

More compelling, the closer you look:
**A contrastive study of ‘reversed’
Comparative Correlatives**

Jakob Horsch

L. Štúr Institute of Linguistics, Slovak Academy of Sciences

jakob.horsch@korpus.sk

**49. Österreichische Linguistiktagung
Emerging Linguists Workshop
December 7, 2025**



Funded by the EU's *NextGenerationEU* program
through the Recovery and Resilience Plan for Slovakia
Project No. 09I03-03-V04-00713 (2024-26)
“Scholarship for excellent researchers (R2)”



PLÁN [OBNOVY]



Funded by

DFG Deutsche
Forschungsgemeinschaft

German Research Foundation

DFG-HO 3904/5-1 & -2

1. The Comparative Correlative (CC)
2. Previous studies
3. Methodology
4. Corpus study results
5. Conclusion

1 The Comparative Correlative

- Bi-clausal construction with complex form and complex meaning:

(1) a. [The more I read,]_{C1} [the more I know]_{C2}

b. [Cuanto más leo]_{C1} [(tanto) más entiendo]_{C2}

CUANTO more read:I (TANTO) more understand:I

c. [Čím víc čítam]_{C1} [tým víc rozumiem]_{C2}

ČÍM more read:I TÝM more understand:I

1 The Comparative Correlative

Complex form:

- Lexically fixed clause-initial elements

(the_{C1} quanto_{C1} /čím_{C1}; the_{C2} / (tanto_{C2}) /tým_{C2})

- Obligatory comparative elements (AdvP, AdjP, NP, PP)
- Optional clause slots (cf. *The more, the merrier.*)

1 The Comparative Correlative

Complex form:

- Lexically/phonologically fixed elements + schematic, open slots
- **Maximally abstract templates** (CULICOVER and JACKENDOFF 1999)

- (2) a. [*the* [...]_{comp. element 1} (clause₁)]_{C1} [*the* [...]_{comp. element 2} (clause₂)]_{C2}
b. [*cuanto* [...]_{comp. element 1} (clause₁)]_{C1} [(*tanto*) [...]_{comp. element 2} (clause₂)]_{C2}
c. [*čím* [...]_{comp. element 1} (clause₁)]_{C1} [*tým* [...]_{comp. element 2} (clause₂)]_{C2}

1 The Comparative Correlative

Complex meaning (HOFFMANN 2017: 351)

- C1 cause/protasis, C2 effect/apodosis (GOLDBERG 2003: 22)
- **asymmetric** cause-effect meaning
($C1 \rightarrow C2$; 'reading more results in understanding more')
- **symmetric** parallel-change-over-time meaning
(‘as you read more during a time period x ,
so you will understand more during a time period y ’)

1 The Comparative Correlative

- Templates like (2) do not account for all formal variations, cf. e.g. (3a-c):

(3) a. [I know (the) more,]_{C2} [the more I read]_{C1}

b. [Entiendo (tanto) más,]_{C2} [cuanto más leo]_{C1}

understand:I (TANTO) more CUANTO more read:I

c. [Rozumiem tým viac]_{C2} [čím viac čítam]_{C1}

understand:I TÝM more ČÍM more read:I

1 The Comparative Correlative

- Templates like (2) do not account for all formal variations, cf. e.g. (3):

(3) a. [I know (the) more,]_{C2} [the more I read]_{C1}

- apodosis precedes protasis
- “reversed” order (McCawley 1988: 176)
- “CC’ ” (Culicover and Jackendoff 1999: 549)
- vs. “canonical” C1C2 order (Hoffmann 2018: 192)

2 Previous studies

- Significant C1C2:C2C1 preference across languages:
 - HOFFMANN (2019): **37:1** (BROWN) and **23:1** (ICE) (EN)
 - HOFFMANN et al. (2020): **15:1** (BNC) (EN)
 - HORSCH (2020): 40:1 (Slovak Web 2011) (SK)
 - no data for Spanish

2 Previous studies

Why? - Principle of Iconicity (BYBEE 2012: 529)

- “(partial) motivation of a construction’s form by its meaning”
(HOFFMANN 2019: 12)
- i.e., C1C2 is preferred because it “mirrors the semantic [cause-effect] interpretation” of the CC (HOFFMANN 2018: 186)
- DIESSEL (2008): sequential iconicity = order of events in discourse reflects [perceived] order of events in real world (cf. HAIMAN 1985)
- “strong predictor of clause order” (LÁZNIČKA 2016: 97)

2 Previous studies

Limitations of corpus studies:

- very different corpora (BROWN, BNC, Slovak Web 2011...)
- CC a mildly frequent construction (30-40:1 pmw)
- small data sets, e.g. BROWN: 4 C2C1 tokens (HOFFMANN 2019: 173)
- **positive data problem:** “just because a construction appears in a corpus it does not automatically follow that it is grammatical”
(HOFFMANN 2011: 1)

2 Research questions

1. What **function** does the C2C1 order serve?
cf. Corollary A to Principle of No Synonymy (GOLDBERG 1995: 67)
2. Can the C1C2 preference be confirmed...
 - a. using different types of empirical data?
 - b. across languages?

