

Processing accounts for gradience in acceptability: The case of multiple *wh*- questions

Snider, N., Arnon, I.,
Hofmeister, P., Jaeger, T.F.
& Sag, I.A.



The big question

- Why are some structures unacceptable?
 - ‘Competence’ (Grammar)
 - ‘Performance’ (Processing complexity)

- (1) The boy the cat the dog bit scratched started crying.
 < (less acceptable)

- (2) The reporter everyone I met trusts is interviewing the president.
 (Bever, 1970)



However,

- Other structures display gradient acceptability when processing factors are manipulated

- **Island constructions** (Kluender 1991, 1992, 1998; Kluender and Kutas 1993)
 - (3) That was the play that the professor wondered whether a student will like.

<
 - (4) That was the play that they wondered whether you would like



Advantages of a processing account

- Accounts for gradience
- Appeals to independently motivated factors
- maximizes explanation and minimizes grammar



A test case: Multiple *wh*-questions

- English: supposedly categorical order constraints

(5a) Who bought what? [lovely]

(5b) *What did who buy? [supposedly ungrammatical ⇒ SUV]

- In contrast, German & Russian described as lacking these constraints:

(6a) Wer hat was gekauft? [lovely]

(6b) Was hat wer __ gekauft? [lovely]



A test case: Multiple *wh*-questions

- Apparently categorical contrast attributed to differences in grammar
 - English has a Superiority Constraint (Chomsky, 1973), German and Russian don't



Overview

- The contrast isn't categorical
- Evidence that choice between wh-orders influenced by well-established processing factors
 - Locality constraints
 - Accessibility constraints
- Evidence from German and Russian
- Explaining cross-linguistic variation



