

Certain ungrammaticality or uncertain grammaticality: Deciding between frequent errors and infrequent grammatical structures

Maayan Keshev & Aya Meltzer-Asscher, Tel-Aviv University

In noisy-channel models of sentence processing, parsing decisions rely not only on the input itself, but also on similar utterances that could have surfaced in the form of the input due to production or perception error. Such a near-neighbor analysis can eventually be adopted if the analysis faithful to the input is considerably less probable [1-2]. We examine the parser's choices when it is confronted by a locally ungrammatical string. We ask whether the parser is willing to ignore an agreement error and construct an ungrammatical representation, or whether it attempts to remain faithful to the input, in anticipation of forthcoming material that will allow an alternative, grammatical analysis. We further investigate whether the parser's choice depends on the prior probability of such alternative analyses. To answer these questions, we examine the processing of Hebrew relative clauses in which the subject position is vacant, and the verb mismatches the filler in number (1). Would the parser construct a subject relative (SR), ignoring the agreement mismatch? Or would it act on the prediction that further input will allow for a grammatical analysis? Word order properties of Hebrew [3] enable us to consider two cases in which further input allows a grammatical object relative (OR) analysis. One such possibility is the occurrence of a post-verbal subject, a grammatical, but rare, structure in Hebrew (1b). Another possibility is interpreting the vacant subject position as a plural impersonal null subject (1c). This structure is relatively common in Hebrew (and considerably more frequent than post-verbal subjects).

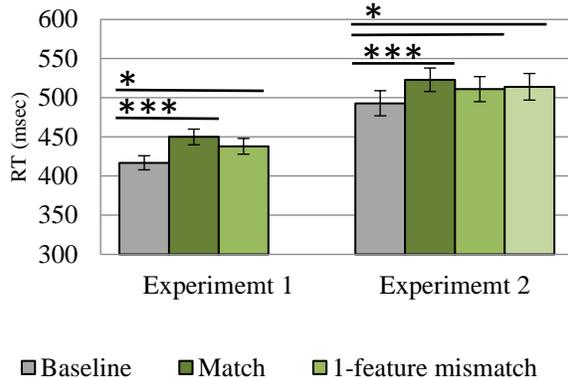
- (1) a. We met the student_i who ____i know the answer. Agreement error
b. We met the student_i who know ____i the teachers well. Post-verbal subject
'we met the student who the teachers know well'
c. We met that student_i who *pro*_{arb} know his_i work. Impersonal null subject
'we met the student who people know his work.'

In four SPR experiments and one acceptability judgment test, we show that the effect of agreement mismatch is modulated by the prior probability of the alternative analysis. In all SPR experiments, the experimental (24-30 sets) as well as the filler items (45-51 items) were globally grammatical. **Experiment 1** included sentences with plural fillers and a singular verb in the relative clause, thus ruling out an impersonal reading (as in 1c), which is restricted to plural verbs in Hebrew. Participants (N=36) tended to ignore agreement and construct an ungrammatical SR, rather than wait for post-verbal subject rendering the structure grammatical. This tendency was evidenced by increased RTs at the post-verbal subject ($p=.03$), relative to an unambiguous baseline condition without a filler-gap dependency. In fact, the mismatch condition aligned with a condition with matching filler-verb dependencies. **Experiment 2** (N=48) replicated the results and extended them to the case of a two-feature (number and gender) mismatch ($p=.037$) (Figure 1A).

Turning to cases in which impersonal null subjects are licensed, **Experiment 3** (N=48) revealed a filled-gap effect at the object position ($p = .009$), suggesting an object relative analysis (with an object gap) was adopted, rather than an SR. A similar increase in RTs was observed ($p=.01$) relative to a matching filler-verb condition, suggesting that an SR is in fact preferred when agreement allows it. In **Experiment 4** (N=36), the filled-gap effect for the mismatching condition replicated in longer filler-verb dependencies (Figure 1B).

These findings provide extension of the noisy channel model to Hebrew, and to number agreement in OR/SR ambiguities. Specifically, the results suggest that during incremental parsing, comprehenders apply elaborate knowledge of the distribution of structures in the language and are willing to construct ungrammatical subject-verb dependencies to avoid improbable grammatical structures. In contrast, the results are not fully compatible with the self-organizing sentence processing approach [4]. This approach suggests that bottom-up activation of competing constructions pulls the interpretation towards a locally coherent (globally ungrammatical) parse. However, in our case, the ungrammatical structure is not locally coherent.

A: Subject relative vs. Object relative with a post-verbal subject



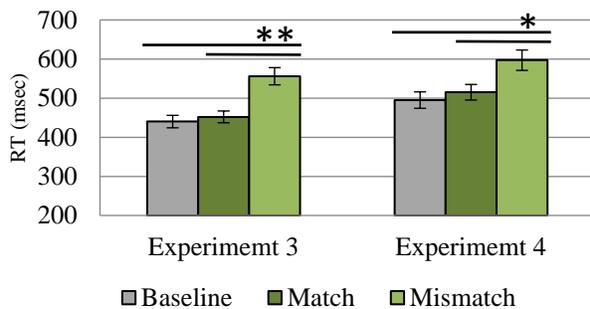
E1: We met the students that, by the end of term, decided(sg) eventually the principal to expel.

E2: We met the students(f) that, by the end of term, decided(sg-m) eventually the principal to expel.

INTERPRETATION: 'We met the students that, by the end of term, the principal decided to expel'.

Baseline: We met because, by the end of term, decided(sg) eventually the principal to expel the students.

B: Subject relative vs. Object relative with an impersonal null subject



E3: We looked for the student that found(pl) eventually the bag of him.

E4: We looked for the student that, at the school yard, found(pl) eventually the bag of him.

INTERPRETATION: 'We looked for the student, the bag of whom was eventually found at the playground'.

Baseline: We looked for the student after they found(pl) eventually the bag of him.

Figure 1. Results of Experiments 1-4. Bracketed abbreviations denote the gender/number marking: pl – plural; sg – singular; f – feminine; m – masculine. Error bars mark +/-1 SE. Analysis was conducted with a linear mixed-model regression; * $p < .05$; ** $p < .01$; *** $p < .001$

References: [1] Levy, et al. (2009). Eye movement evidence that readers maintain and act on uncertainty about past linguistic input. *PNAS*, 106. [2] Gibson, et al. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *PNAS*, 110. [3] Shlonky, (1997). Clause structure and word order in Hebrew and Arabic. OUP. [4] Tabor, et al., (2004). Effects of merely local syntactic coherence on sentence processing. *JML*, 50.