3 Methodology

Acceptability judgments:

- English, Spanish, Slovak
- Pen-and-paper questionnaires, on site (Trinity College ($n=39$), University of Santiago de Compostela ($n=39$), Comenius University ($n=40$))
- Variable ORDER (levels: C1C2, C2C1)

3 Methodology

Magnitude Estimation (ME) (BARD et al. 1996; COWART 1997; HOFFMANN 2013)

- subjects generate their own scales
- results are made comparable ('centered') using z-scores
- grammatical/ungrammatical fillers: 'yardsticks'
- assumption: low-frequency phenomena (C2C1) rated worse than high-frequency phenomena (C1C2) but better than ungrammatical fillers

3 Methodology

Questionnaire design

- Lexicalizations to test each condition multiple times (C1C2/C2C1)
- Latin squares method: 3 material sets (MS) with 9 test items, 9 grammatical fillers and 9 ungrammatical fillers → 27 items per questionnaire
- randomization and blocking
- Instructions in English (at least B2)

3 Methodology

Data analysis: Mixed-effects modelling

- Stepwise regression/backward selection (KUZNETSOVA et al. 2015, 2020)
- Tested effects:
 - fixed effects: ORDER, THE_IN_C2 (English), TANTO_IN_C2 (Spanish), POSITION_COMPARATIVE_ELEMENT (Slovak)
 - random effects: SUBJECT, GENDER, LEXICALIZATION, MATERIAL SET
 - + Interactions

4 Results

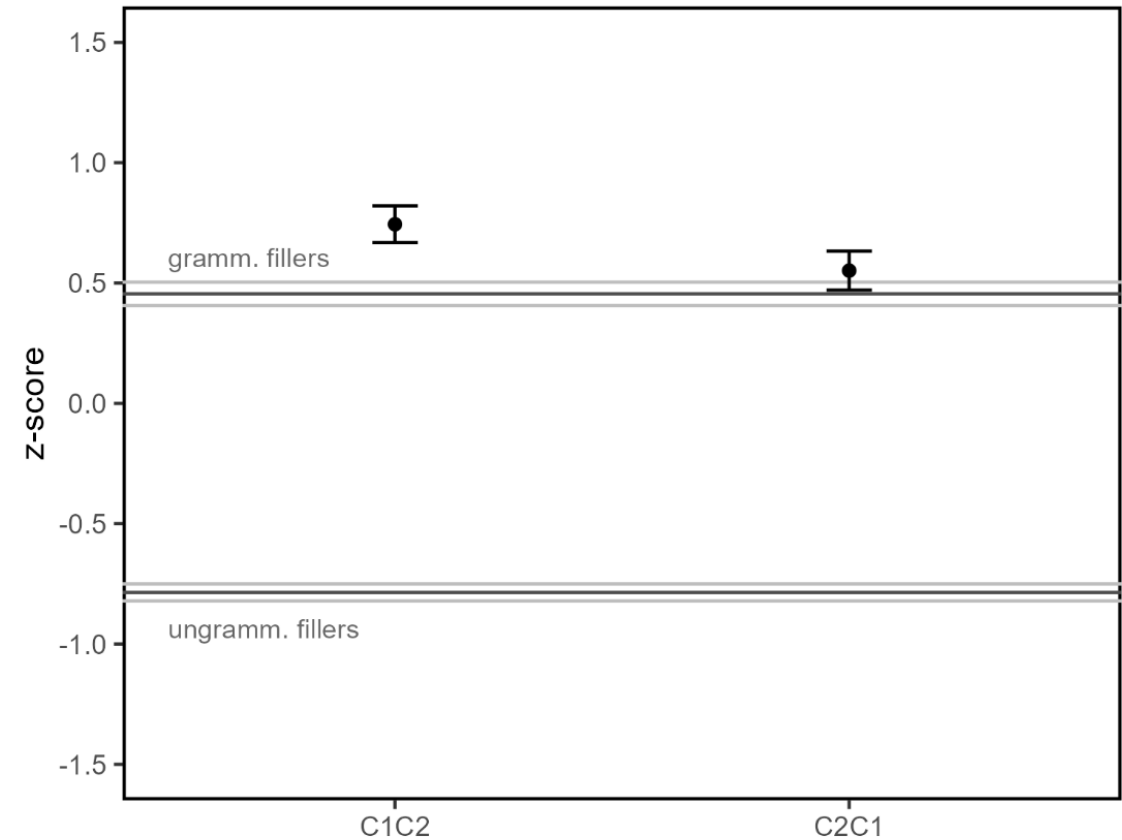
English

Random effects:

