Expectations for Novelty: The Role of Information Structure in Syntactic Processing

Hans Wilke, Jet Hoek, Bettelou Los, Antonella Sorace, Hannah Rohde (University of Edinburgh)

A longstanding question in linguistics centres on how comprehenders infer syntactic structure and what factors influence this process. Relative Clauses (RCs) have provided a useful testing ground for such questions, largely because their structural complexity challenges comprehenders' cognitive capacities and ability to hold syntactic components in memory to build long-distance dependencies. The observed complexity of different types of RCs has been attributed to their syntactic structure (Gibson et al., 2013). However, it has been shown that syntactic explanations alone fail to make the right predictions of comprehenders' processing difficulties (Gibson et al., 2005; Gordon et al., 2001). Here we investigate the influence of information structure (IS) on comprehenders' RC processing. We have designed a self-paced reading experiment in which the IS of RC structures is manipulated to either provide given information in the main clause and new information in the dependent clause, or vice versa. This is done for two types of frequently contrasted RC-structures: subject-modifying and object-modifying RCs. The goal is to test whether comprehenders have expectations about the appropriate information status of RCs when ordered early or late in the sentence. The materials allow us to contrast predictions of syntax-only theories of RC processing with those that incorporate information structural constraints. We are currently in the process of data collection/analysis.

The complexity of relative clauses

Prior work on RC complexity focuses primarily on syntactic explanations for why some RCs are easier to process than others. Two types of RCs that are often compared are subject-extracted (*The reporter who attacked the senator*) and object extracted (*The reporter who the senator attacked*). Object-extraction is often assumed to be more complex than subject extraction, given assumptions such as:

(i) Perspective shifts are harder to process
(ii) Canonical word order is easier to process
(iii) Subject positions are more accessible
(iv) Longer distance-dependencies require more storage (Gibson et al., 2013)

When subject-extracted and object-extracted RCs are compared, assumptions (i-iv) would predict that (1) should be easier than (2): It has fewer perspective shifts, canonical word order, it relativises the subject, and it has shorter dependency lengths. However, in processing studies, (1) has been found to be read more slowly than (2) (Gibson et al., 2005).

(1) The reporter ignored the president, *who attacked the senator*.
(2) The reporter, *who the senator attacked,* ignored the president.

Alongside the evidence establishing processing difficulty for sentence-final RCs like (1), it has also been shown that sentence-final RCs need not always cause difficulty. Gibson et al. (2005) postulate that foregrounded information is more easily processed late in a sentence and that certain RCs, namely non-restrictive RCs, can be understood as assigning their content a foregrounded status (see also Santi et al., 2019). Gibson et al. (2005) compared regular restrictive RCs, marked with *that* in American English as in (3), with non-restrictive RCs, marked with *who*, separated from the noun with a comma as in (4):

(3) An actor insulted the director *that the critics praised at a banquet*
(4) An actor insulted the director, *who the critics praise at a banquet*

Results show that sentences with foregrounded RCs at the end of the sentence were parsed faster than sentences with backgrounded RCs at the end of the sentence. However, this manipulation of IS also introduces a syntactic difference between conditions. Furthermore, the subject-modifying counterparts of (3) and (4) were processed more quickly overall, regardless of their IS. It is therefore still not clear whether IS influences RC processing when syntactic structure is held constant.

Two of the previously mentioned assumptions are relevant to the current study; (i) and (iv), and we investigate two IS-related hypotheses: the Information Flow Hypothesis (Gibson et al., 2005), which postulates that new information is more easily processed later in a sentence, and an information status-clause type mapping based on corpus-based evidence that shows that new information tends to be expressed in main clauses, while old information tends to be presented in dependent clauses (e.g., Diessel, 2001). This

results in the following table which shows for each condition in our materials (see (5)) which hypothesis predicts it to be easier to process compared to the other conditions:

	(a)	(b)	(c)	(d)
perspective shifts	×	✓	×	✓
long-distance-dependency	✓	×	✓	×
information structure(1) order: first given, then new	✓	✓	×	×
information structure(2) new: main clause, given: dependent clause	×	✓	✓	×

Table 1: predictions for each condition based on the different hypotheses

Method

We designed an experiment where RC processing difficulty is measured via a self-paced reading task. The self-paced reading task will be presented online via Ibex farm, and we will recruit 60 monolingual speakers of English via Amazon mechanical Turk. The experiment contains 32 items in four conditions:

(5) Experimental materials

(a) object-modifying, subject extracted RC. Main clause: *given*, dependent clause: new
(b) subject modifying, subject extracted RC. Main clause: *given*, dependent clause: new
(c) object-modifying, subject extracted RC. Main clause: new, dependent clause: *given*
(d) subject-modifying, subject extracted RC. Main clause: new, dependent clause: *given*

IS-status (given/new) is manipulated by a context that precedes the RC-structure:

Intro	My dad recently bought a new camera, and there is no hiding from it. He brings it everywhere he goes, and tries to turn every situation into a hallmark-moment for the family photo album. When we went out for tea the other day,
(a)	*my dad took a candid picture of my mom*, who was eating a lemon cheesecake
(b)	my dad, who was eating a lemon cheesecake, *took a candid picture of my mom*
Wrap-up	He accidentally had the flash turned on, which startled my mom. She knocked her tea over the cheesecake, and made him delete the picture immediately.

Intro	My $mom_{(c)}$/My $dad_{(d)}$ is obsessed with cheesecake. Every time we go out for tea or coffee, she'll get a slice of whatever cheesecake is available, but she is partial to citrus flavours. When we went out for tea the other day,
(c)	my dad took a candid picture of my mom, *who was eating a lemon cheesecake*
(d)	my dad, *who was eating a lemon cheesecake*, took a candid picture of my mom
Wrap-up	He accidentally had the flash turned on, which startled my mom. She knocked her tea over the cheesecake, and made him delete the picture immediately.

The residual reading times will be analysed with a mixed effects model. The results will give us insights into the role IS plays in RC complexity. The current design not only allows us to measure whether IS is as (or more) important for predicting complexity as syntactic factors, but it will also show whether information structural needs are better satisfied by the order of given/new information in a sentence, or by the distribution of given/new information in the dependent/main clause respectively.

References

Diessel (2001). The Ordering Distribution of Main and Adverbial Clauses: A Typological Study. *Language* 77(2), 433-455.

Gibson et al. (2013). The processing complexity of English relative clauses. In M. Sanz, I. Laka, & M. K. Tanenhaus (Eds.), *Language Down the Garden Path* (pp. 149–173).

Gibson et al. (2005). Reading relative clauses in English. *Cognitive Linguistics*, *16*(2), 313–353.

Gordon et al. (2001). Memory interference during language processing. *JEP:LMC*, *27*(6), 1411–1423.

Santi, A., Grillo, N., Molimpakis, E., & Wagner, M. (2019). Processing relative clauses across comprehension and production: similarities and differences. *Language, Cognition and Neuroscience*, *34*(2), 170–189.