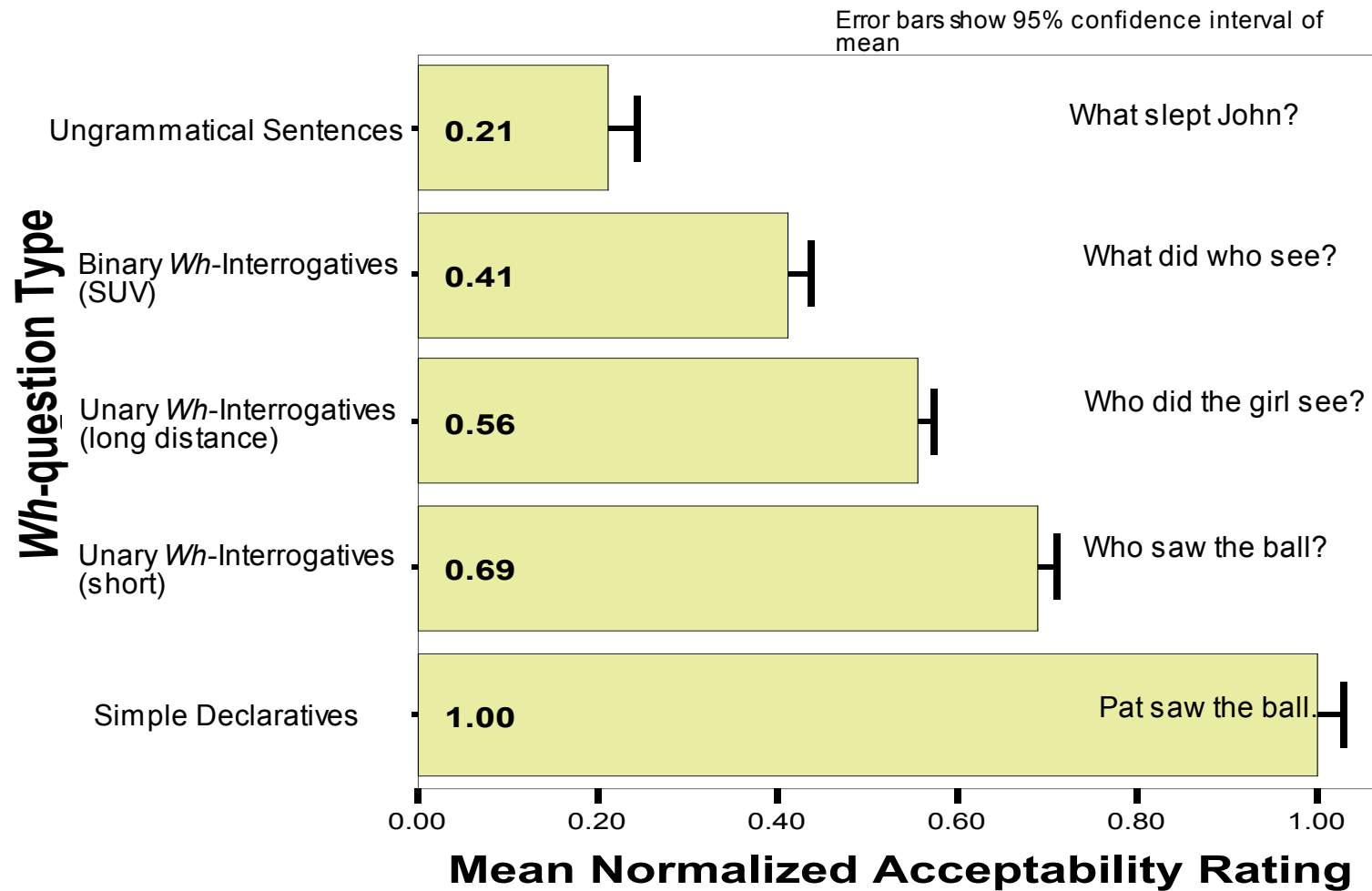
SUVs naturally occur

- Arnon et al. (2005) point out ‘Superiority-violating examples occurring naturally in corpora, e.g.

What changed recently? If the system has been working until the last month or so, and suddenly is going beserk [*sic*], something has changed to trigger the corruption. What did who do (with or without your knowledge) to these systems.

- Clifton et al. (to appear) report a similar set of examples

SUVs rated differently from ungrammatical sentences





Noted exceptions

- Which book did which student read?
- D-linking (Pesetsky 1987)

Why are SUVs unacceptable?



Wh-processing hypothesis (WPH)

- *Given the choice between several grammatical wh-orders, speakers disprefer those which (given the context) are associated with a greater processing cost*
- I. Gaps that are further from the filler are harder to process (Gibson, 1998; 2000)
- II. Less accessible fillers are harder to process
- III. Less accessible intervenors are harder to process (Warren & Gibson, 2002)



Locality Constraints

- Effect of distance (=locality) on acceptability, frequency and grammaticalization of long-distance dependencies observed in Gibson (1991;1998;2000) Hawkins (1994; 1999), and many others



Accessibility

- Words associated with high cognitive cost (=low accessibility) between filler and gap increase difficulty (Gibson, 1998; Alexopoulou & Keller, 2003)
- *Which* phrases are markers of high accessibility (cf. Ariel, 2001). This is supported by the findings of Frazier & Clifton (2002).



The evidence

- Tested WPH in English multiple wh-questions
- Assumption: increased processing difficulty reduces acceptability (Fanselow & Frisch, to appear)
- One acceptability judgment survey using M(agnitude) E(stimation) and one self-paced reading-time study



Experimental techniques

■ Magnitude Estimation

- Judgments were elicited over WWW using Magnitude Estimation (Bard et al., 1996) with WebExp software (Keller et al., 1998)
- Participants set their own continuous scale of acceptability, based on a reference sentence
- 36 participants; unpaid; recruited from e-mail lists

■ Reading times

- Measured using self-paced moving window
- A comprehension question followed each test item
- 41 participants; paid; MIT undergrads

Experiment 1: Locality

Hypothesis

- Longer filler-gap distances (measured in terms of new discourse referents, DLT:Gibson, 2000)
 - ⇒ higher processing cost
 - ⇒ lower acceptability judgments

Which man did [the girl] see ___? >

1

Which man did [the girl] in [the bar] on [California Ave.] see ___?

