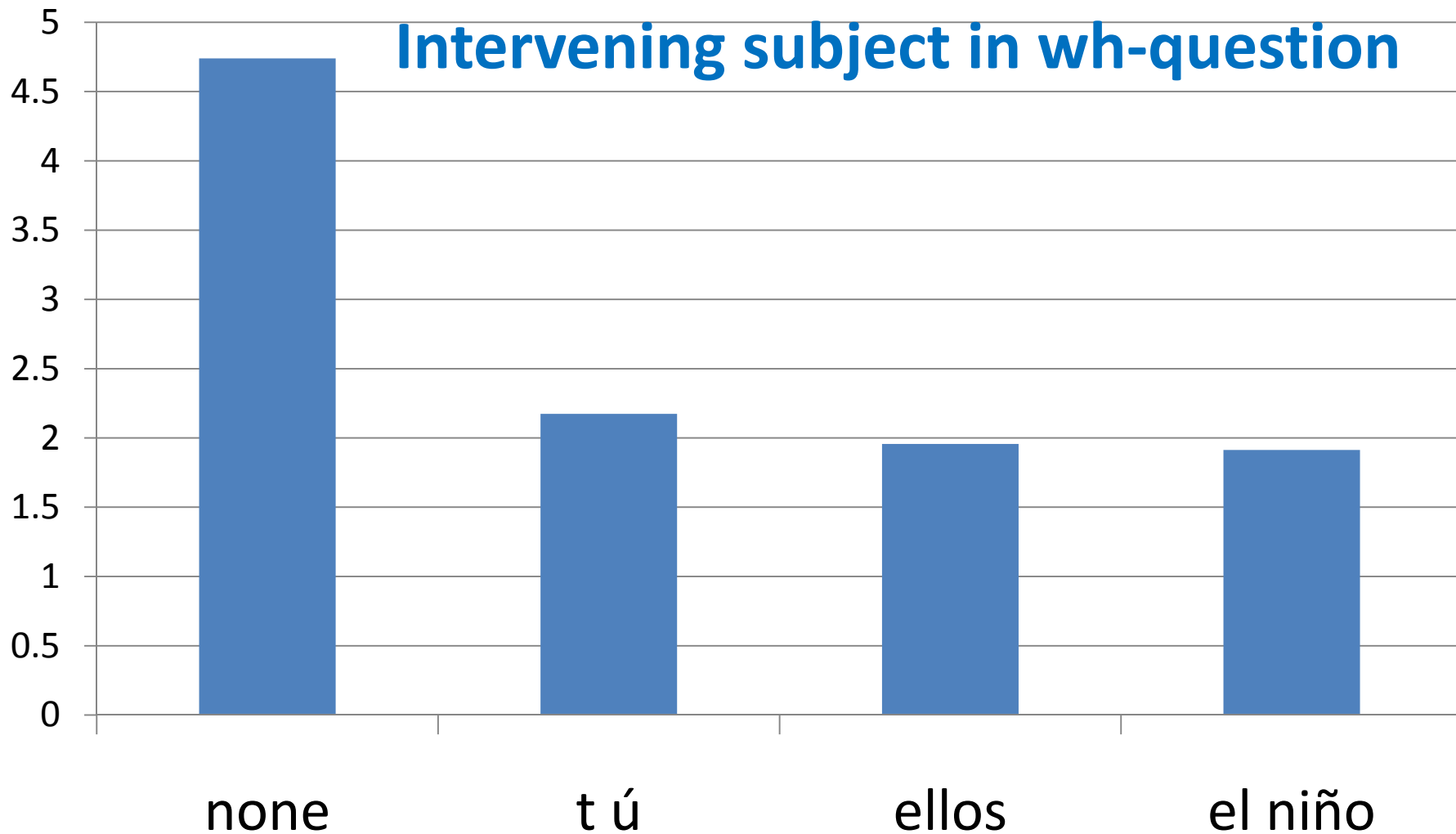
wh 2p V >

wh 3p V >

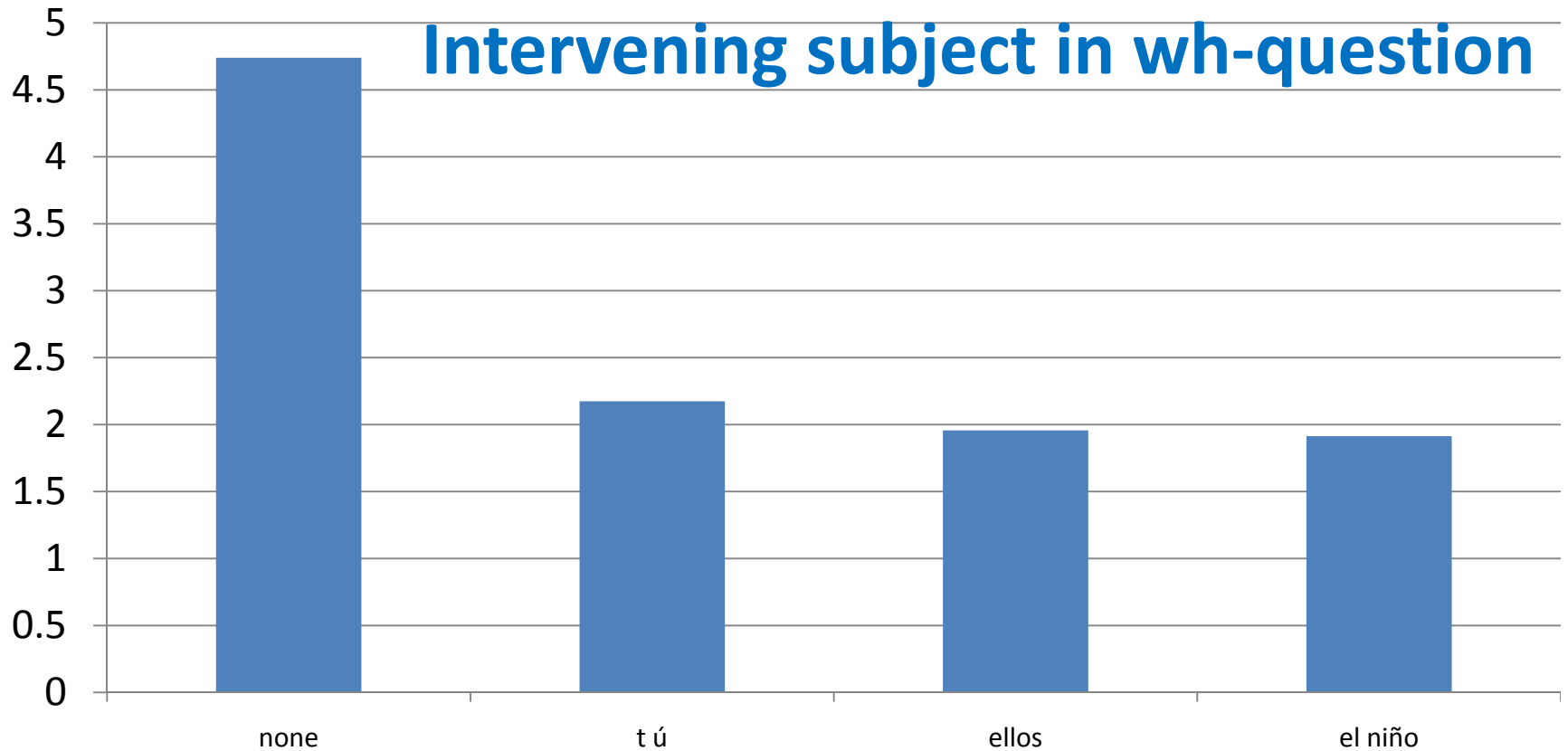
wh lexical V

- This matches exactly the hierarchy that we saw earlier.

