



Dissociation between Production and Comprehension Complexity

T. Florian Jaeger (Ling. Dept., Stanford & MIT), Ev Fedorenko, & Edward Gibson (Dept. of BCS, MIT)

Contact: tiflo@stanford.edu (Florian)



Is Comprehension Exclusively Driven by the Frequency Distributions in the Input?

Frequency-based theories of sentence processing argue that comprehension complexity may be reduced to distributional patterns in the comprehender's language input (e.g., MacDonald, 1999). We present evidence challenging the strongest version of this view. We argue that **while comprehenders are sensitive to distributional patterns in the input, comprehension complexity is not entirely reducible to these patterns.**

The phenomenon: Relativizer Omission

English restrictive object-extracted relative clauses (ORCs), indicated by [...] in (1), exhibit optional relativizer omission (i.e. they can be realized with or without a relativizer, e.g. *that*):

(1) The journalist [(that) the editor congratulated _] won a prize.

Both distributional patterns of relativizer omission in corpora and the effect of relativizer omission on comprehension have been studied extensively (we briefly summarize these findings below).

This makes ORCs the ideal test-case for theories of sentence processing that attempt to reduce comprehension complexity to frequency distributions in the input of the comprehender.

Relativizer Frequency in the Input

Many corpus studies have shown that relativizers likelihood increases with an increase in the complexity of the ORC.

Among many other factors, relativizer likelihood correlates with the complexity of the ORC's subject:

- The less accessible the subject is, the more likely is a relativizer (e.g. Biber et al., 1999; Jaeger & Wasow, 2005; Race & MacDonald, 2003), cf. *the editor* in (1) vs. *you* in (3):

(2) The journalist [(that) you congratulated _] won the prize.

The Effect of Relativizers on Comprehension

Race & MacDonald (2003, Experiment 3) show that:

- If the ORC's subject is a pronoun, as *you* in (2) above, a relativizer helps comprehension (as evidenced by decreased reading times on the ORC's verb).

- But if the ORC's subject is a common noun phrase, as *the editor* in (1) - (2) above, a relativizer does *not* help comprehension.

This mirrors the distribution of relativizers in corpora.

Race & MacDonald (2003) conclude that relativizers always help comprehension only in environments where they are used in production (and therefore, where they occur in the comprehender's input).

Summary & Definition of Goal

The results of Race & MacDonald (2003) provide support for the claim that comprehenders are sensitive to distributional patterns in their language input.

However, it does not follow that all or even most of comprehension complexity is driven by distribution in the input.

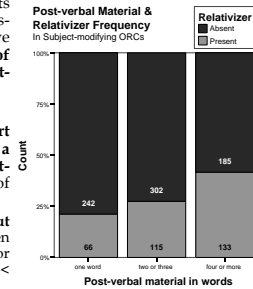
The question thus is: *Can comprehension complexity be reduced to comprehenders' sensitivity to distributional patterns in their language input?*

An Unexpected Distributional Pattern

Race & MacDonald (2003; also Jaeger, et al., 2005) also find that relativizer likelihood correlates with the complexity of the ORC's post-verbal region: **The more post-verbal material, the more likely is a relativizer**, cf. (3) vs. (1):

(3) The journalist [(that) the editor congratulated _ *yesterday during a meeting*] won a prize.

To evaluate the effects of the presence/absence of a relativizer with regard to the complexity of the post-verbal material on comprehension difficulty, and to test whether these effects reflect the relativizer distribution in corpora, we restricted the domain of investigation to subject-modifying ORCs, cf. (3).



We found additional support for the post-verbal effect in a study of 979 subject-modifying ORCs (out of more than 5,300 ORCs):

- The effect was subtle but significant for both written ($\chi^2=10.7$, $p < 0.01$) and for spoken language ($\chi^2=13.3$, $p < 0.01$)

So, do relativizers ease the comprehension of long post-verbal domains or integration of ORCs with long post-verbal domains?

Comprehension: Self-paced Reading Experiment

Design, Participants, & Procedure

Design - 2 x 3

Factors:

- Relativizer presence (REL) (object-extracted relative clauses (ORCs) with or without relativizer)
- Complexity of post-verbal domain in RC (POSTV) (one, two, or three 3-word prepositional phrases (PPs) attached to ORC's verb)

Procedure: Self-paced word-by-word reading with a moving window display.

