Making a good mismatch – (How) syntactic cues license voice mismatches in VP ellipsis

Robin Lemke, Lisa Schäfer, Ingo Reich

Voice mismatches under Verb Phrase Ellipsis

Voice mismatches can be acceptable under VP ellipsis (1a), but this is not the case for all connectors (1b) (Kehler 2000)

(1) a. *This problem was <u>to have been looked into</u>, **but** obviously nobody did ⟨look into this problem⟩.*
b. **This problem was <u>looked into</u> by John, **and** (similarly) Bob did ⟨look into this problem⟩, too.*

Discourse structure account (Kehler 2000)

- Different discourse relations between conjuncts require categorially different resolution mechanisms (syntactic/semantic)
- Only *resemblance* relations (1b) require morphosyntactic identity ((1a) is an instance of a *cause-effect* relation)

Information-theoretic account

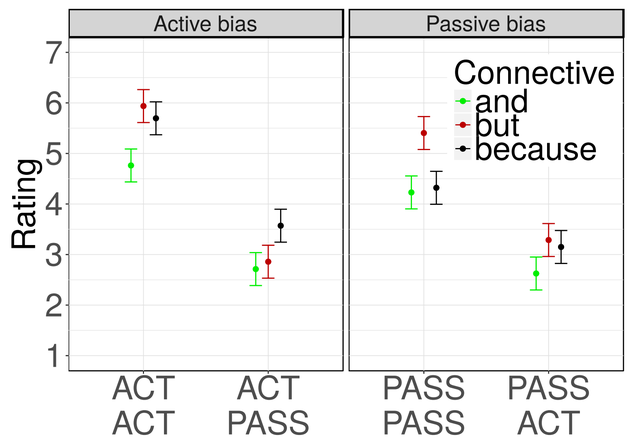
- Information (**$-\log_2 p$ (word | context)**) indexes processing effort (Hale 2001)
- Uniform Information Density (Levy & Jaeger 2007): Avoid peaks and troughs in the ID profile
- Voice mismatches are more acceptable the more likely the target is in context (including connector)

Experiment 1 – *and* vs. *but* vs. *because*

(2) a. *Joshua **didn't** give Sarah private lessons in Mandarin **(and | but | because)** Jacob **(did | was).*** (ACTIVE BIAS)
b. *Sarah **wasn't given** private lessons in Mandarin by Joshua **(and | but | because)** Jacob **(did | was).*** (PASSIVE BIAS)

Method

- **BIAS × CONNECTOR × MISMATCH** (2 × 2 × 3), BIAS tested between subjects
- 96 subjects, 30 items, 78 fillers
- Web-based (`prolific.ac`)
- Analysis with CLMMs in `R` (`ordinal`, Christensen 2015)



Results

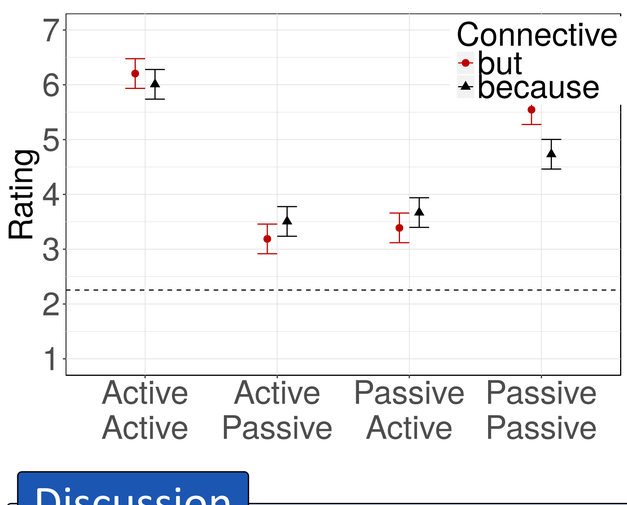
- CONNECTOR:MISMATCH interactions: mismatches with *because* improve compared to *but* ($z_{act} = 4.89$, $z_{pass} = 5.74$, both $p < 0.001$)
- The same holds for *and* vs. *but* ($z_{act} = 6.39$, $z_{pass} = 3.11$, both $p < 0.01$)

Experiment 2 – *but* vs. *because*

Method

- **2 × 2 × 2 design: BIAS × CONNECTOR × MISMATCH**
- 64 subjects, 32 items, 78 fillers

(3) *Joshua **didn't** give Sarah private lessons in Mandarin **(but | because)** Jacob **(did | was).***



Results

Mismatches with *because* are more acceptable than mismatches with *but* ($z = 7.38$, $p < 0.001$)

Discussion

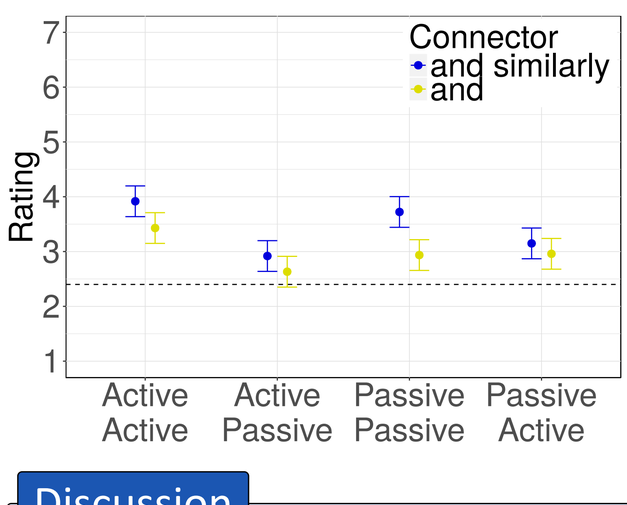
- Effect expected under Kehler's account if *but*, but not *because*, encodes a resemblance relation
- Effect expected under the UID account if a parallel continuation is more likely given *but* than *because*

Experiment 3 – *and* vs. *and similarly*

Method

- **2 × 2 × 2 design: BIAS × CONNECTOR × MISMATCH**
- 64 subjects, 32 items, 78 fillers

(4) *Joshua **gave** Sarah private lessons in Mandarin **(and | and similarly)** Jacob **(did | was).***



Results

Mismatches with *and* are less acceptable when *similarly* is inserted ($z = -4.55$, $p < 0.001$)

Discussion

- Effect of *similarly* unexpected under Kehler's account: Resemblance relation in both CONNECTOR conditions
- Effect expected under the UID account if *similarly* increases the probability of a parallel continuation

Selected references • Christensen, R. H. B. (2015). *ordinal – Regression models for ordinal data. R package version 2015.6-28.* `http://cran.r-project.org/package=ordinal/` • Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. *Proceedings of NAACL (Vol. 2)*, 159–166. • Kehler, A. (2000). Coherence and the resolution of ellipsis. *Linguistics and Philosophy*. 23, 533–575. • Levy, R. & Jaeger, T.F.(2007). Speakers optimize information density through syntactic reduction. In: Schlökopf, B.; Platt, J. & Hoffmann, T. (Eds.) *Advances in neural information processing systems* 19, 849–856.