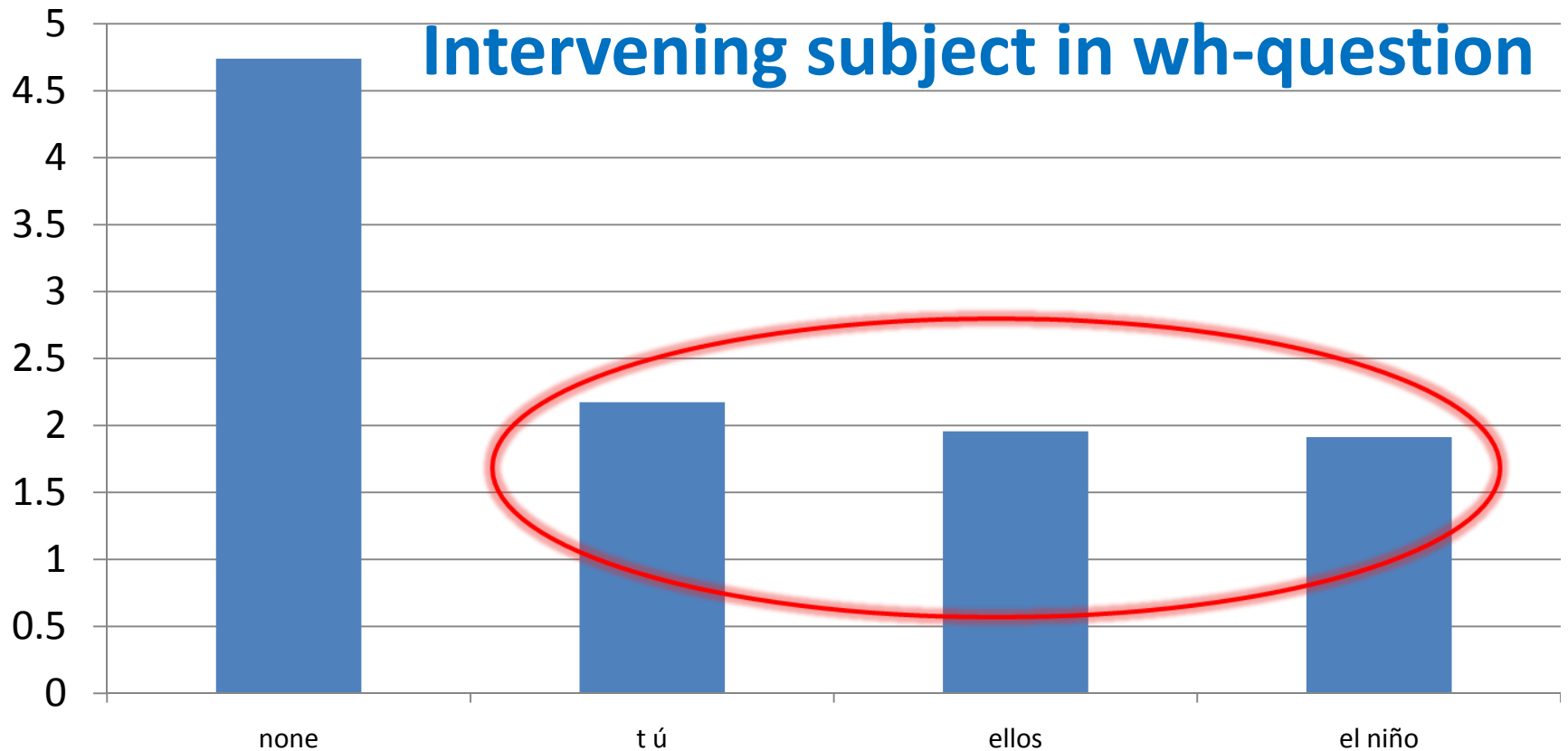
Intervening subject in wh-question



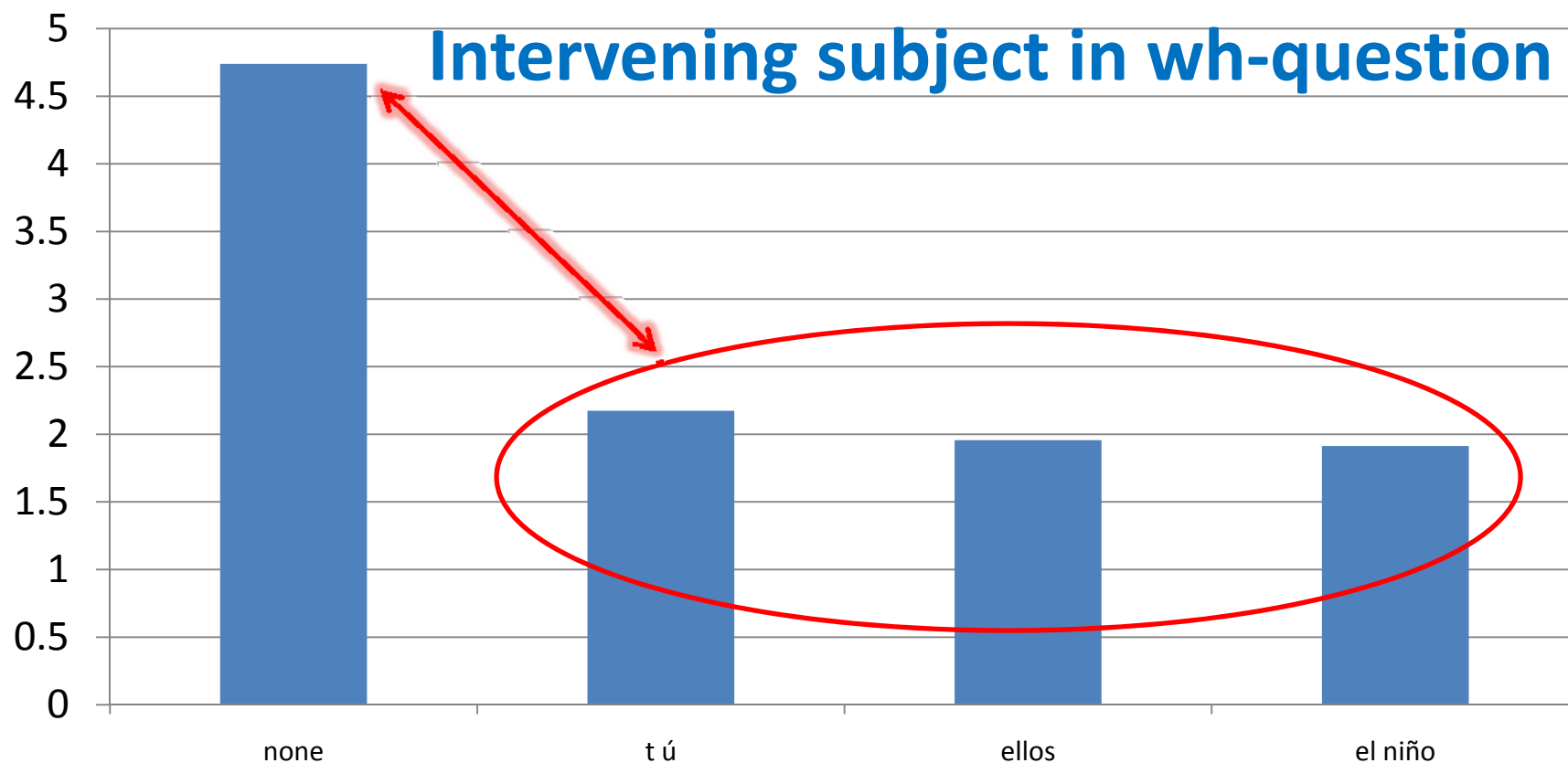
Two questions



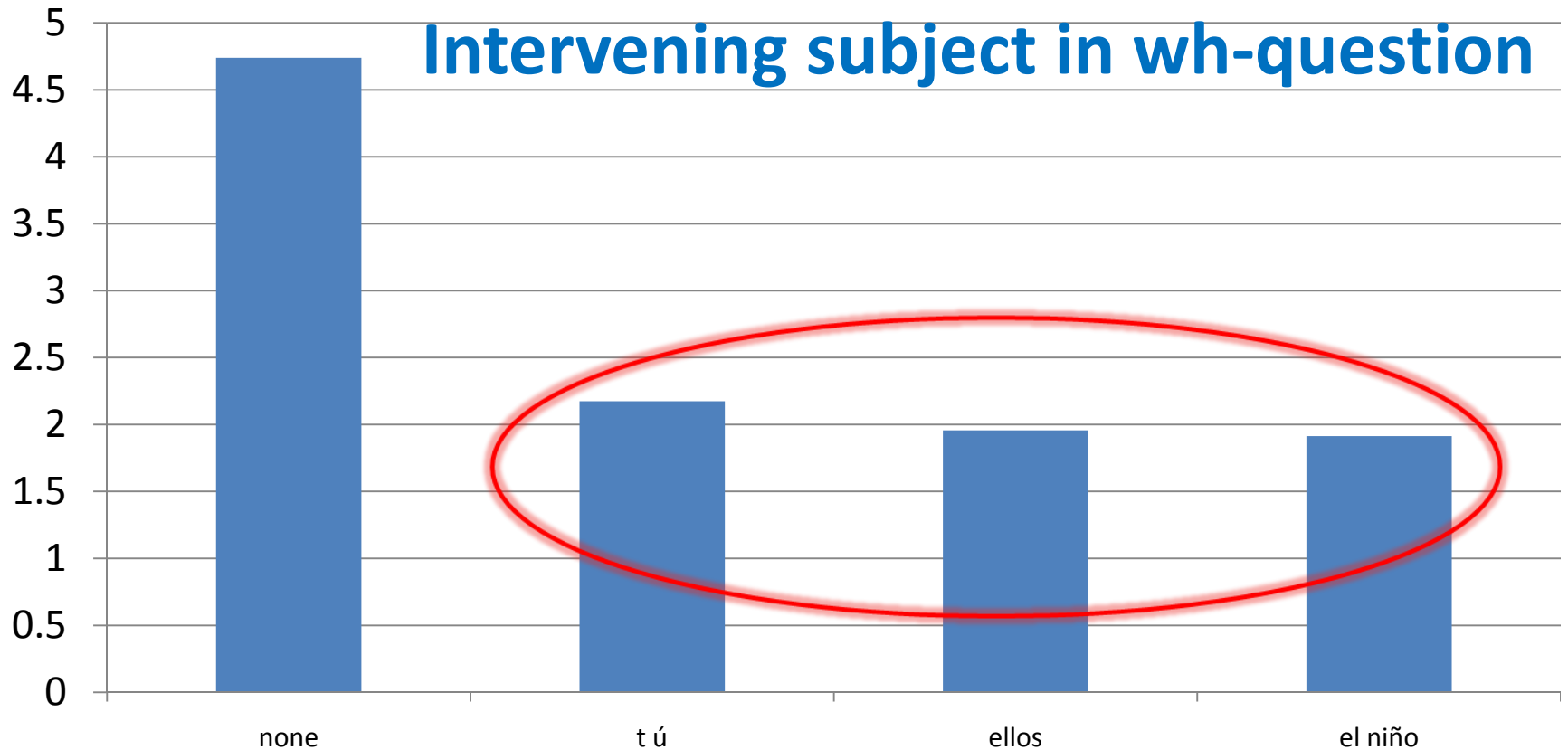
#1: Processing affects acceptability?



#2: Is this also a processing effect?



Let's start with this



Effect of filler: Complex *wh*-phrases

- Complex *wh*-phrases increase acceptability in ways that suggest a processing account (Kluender 1998):

(5) [Which of those boys] did you ask
whether anyone had talked to?

vs. ??[Which of those boys] did you ask
when the judge had talked to?