3



Experiment 1: Locality

- Single wh-phrases
- Examined difference between examples like (a)-(d) below in a Latin square design, varying extraction type and PP attachment


a. *Which man ___ saw the girl?* (SBJ-WH, 0 INT)

b. *Which man ___ saw the girl in the bar on California Ave?*
(SBJ-WH, 0 INT)

c. *Which man did **the girl** see ___?* (OBJ-WH, 1 INT)

d. *Which man did **the girl in the bar on California Ave.** see ___?*
(OBJ-WH, 3 INT)

Experiment 1: Locality Results

- a. *Which man ___ saw the girl?* (SBJ-WH, 0 INT) n.s.
 - b. *Which man ___ saw the girl in the bar on California Ave?* (SBJ-WH, 0 INT)
 - c. *Which man did **the girl** see ___?* (OBJ-WH, 1 INT) *
 - d. *Which man did **the girl in the bar on California Ave.** see ___?* (OBJ-WH, 3 INT)
- 

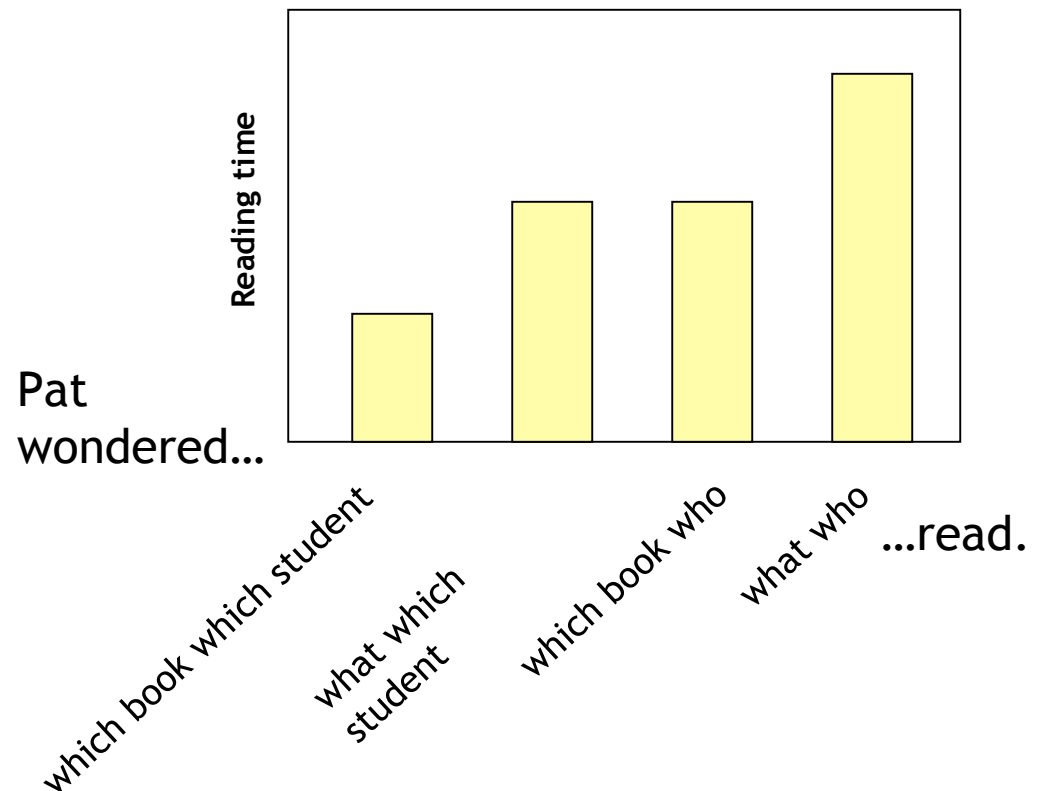
Experiment 2+3: Accessibility

- ME Acceptability experiment
- Reading time experiment
- Same stimuli

Hypothesis

SUVs with high accessibility fillers and interveners should be **more acceptable** and **read faster** than those with low accessibility fillers and interveners