- PARTICIPANT: n.s. (singular fit)
- MATERIAL SET: 0.83 n.s.
- LEXICALIZATION: 0.0003376 ***

Fixed effect:

- ORDER: 0.07 N.S.



ORDER	z-score	SD	SE
C1C2	0.74	0.78	0.08
C2C1	0.55	0.83	0.08

Table 1. Z-score means of ORDER, English ($n=39$)

4 Results

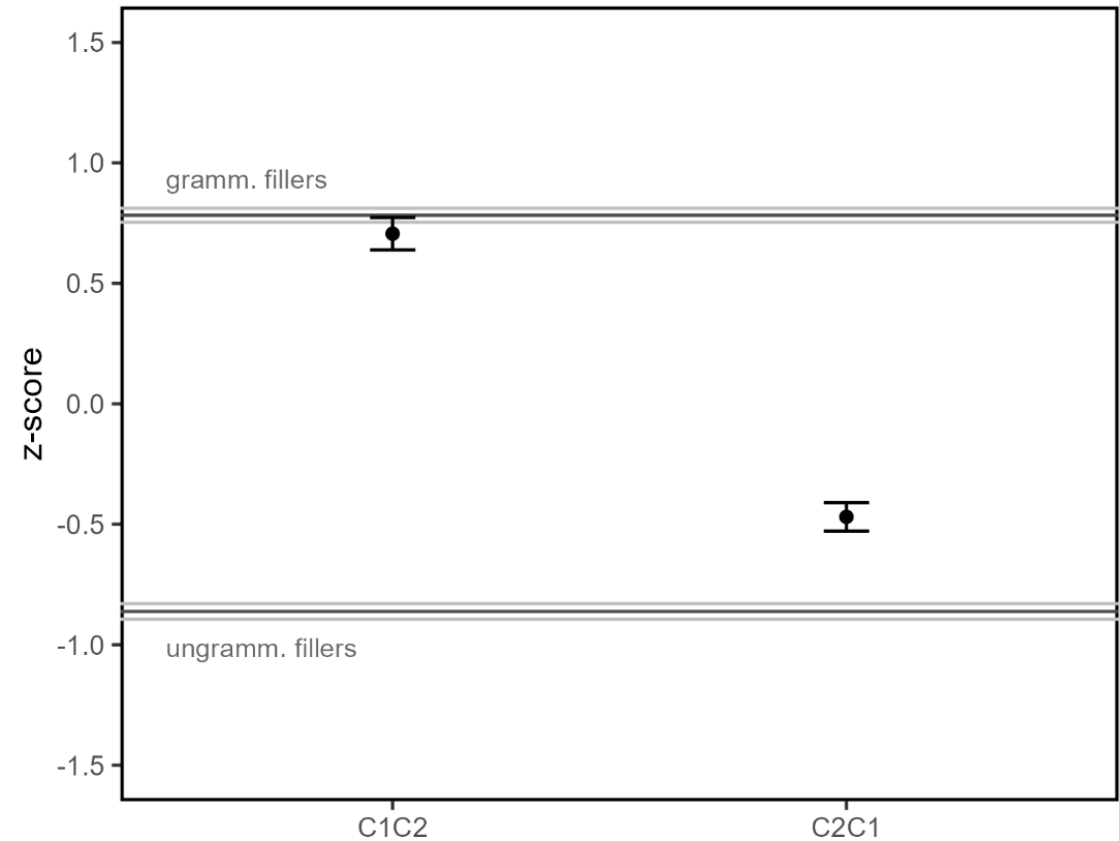
Spanish

Random effects:

- PARTICIPANT: n.s. (singular fit)
- MATERIAL SET: n.s. (singular fit)
- LEXICALIZATION: 0.001627 ***

Fixed effect:

- ORDER: $< 2e-16$ ***



ORDER	z-score	SD	SE
C1C2	0.71	0.73	0.07
C2C1	-0.47	0.64	0.06

Table 1. Z-score means of ORDER, Spanish ($n=39$)