Effect of filler: Complex *wh*-phrases

- Bare *wh*-words in Italian prefer closest gap; complex *wh*-phrases do not (De Vincenzi 1991):

(6) Chi ha chiamato Giovanni?

‘Who called Giovanni?’ (preferred) or

‘Who did Giovanni call?’

(7) Quale ragazza ha chiamato Giovanni?

‘Which girl called Giovanni?’ or

‘Which girl did Giovanni call?’ (no preference)

Effect of filler: Complex *wh*-phrases

- Complex *wh*-phrases are less susceptible to superiority (Arnon et al. 2007):

(8) What did who read?

(9) Which book did which student read?

Effect of filler: Complex *wh*-phrases

- Complex *wh*-phrases are more accessible antecedents for pronouns.

(10) **Who** did Bradley send a rifle to when **he** was threatened?

(11) **Which guy** did Bradley send a rifle to when **he** was threatened?

(Frazier and Clifton 2002)

Conclusion

- Complex *wh*-phrases survive in working memory at a higher activation level (Kluender 1998, Hofmeister 2007).

We now predict

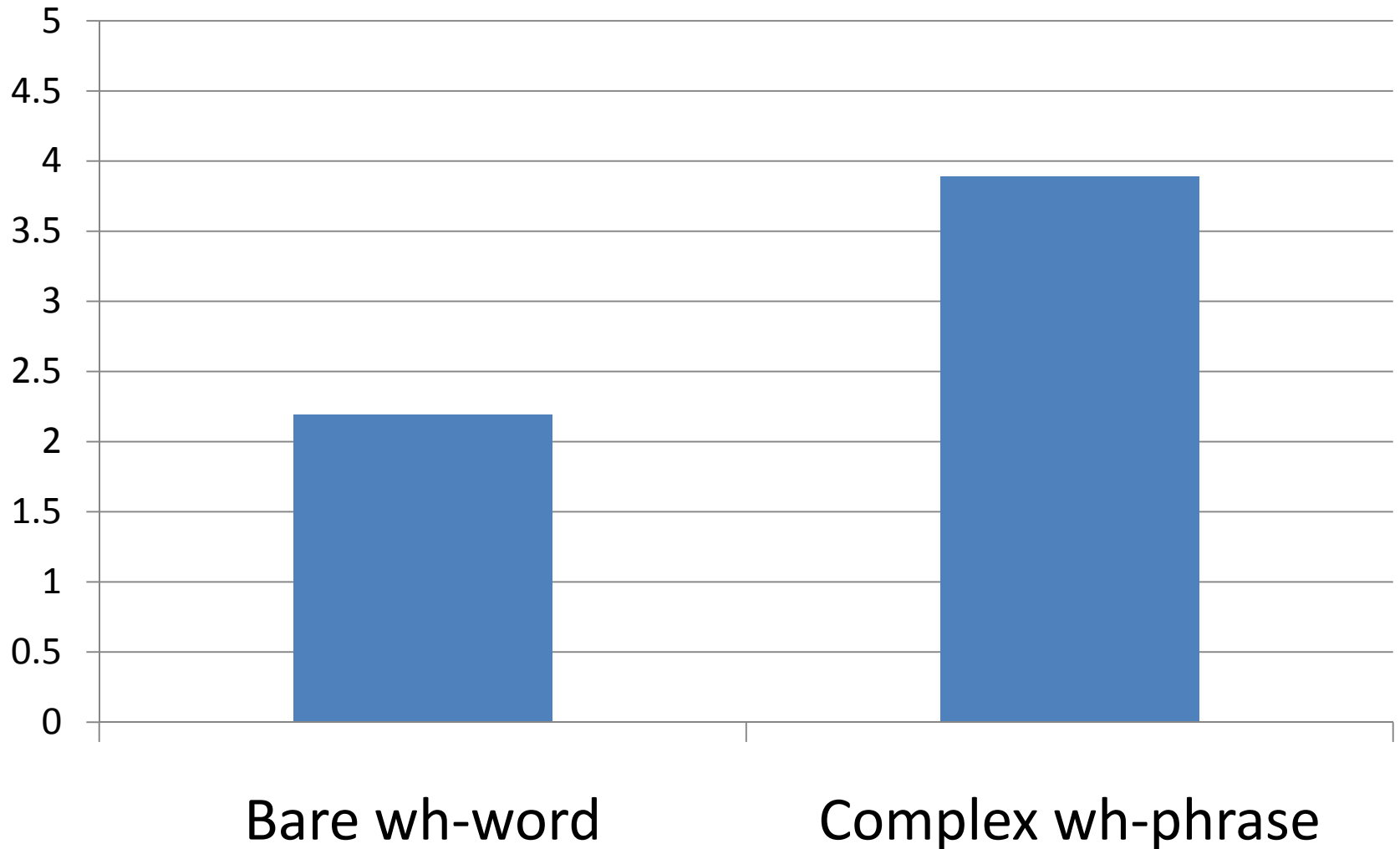
(12) ***Qué** Ana leyó?

‘What did Ana read?’

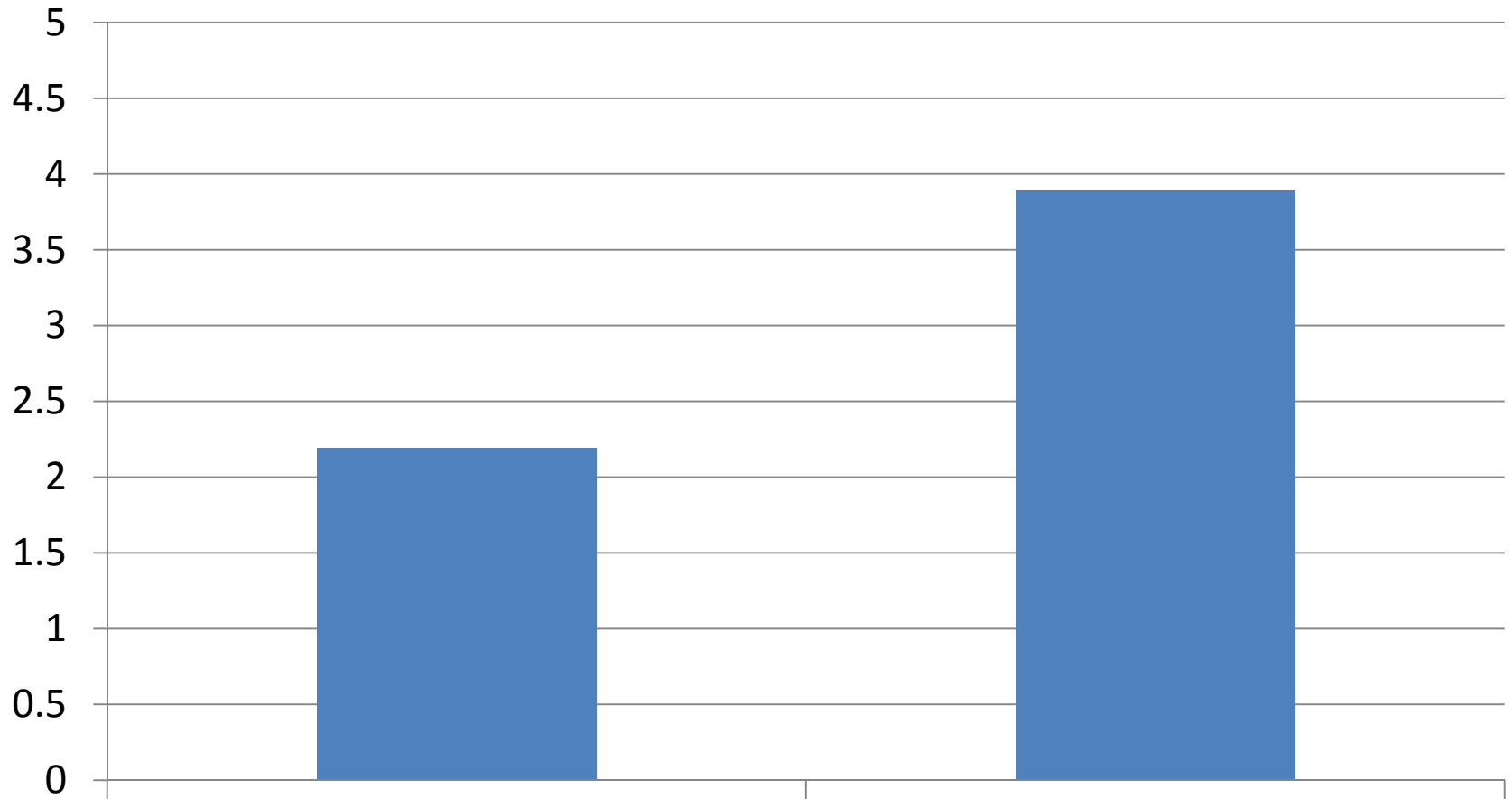
(13) **Cuáles de esos libros** Ana leyó?

‘Which of those books did Ana read?’

wh-phrase in *wh*-questions



wh-phrase in *wh*-questions



Bare wh-word

Complex wh-phrase

N=26

Effect of filler: Individual *wh*-words

- *Wh*-words that do not produce filler-gap dependency should be oblivious to intervening material:

(14)a. **Por qué** Miguel trabaja tanto?

b. **Por qué** trabaja tanto Miguel?

‘**Why** does Miguel work so much?’

Effect of filler: Individual *wh*-words

- Different *wh*-words may have different initial activation levels in working memory. This should mean varying ability to tolerate intervening material (more below).

Effect of filler: Individual *wh*-words

- **Interveners similar to filler** pose greater processing burden (Gordon, Hendrick and Johnson 2004): *what/who* vs. *where/when*.