Participants: 39 subjects

Materials & Regions

36 sets of experimental items with object-extracted subject-modifying relative clauses (ORCs), each with 6 versions as shown below, where shading indicates the critical regions for the analysis (for more details, see Analysis & Results):

- a. Relativizer present, 1 post-verbal PP in the ORC
The player [that the coach met at 8 o'clock] bought the house near the gym by the river, but ...
- b. Relativizer present, 2 post-verbal PPs in the ORC
The player [that the coach met by the river at 8 o'clock] bought the house near the gym, but ...
- c. Relativizer present, 3 post-verbal PPs in the ORC
The player [that the coach met near the gym by the river at 8 o'clock] bought the house, but ...

- e. Relativizer absent, 1 post-verbal PP in the ORC
The player [_ the coach met at 8 o'clock] bought the house near the gym by the river, but ...
- f. Relativizer absent, 2 post-verbal PPs in the ORC
The player [_ the coach met by the river at 8 o'clock] bought the house near the gym, but ...
- g. Relativizer absent, 3 post-verbal PPs in the ORC
The player [_ the coach met near the gym by the river at 8 o'clock] bought the house, but ...

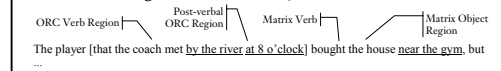
In addition to the 36 items, the experiment contained 69 fillers, totaling 105 items per list. All items were balanced over lists using Latin-square design.

Predictions

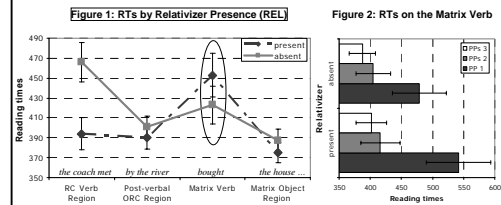
Comprehension theories that reduce comprehension complexity to distributional patterns in the input predict that **relativizers should ease comprehension of post-verbal material in the ORC or the next point of integration following it (i.e. the matrix verb).**

Analysis & Results

While we found the expected effect of REL on the ORC's verb (cf. Race & MacDonald, 2003), we did not find the predicted effect of REL or REL x POST-V on the critical regions (post-verbal ORC Region & Matrix Verb):



- Figure 1 summarizes the effect of REL over all four regions of interest.
- Figure 2 summarizes the effect of POST-V on the matrix verb region.



Analysis details for critical regions (only relevant effects are mentioned):

- Post-verbal Region:

Neither a main effect of REL nor an interaction ($F_s < 1.7$, $P_s > 0.2$).

- Main Verb:

Significant effect of POST-V ($F(1, 76)=9.0$; $F(2, 70)=9.5$; $P_s < 0.001$). RTs on the main verb decreased with more post-verbal material in the ORC (1 PP: 410ms vs. 2 PPs: 410ms vs. 3 PPs: 394ms).

The effect of REL (present: 453ms vs. absent: 423ms) was marginal by item, but did not reach significance by subjects ($F(1, 38)=2.2$, $p=0.15$; $F(2, 35)=2.8$, $p=0.1$).

Crucially, the REL effect is in the opposite of the predicted direction and there was no interaction ($F_s < 1$).

All other regions did not show any effects bearing on the current issue. But we did get all effects expected from the literature, such as a significant effect of REL on the RC region, where a relativizer eases comprehension.

Discussion & Conclusions

Frequency-based theories of comprehension in which comprehension effects are reduced to distributional patterns in the input make the wrong predictions:

- The two critical regions (the ORC's post-verbal domain and the matrix verb) did neither exhibit an effect of relativizer presence (REL) nor an interaction between relativizer presence and the complexity of the ORC's post-verbal material (POST-V).

- Numerically, the REL effect on the matrix verb points in the opposite of the predicted direction: Relativizers did not improve comprehension on the matrix verb.

Dissociations between production and comprehension like the one presented here confirm the need to study both comprehension and production.