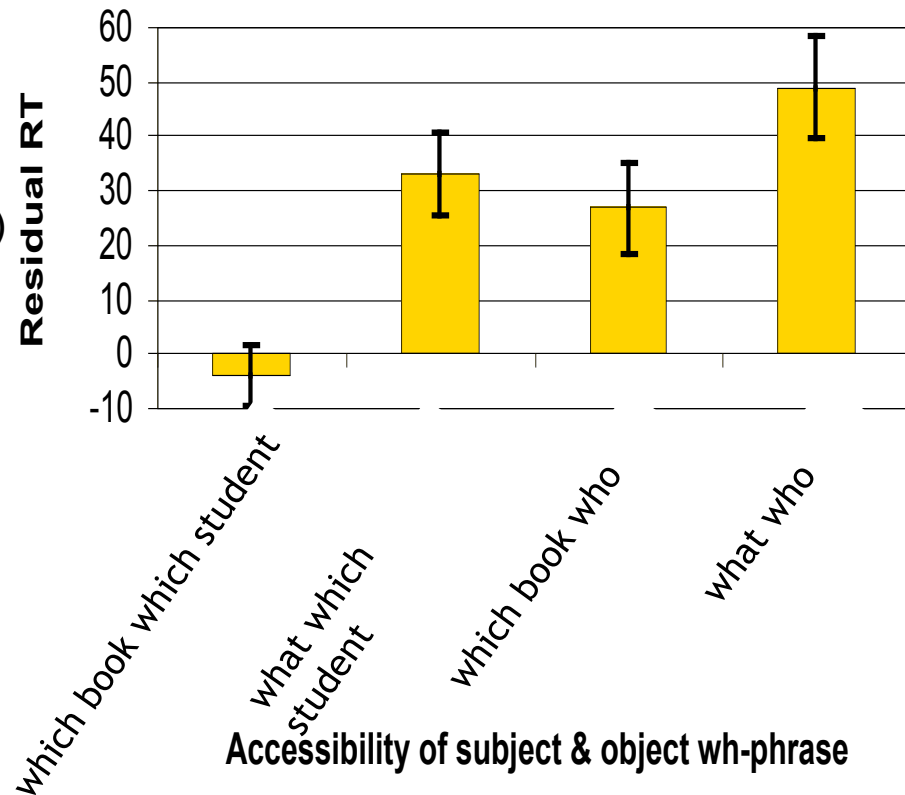
Prediction for Filler x Intervener



Results: Accessibility and reading times

- more accessible fillers result in faster processing at the verb
($F1(1,40) = 17.7$, $p < .001$, $F2(1,19) = 12.3$, $p < .003$)
- Same for more accessible interveners
($F1(1,40) = 10.5$, $F2(1,19) = 11.5$, $P_s < .01$)
- ME study showed similar effects

Accessibility effect on reading time





Summary of English data

- We have shown that a historically grammatical contrast involves a gradient space of judgments that (partially) depends upon locality and accessibility
- Considerable amount of variation can be explained by processing factors introduced to explain other phenomena in sentence processing
- Our results so far may only account for part of the observed variation
 - frequency effects of verbs (item variation), plausibility, clause boundaries, and contextual factors are likely also relevant
- SUVs cannot be treated simply as ungrammatical on account of gradience, corpus data, and experimental evidence indicating the role of processing complexity



So far...

- The processing account can better explain the gradient of the data and its sensitivity to processing factors