(15)a. ***Qué** Juan leyó en la biblioteca?

‘**What** did Juan read in the library?’

b. ***A quién** María vio en el parque?

‘**Who** did María see in the park?’

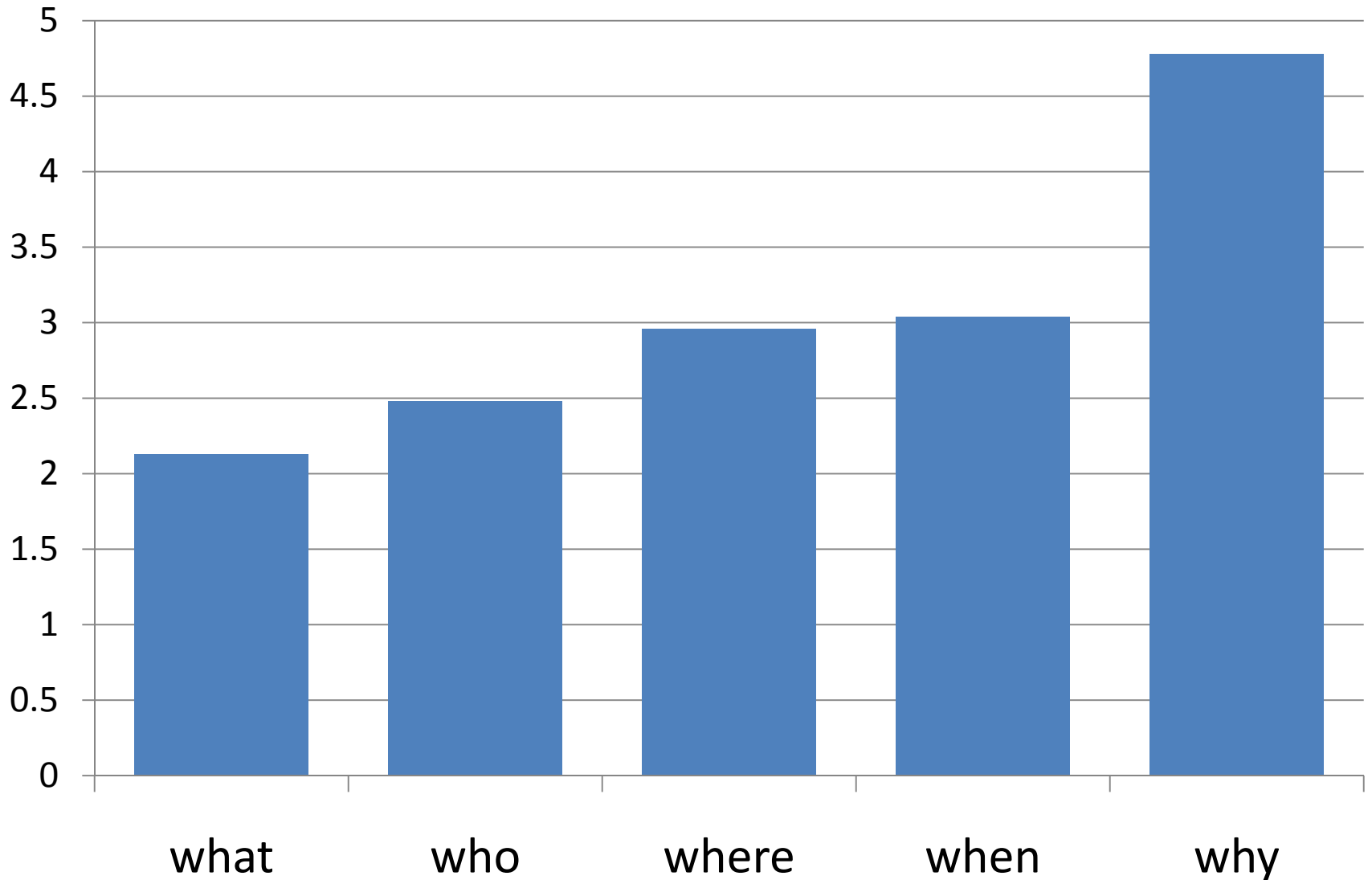
c. ?***Dónde** Ana compró el periódico?

‘**Where** did Ana buy the newspaper?’

d. ?***Cuándo** José escribió la carta?

‘**When** did José write the letter?’

Nominal vs. adverbial *wh*-word



Nominal vs. adverbial *wh*-word



Argument/adjunct asymmetry?

Common claim in literature, but:

- *Where* and *when* are still seriously degraded.
- Minimal pairs such as (16) do not show a contrast:

(16)a. *Dónde tú pusiste el libro?*

‘Where did you put the book?’

b. *Dónde tú compraste el libro?*

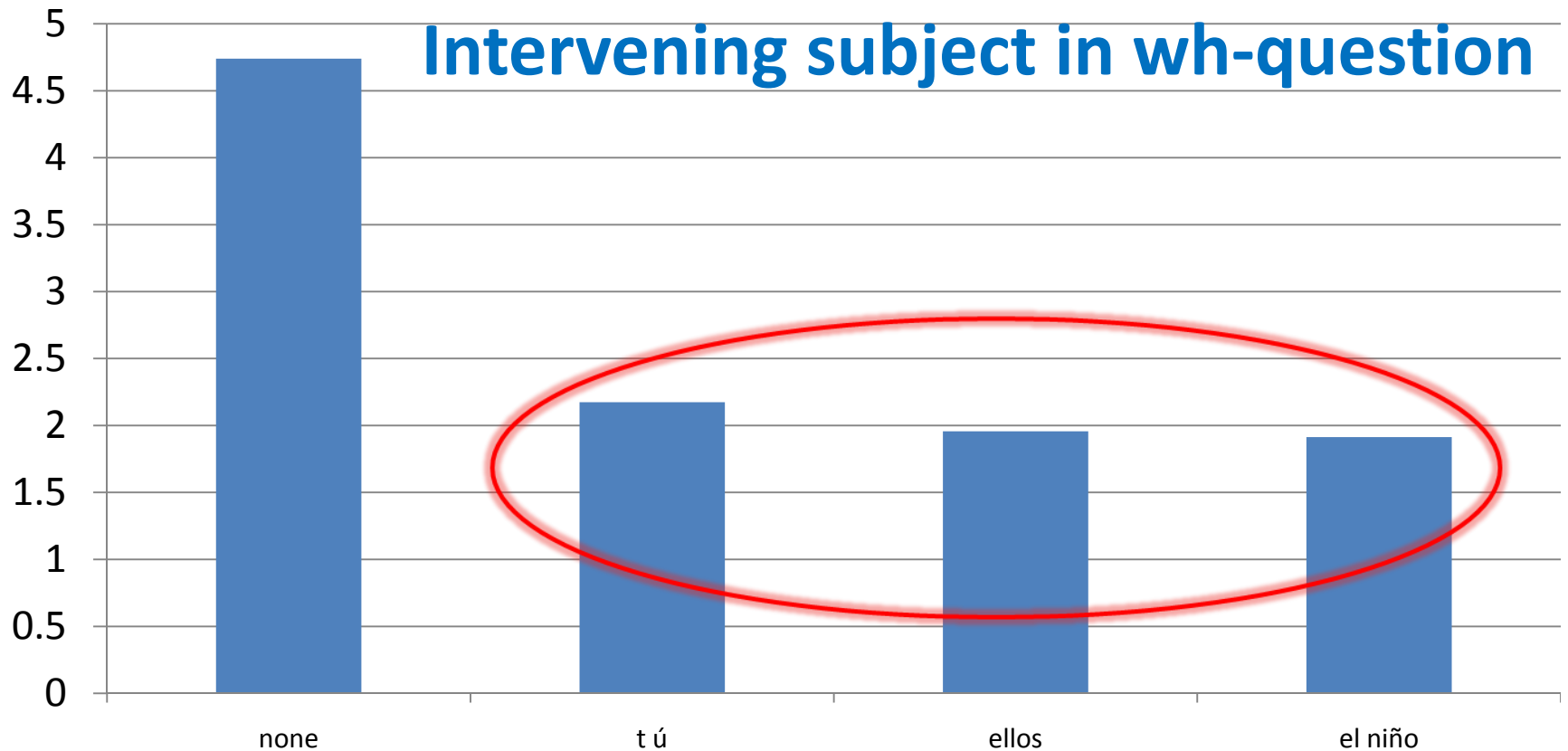
‘Where did you buy the book?’