4 Results

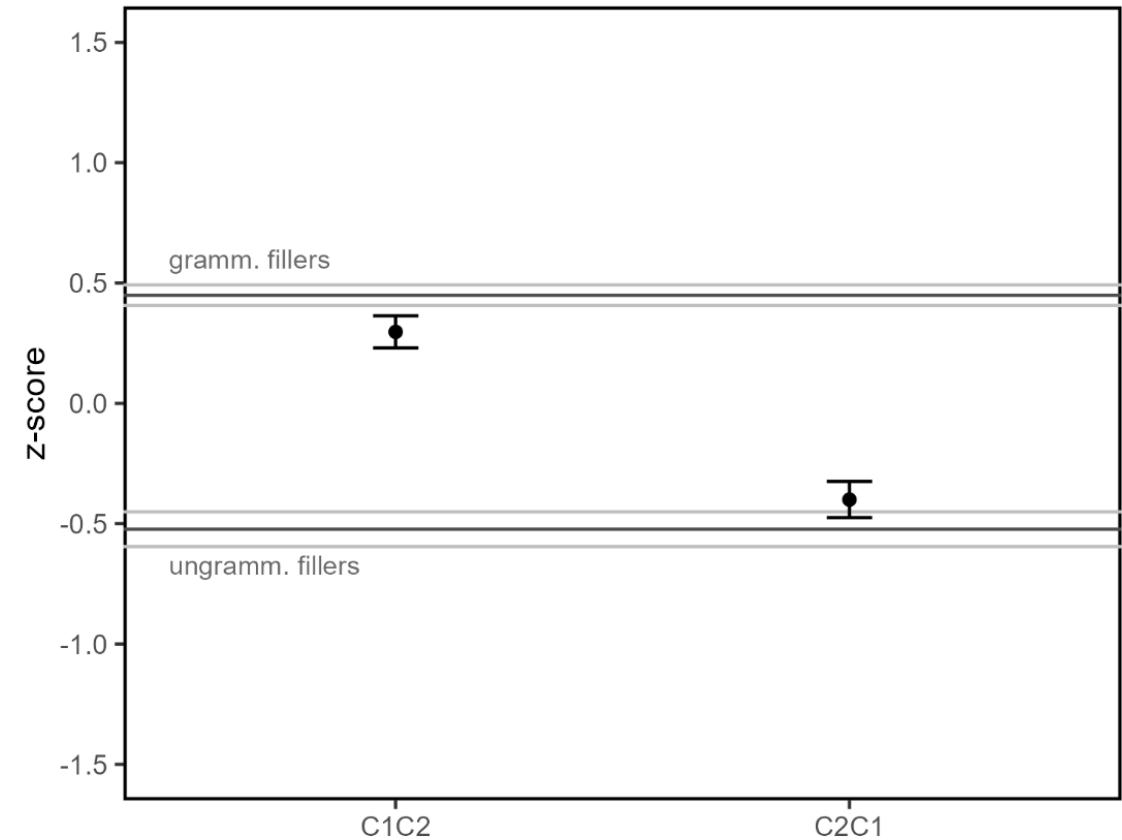
Slovak

Random effects:

- PARTICIPANT: 0.39 N.S.
- MATERIAL SET: 0.00008656 ***
- LEXICALIZATION: 0.0003003 ***

Fixed effect:

- ORDER: $1.2e-13$ ***



ORDER	z-score	SD	SE
C1C2	0.30	0.72	0.07
C2C1	-0.40	0.81	0.08

Table 1. Z-score means of ORDER, Slovak ($n=40$)

5 Conclusion

- Across languages, C1C2 clearly preferred over C2C1
- However ORDER not a significant effect in English
- In line with Principle of Iconicity: cause-effect semantics
- Function: Emphasis?
- Principle of Emphasis: unusual constituent orders that “strike the recipient as more or less out of the ordinary [...] fulfil an additional communicative purpose” (FIRBAS 1964: 117)

5 Conclusion

- Contrastive focus on C2 = effect semantics foregrounded (HORSCH 2020: 35)
- General foregrounding of asymmetric cause-effect semantics
- “the use of a focus particle [e.g. *even*] in C2 is more acceptable in CC’ constructions” in English, cf. (4) and (5) (HOFFMANN 2019: 157)

(4) [The more you think about it,]_{C1} [the more interesting it **even** becomes]_{C2}

(5) [It becomes **even** more interesting,]_{C2} [the more you think about it]C1

References, abstract, slides, spreadsheets...