BUT: Cross-linguistic challenge

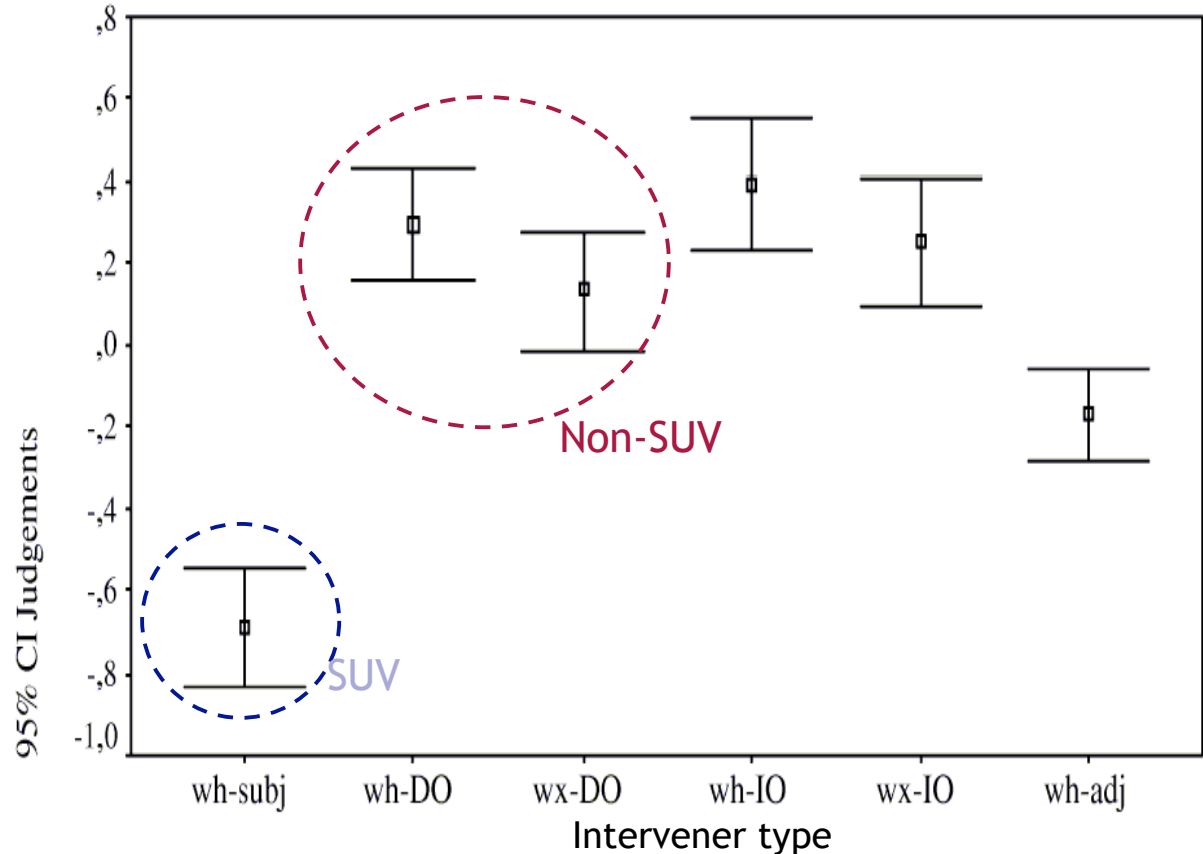
- If it's processing preferences, why the difference between languages?
- Re-evaluate the claims with more precise measures
- Review cross-linguistic data: Russian and German

Superiority in German

Superiority has been argued not to exist in German (Grewendorf 1988, Müller 1991)