Conclusion

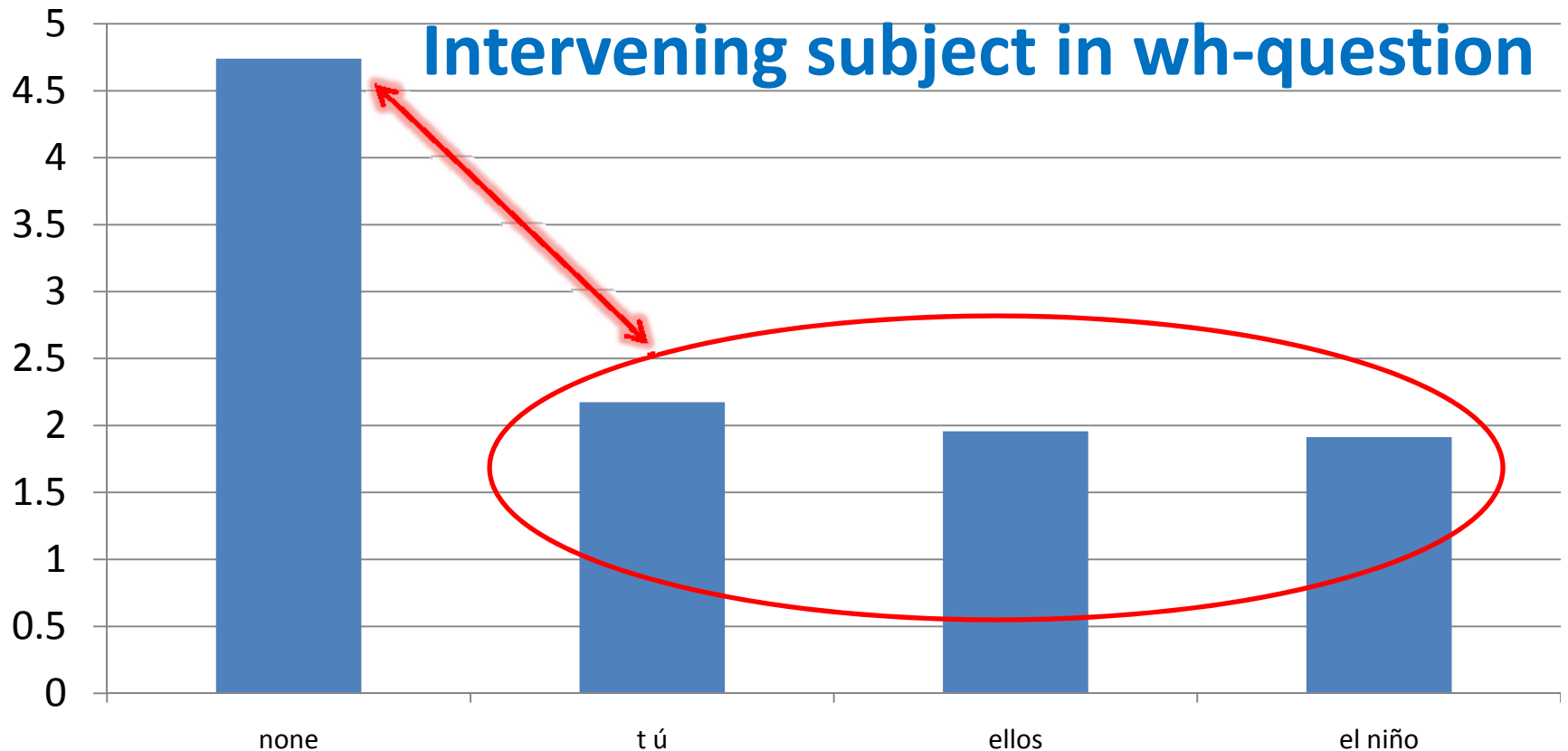
We have found the following hierarchy in ability to tolerate an intervening preverbal subject:

- (17) why >
- complex *wh*-phrase >
- (how >)
- where/when >
- what/who

Processing does affect acceptability



Can it also account for this?



Intervener vs. no intervener

(1)a. *Qué **Juan** leyó en la biblioteca?

b. Qué leyó **Juan** en la biblioteca?

‘What did Juan read in the library?’

- Our model clearly predicts the contrast in (1), but does it predict the **size** of the contrast?

Scenario that would have to occur

Qué Juan

Processing *Juan* consumes many
resources.

Difficult to maintain *qué* at high
activation level in working memory

Qué Juan

Difficult to retrieve *qué* and
resolve dependency.

Qué Juan leyó

Qué Juan leyó

Difficult to retrieve *qué* and resolve dependency.

Level of processing difficulty is so high that sentence is perceived as unacceptable.

Is it plausible that subject consumes so many resources?

- Overt pronoun subjects refer to relatively inaccessible discourse entities and appear to be more difficult to process (Callahan 2007). This effect may extend to lexical DPs as well.

(18) Cuando él trabaja, Juan no bebe.

‘When HE works, Juan doesn’t drink.’

(coreference discouraged) (Luján 1999)

Is it plausible that subject consumes so many resources?

Preverbal subjects have a default categorial (as opposed to thematic) interpretation. Subject refers to highly individuated discourse entity, plausibly more costly for processing.

(19)a. *Estudiantes llegaron.

b. Llegaron estudiantes.

‘Students arrived.’

(Byrne 1998)

Is it plausible that subject consumes so many resources?

Preverbal subjects have a default categorical (as opposed to thematic) interpretation. Subject refers to highly individuated discourse entity, plausibly more costly for processing.

(20) a. Podemos presentar una carta que redacte una estudiante.

b. ?? Podemos presentar una carta que una estudiante redacte.

“We can present a letter that a student can write.”
(Gutiérrez Bravo 2007)

Conclusion

Overt, preverbal subjects refer to discourse entities that are:

**relatively inaccessible and
highly individuated.**

This plausibly exacts a high processing cost (Warren and Gibson 2002).

Is this enough to account for the full depth of unacceptability of (1a)?

(1)a. *Qué **Juan** leyó en la biblioteca?

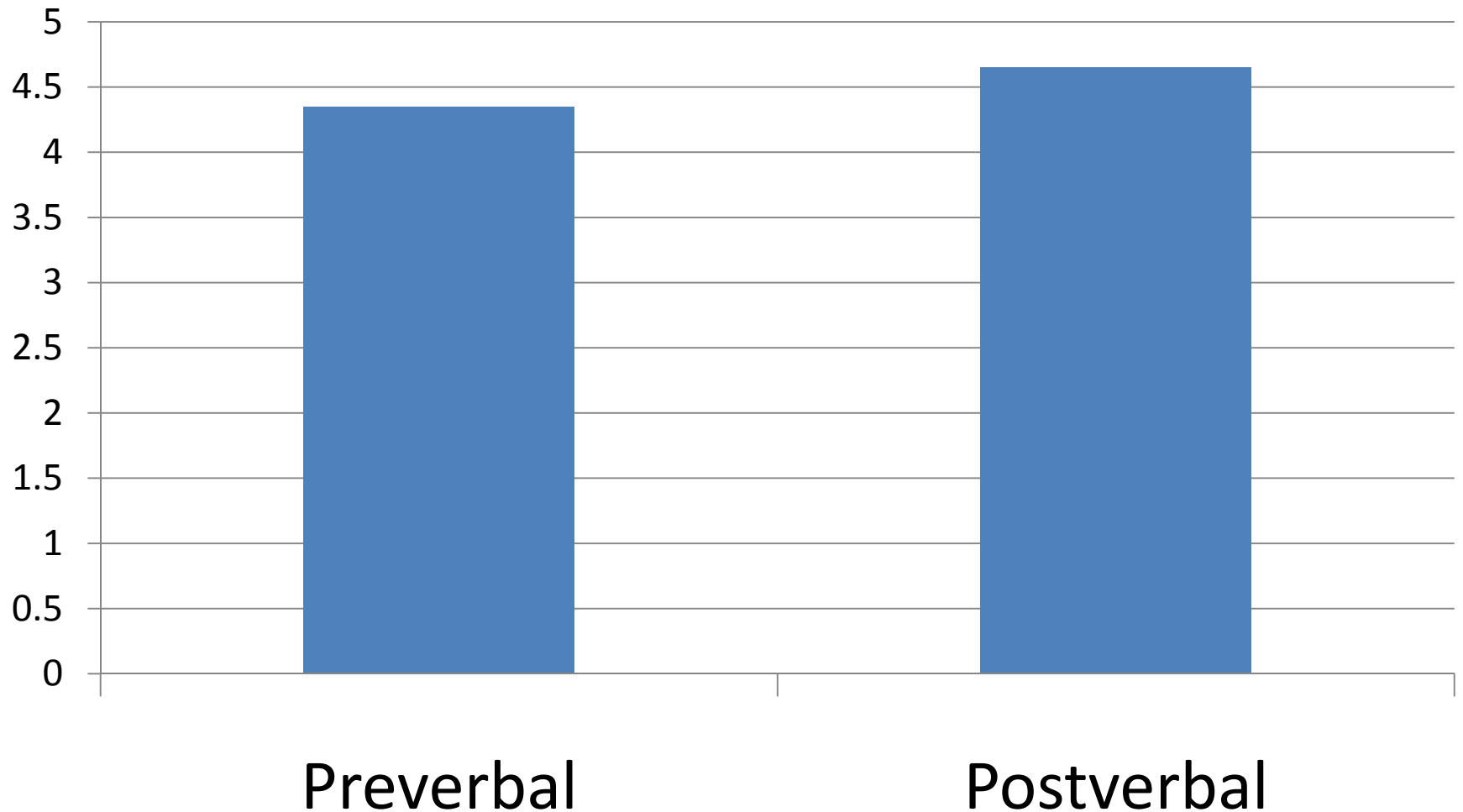
Perhaps, though effect is much weaker with extraction out of embedded clause:

(26)a. Qué dices que **tus papás** compraron?

b. Qué dices que compraron **tus papás**?

‘What do you say that your parents bought?’

Embedded subject in long-distance movement



N=26

Additional factors at work in the case of the matrix subject?

- Semantic/information structure incompatibility between preverbal subjects and interrogatives (Escandell Vidal 1999 and Gallego 2006)?
- Note that VSO order is marginal as a declarative, but possible as an interrogative:

(21)a. ?? Compró Ud. esta casa.

b. Compró Ud. esta casa?

Additional factors at work in the case of the matrix subject?

- *Wh*-questions with SV order do not have a “true interrogative” interpretation?
(Gallego 2006)

(22)a. Por qué Mario compró la casa?

b. Por qué compró Mario la casa?

‘Why did Mario buy the house?’

Additional factors at work in the case of the matrix subject?

- So there may be other factors that make the postverbal preference stronger in the main clause than in the embedded clause.

What we have seen so far

- Inversion effect largely due to independently needed processing considerations
- Other factors (semantics, information structure) may also play a role
- Syntax plays only an indirect role in the inversion effect: *wh*-movement, placement of subject
- This analysis made possible/encouraged by experimental techniques

This is all good...

In fact, it is **too** good.

We can explain why (1a) is **not** possible:

(1)a. *Qué **Juan** leyó en la biblioteca?

But what about languages where it **is** possible?

We can address this even within Spanish...

Variation in Spanish

Overt Pronoun Rate

16% Ciudad Juárez, Mexico (Strongman (1995))

20% Caracas, Venezuela (Bentivoglio (1987))

25% Madrid, Spain (Enríquez (1984))

38% Santiago, Chile (Cifuentes (1980-1))

39% Los Angeles, USA (Silva-Corvalán (1977))

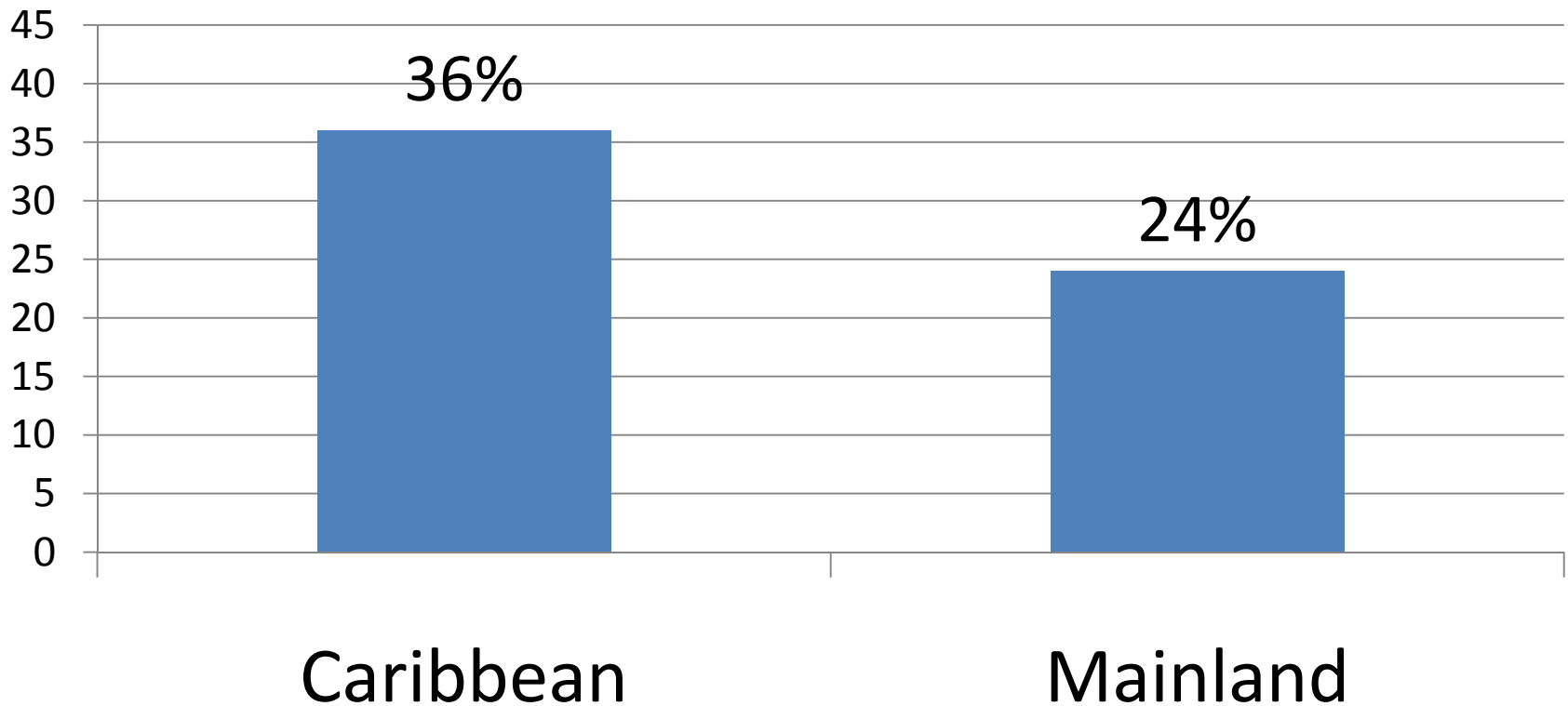
Variation in Spanish

These comparisons across corpora are suggestive, but not perfect.

Better...

Otheguy, Zentella and Livert (2007)

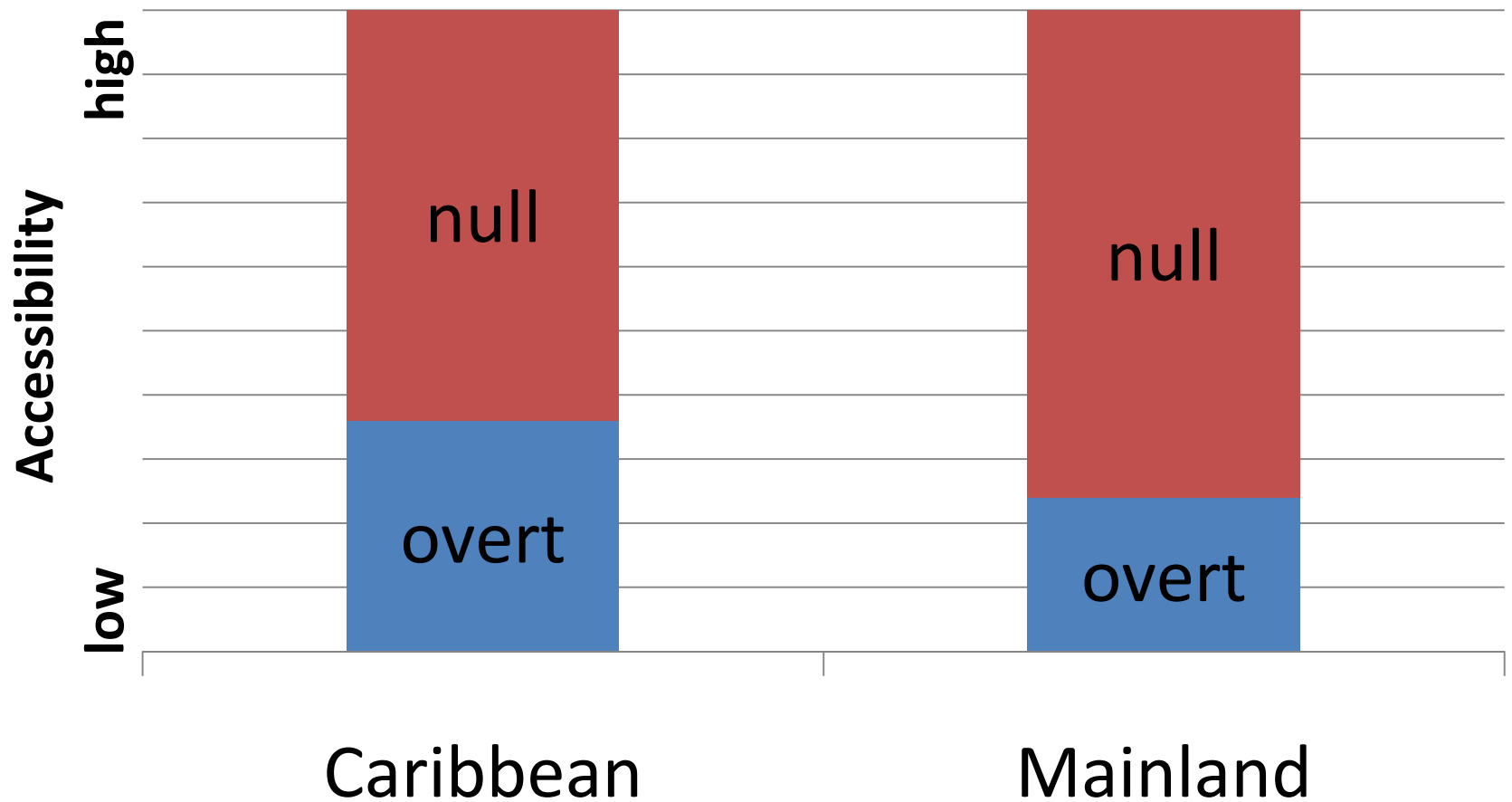
Overt Pronoun Rate



What are we to make of this?

- Overt pronouns: less accessible referents
- Can't be that Caribbeans systematically talk about less accessible individuals.
- Must be that for them, overt pronoun has more accessible referent than for mainlanders.

What are we to make of this?



Preverbal/postverbal subjects

Usual claim: Higher overt pronoun rate = more preverbal subjects.

If so, a subject doesn't need to be as individuated to be preverbal in these varieties.

In these varieties, overt preverbal subject will be:

more discourse-accessible

less individuated

than in mainland varieties.

And thus easier to process.

Clear prediction

- In these varieties, *Wh*-phrase will be better able to tolerate intervening subject.