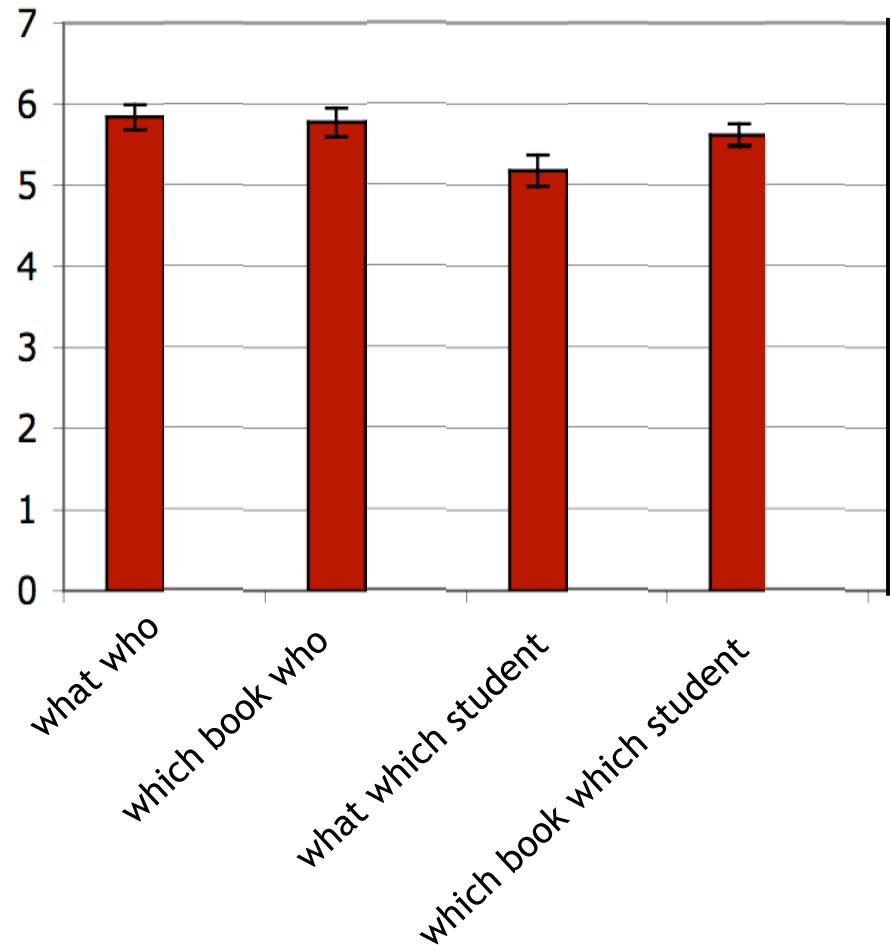
However, Featherston (2005) showed


- 1) SUVs are worse than non-SUVs
- 2) More accessible interveners improve acceptability (F1=5.71, p=0.025; F2=51.65, p<0.001)



Superiority in Russian

- Argued not to exist in Russian
- Fedorenko and Gibson (2006) found no effect of wh-order (Superiority)
- Also no effect of filler or intervener accessibility





Summary of X-ling Facts

- English multiple *wh*-phrases display gradient acceptability, depending on locality and accessibility of filler and intervening *wh*-phrases
- The same is true for German, although the constructions are still more acceptable
- No measurable effect in Russian



How to explain cross-linguistic variation within a processing account?

- The Competition Model (MacWhinney & Bates, 1989): cross-linguistic differences in processing influenced by the availability and reliability of cues in the language.
- Different cost associated with dispreferred options, depending on other cues.



What motivates the preferences?

- Influence the ease of resolving a dependency

- Language-specific features that might influence resolution ease
 - Case-marking: lets hearer partially integrate
 - thus renders the violation of the other preferences less costly



The case for case...

- English - no case marking
- German and Russian - case marking
 - Evidence that case marking is more ambiguous in German than in Russian and used less reliably in online sentence processing (Kempe & MacWhinney, 1999)
- The effect of the preferences: English > German > Russian attributed to the different strength of case marking as a cue



Predictions for Russian

- Effects of processing preferences should be measurable when case cues are not available
- Predict ‘superiority effect’ in multiple *wh*-constructions with two inanimate NPs



How to accommodate cross-linguistic variation

- Cross-linguistic variation as an interaction of general preferences and language specific features



Summary

- Presented a processing account for the unacceptability of SUVs based on independently motivated processing factors
- Better accounts for the gradience of the data
- Doesn't require construction specific grammatical constraints
- Need to re-examine what we consider ungrammatical and why



Thanks to

Ted Gibson

Ev Fedorenko

Bruno Estigarribia

Jeanette Pettibone

David Pesetsky

Colin Phillips

Sam Featherston

Experiment 1: Locality

- Object extractions (more interveners) are judged as somewhat less acceptable than subject extractions ($F(1,35) = 4.9, p < .05$; non-significant by items, $F(1,35) = 2.5, p = .12$).

Figure 2: Overall Effect of EXTRACTION