Et voilà

- (23)a. Qué **yo** les voy a mandar a esos muchachos?
‘What am I going to send to those boys?’
- b. Qué número **tú** anotaste?
‘What number did you write down?’
- c. Qué **ese letrero** dice?
‘What does that sign say?’
- d. Cuánto **un medico** gana?
‘How much does a doctor make?’ (from Toribio (2000))

Hierarchy of acceptability in literature

- Data disagreements in literature, but Ordóñez and Olarrea (2006) note a hierarchy:

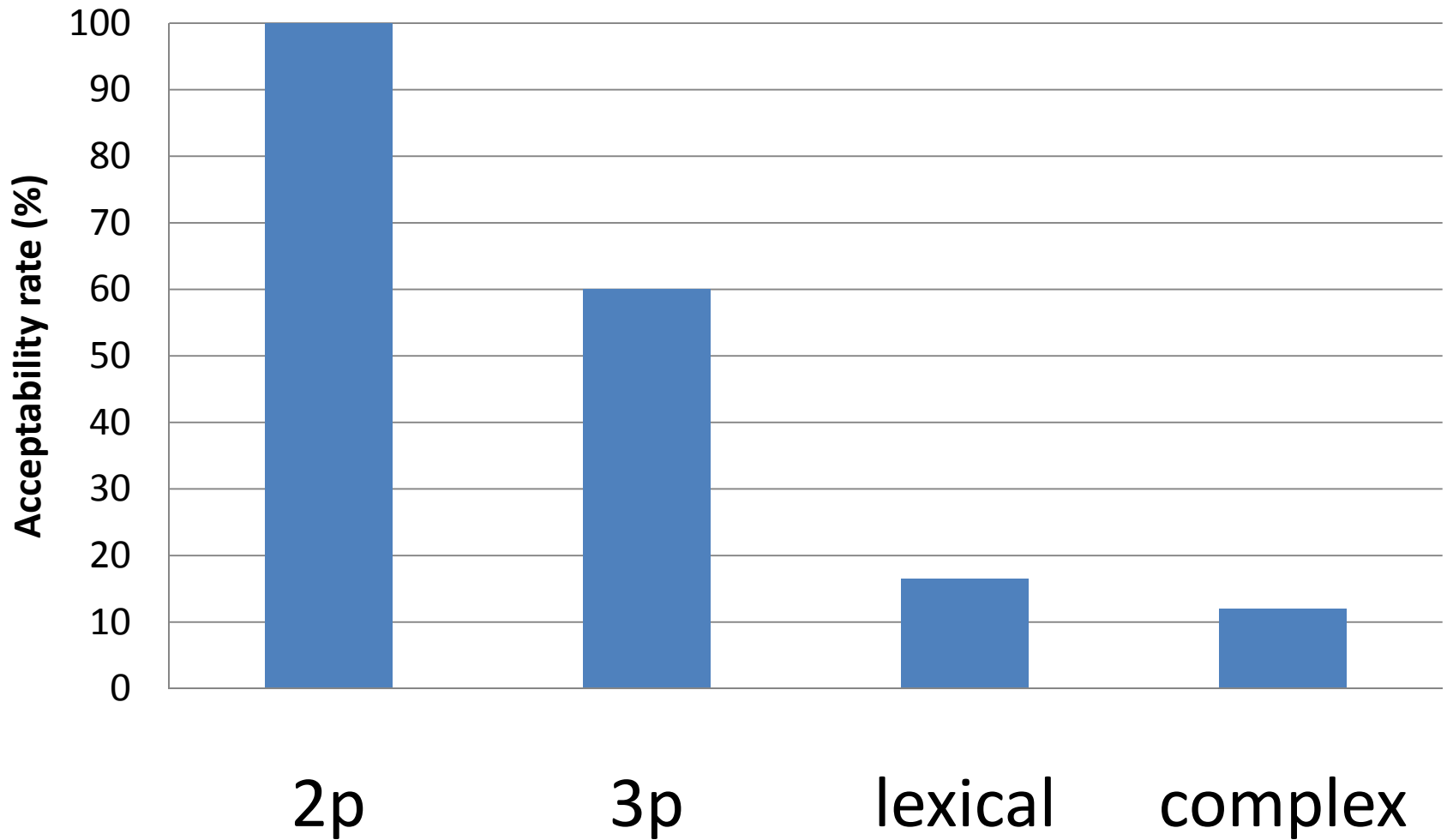
(24) 2p pronoun > 3p pronoun > lexical

- This is exactly the hierarchy we saw in standard Spanish.
- Disagreements about judgments = gradience

Gradience in Dominican Spanish

- Ordóñez and Olarrea (2006) find gradience in a survey of acceptability among Dominicans (N=65):

Acceptability of intervening subject in Dominican Spanish



How to talk bad in Dominican:

Stress, conjoin, or modify subject:

(25)a. *Qué **TÚ** quieres?

‘What do YOU want?’

b. *Qué **tú y él** comieron?

‘What did you and he eat?’

c. *Qué **tú mismo** comes?’

‘What do you yourself eat?’ (from Ordóñez and Olarrea (2006))

In all three of these cases, increased resources for processing subject would be expected.

Intervening subject is still dispreferred

- In corpus of Puerto Rico Spanish, only 10% of *wh*-questions with overt subject have the subject in an intervening position.

(Gutiérrez Bravo (2007))

- Consistent with idea that preverbal subjects cause processing difficulty, even in Caribbean.

Conclusion: Caribbean fits well with
our analysis.

- Properties of *wh*-questions follow from properties of overt preverbal subjects.

How it works in Caribbean:

Qué él

This subject has:
higher accessibility
lower individuation
than Mainland Spanish.
So it's easier to process.

Qué él

Relatively easy to retrieve *qué*
and resolve dependency.

Qué él leyó

Qué él leyó

Relatively easy to retrieve *qué*
and resolve dependency.

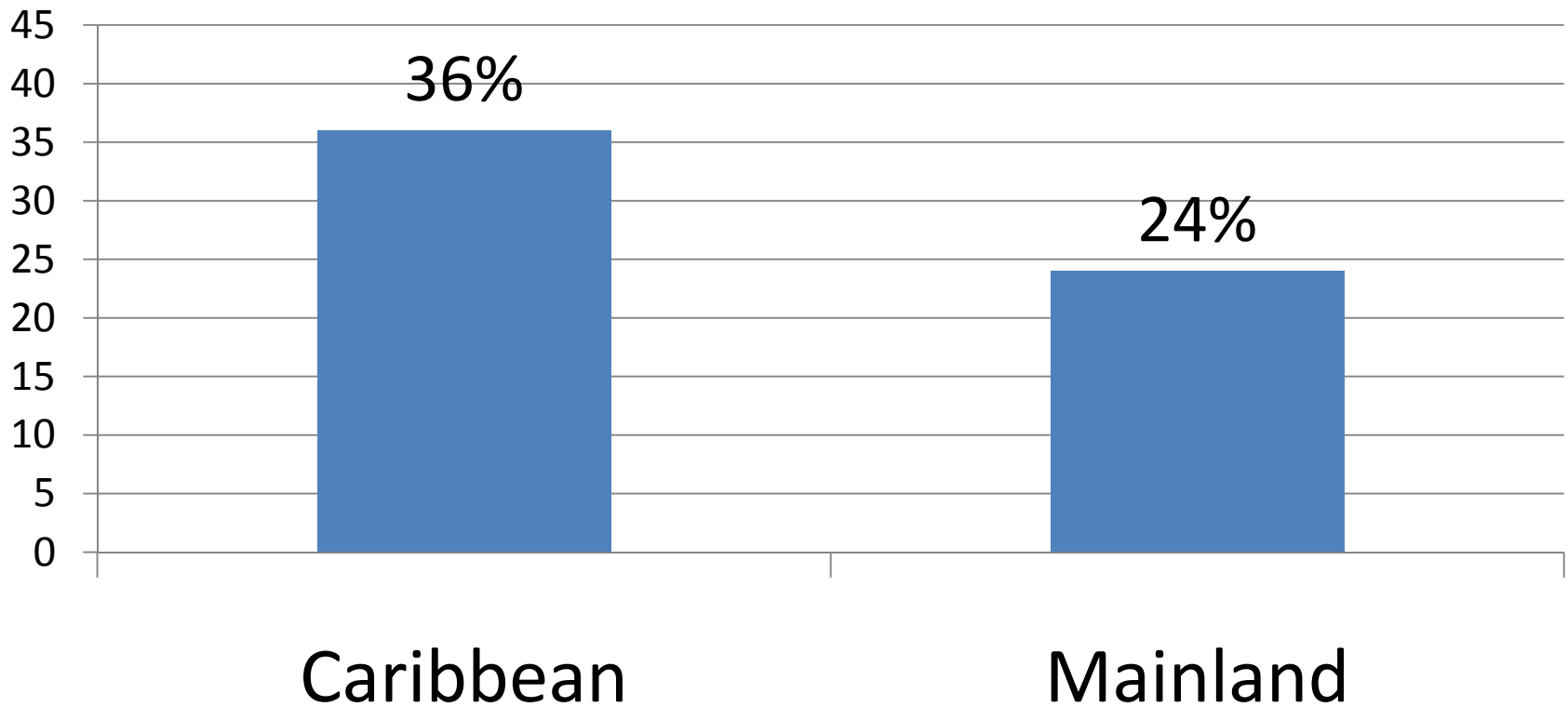
Acceptability is
thus higher than in
Mainland Spanish.

Aspects of this analysis to notice:

- Both dialects permit same choices for subject: null vs. overt, preverbal vs. postverbal.
- Discourse consequences differ, and processing consequences thus differ too.
- Locus of variation: How discourse makes use of possibilities afforded by the syntax.

So far: Caribbean vs. Mainland

Overt Pronoun Rate



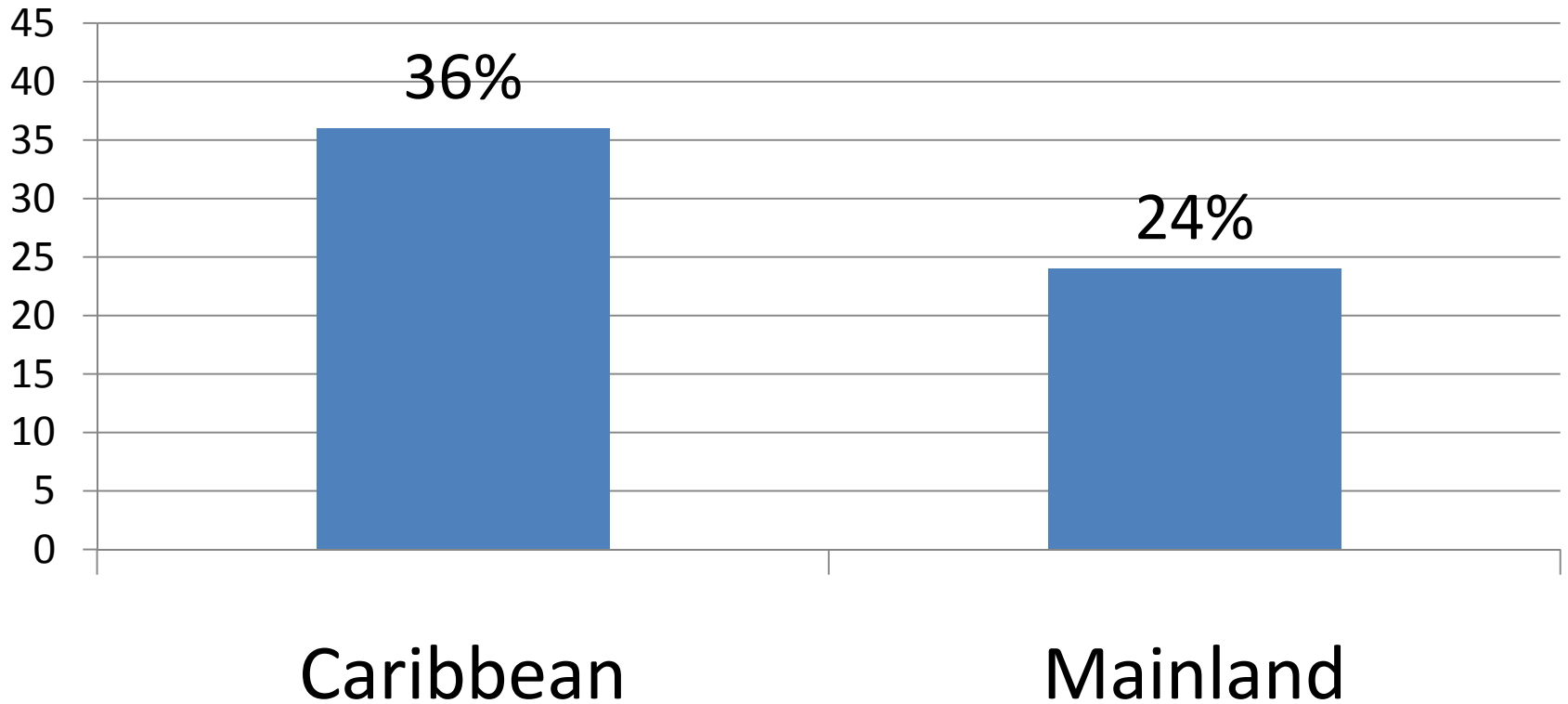
Is this a binary distinction?

- Literature has generally treated it this way.
- Fits in well with concept of parametric variation.
- But is this right?

The facts are that:

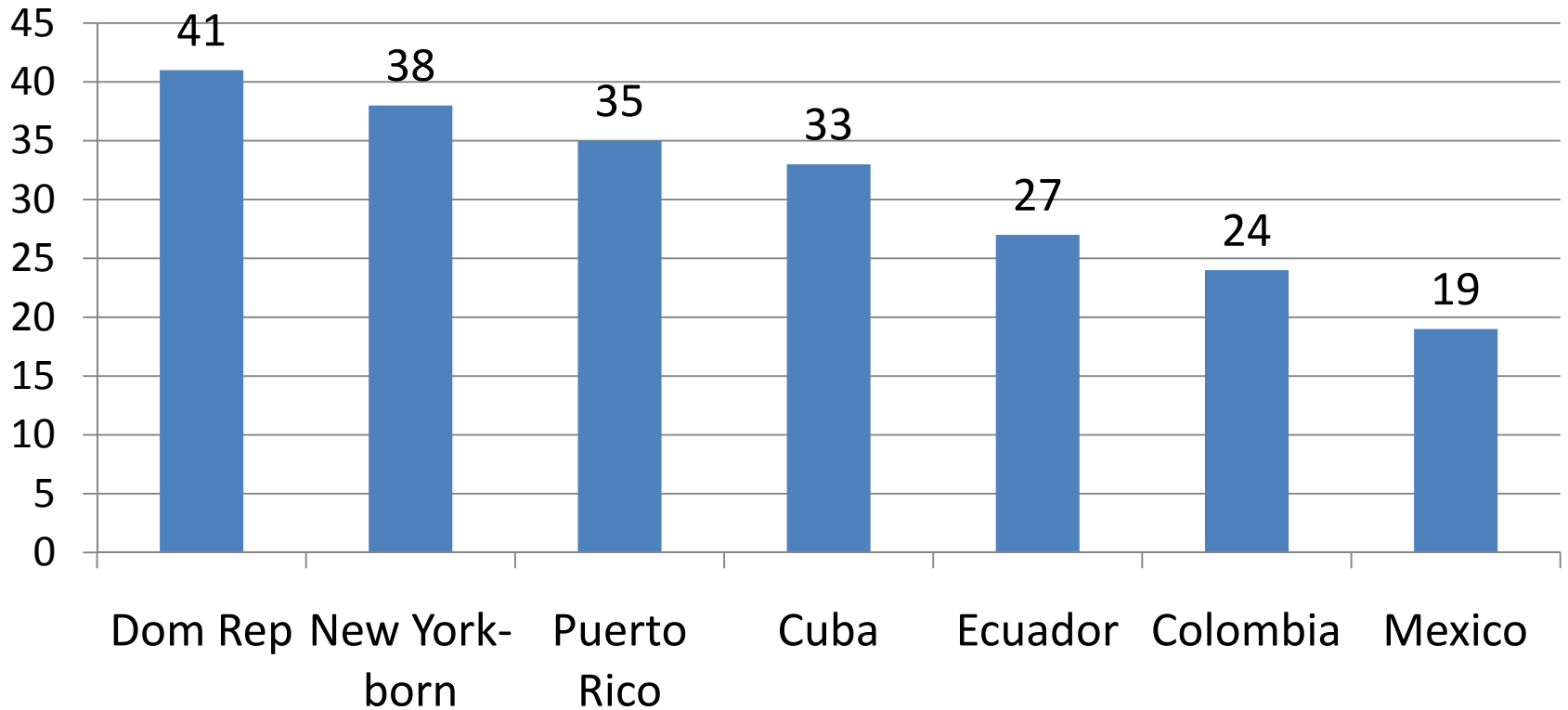
Instead of this:

Overt Pronoun Rate



We really have this:

Overt Pronoun Rate



What is going on?

- Presumably reflects differences in discourse properties of overt subject pronouns:
 - How accessible
 - (How individuated)
- We then predict small differences from region to region in terms of how disruptive an intervening overt subject is for processing a filler-gap dependency.

wh-words

- We do NOT expect much variation with regard to the *wh*-words themselves.

(26) why >

(how >)

where/when >

what/who

We predict the following varieties:

(27) *Wh*-words which tolerate an intervening overt subject:

- | | | | | |
|----|-------------|-------------|--------------------|------------------|
| a. | *why | *how | *when/where | *what/who |
| b. | why | *how | *when/where | *what/who |
| c. | why | how | *when/where | *what/who |
| d. | why | how | when/where | *what/who |
| e. | why | how | when/where | what/who |

These are the 5 attested patterns

- Patterns such as the following were NOT found: Bakovic (1998)

(28)

- e. **why** **how** ***when/where** **what/who**
- f. ***why** ***how** ***when/where** **what/who**

All other things being equal

We should find the same correlation across
Romance and beyond:

(29) The more resources required to process an
intervening subject, the less tolerated an
intervening subject will be.

Brazilian Portuguese may be a good example of this

- Subjects are generally overt and preverbal.
- Thus overt subjects are not limited to inaccessible referents,
- and preverbal subjects are not limited to highly individuated referents.
- Processing overt preverbal subjects should thus not be particularly costly.

Brazilian Portuguese may be a good example of this

- As expected, then, intervening subjects are well tolerated in *wh*-questions:

(30) Quem **a Maria** viu?
‘Who did Mary see?’

- In fact, the syntax provides no way to avoid having an intervening subject.

Finally, a note about acquisition

Example of poverty of the stimulus problem:

(1)a. *Qué **Juan** leyó en la biblioteca?

- How does child know that preverbal subject is disallowed here, even though preverbal subjects are generally fine?
- Would seem to require negative evidence.

In the present analysis

- Child needs to figure out that overt preverbal subjects have special discourse properties.
- From this it will follow that overt preverbal subjects are costly to process and that (1a) is unacceptable.
- Children **do** have access to this kind of evidence, so the unacceptability of (1a) follows (Goodall 2007). The poverty of the stimulus problem disappears.

Conclusion

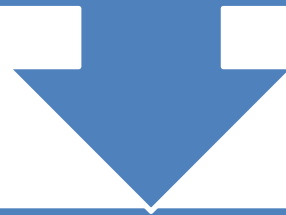
- Experimental methods allowed us to detect subtle gradience in judgments.
- Accounting for this gradience involved invoking processing factors.
- These same processing factors account for much of the inversion effect.

Conclusion

- We can then account for inversion effect with straightforward syntax:
wh-movement, placement of subject.
- We predict observed variation in a natural way, without parameters.
- Possible solution to a poverty of the stimulus puzzle.

Moral of the story

Small changes in methodology



can lead to big changes in view of
grammar.

Thank you!