

What to expect when you are expecting an antecedent: processing cataphora in Dutch

Anna Giskes (NTNU) and Dave Kush (University of Toronto; NTNU)

Background: During incremental processing, the parser cannot fully interpret cataphors like *he* in (1) until their antecedent is encountered. Past research has argued that the parser expects the antecedent in the next available syntactic position, often the main subject [e.g., 1-4]. Evidence comes from Gender- and Number Mismatch Effects (G/N-MME): in manipulations like (2), readers slow down at a gender-mismatching subject NP compared to a matching NP.

(1) When he_i had sown the field, the farmer_i baked pancakes for the children.

(2) When **he/she** had sown the field, the boy...

Previous research is uninformative about how far in advance the parser predictively commits to an antecedent in a specific position. MMEs in (2) are compatible with (i) a parser that predictively builds the antecedent in subject position *before* receiving bottom-up input of the subject. However, MMEs may also reflect a parser that (ii) waits until *after* encountering a subject NP to posit coreference (but before gender features are processed bottom-up) [3]. Previous studies do not allow us to tease these two options apart, because MMEs occur at/after the subject NP. We constructed a test of the two hypotheses in Dutch, a V2 language with subject-verb number agreement. We reasoned that if the parser predictively commits to and builds an antecedent in main clause subject position, this should trigger a prediction of matching number agreement on the main verb. Because Dutch is V2, the finite verb will precede the main subject in sentences with fronted adjunct phrases like (1). We therefore looked for N-MMEs between a cataphor and the main verb as evidence for advance prediction of the subject.

Self-paced reading experiments: (exp 1: n=80; exp 2: n=160) We manipulated number-match between the main clause subject and a cataphor in a fronted adjunct clause (Table 1: main clause verb *bakte* underlined). In a control we replaced the finite subordinate clause with a participial clause without a cataphor. The participial clause was ambiguous regarding the number of its implicit 'PRO' subject, thus providing a baseline without an expectation for number. In experiment 2, we added 10 separate items manipulating the gender of the main clause subject (underlined in Table 1), in order to replicate the classic G-MME.

Results See Fig. 1-3. In both SPR experiments, the number manipulation did not yield a significant mismatch effect at the verb or in the spillover regions (maximal LMEMs on log-transformed RTs; for all models, $t < 1.5$). The largest trend towards a N-MME (19 ms) was observed in experiment 2 at the critical main verb ($t = 1.36$). In contrast, we observed a large G-MME in the spillover region for the gender manipulation (85 ms GGME, $t = 5.98$).

Conclusion: The absence of significant NMMEs at the V2 verb suggest that cataphors do not trigger an 'early' prediction of a matching NP in main subject position. The strong G-MME in the same studies suggest that participants still had strong expectations for an antecedent in main subject position. These results are consistent with a parser that does not make a predictive syntactic commitment to locate an expected antecedent in subject position. They are also in line with a parser that does predict the subject to some extent, but does not execute all knock-on consequences following from that prediction. Furthermore, it is possible that the (degree of) prediction varies for number- and gender features, in line with relatively small and late N-MMEs in previous research [2].

The results suggest that at least some active parsing strategies triggered by long-distance dependencies do not reliably entail predictive building of syntactic structure.

References: [1] Kazanina, N. et al. (2007). *JML* 56.3 (2007): 384-409. [2] Van Gompel, R.P. & Liversedge, S.P. (2003). *JEP: Learning, Memory, and Cognition*, 29(1), 128. [3] Drummer, J.D. & Felser, C. (2018). *JML*, 101, 97-113. [4] Kush & Dillon, B. (2020). OSF Preprints, 14 Aug 2020.

Table 1: SPR item set (24 items) for SPR (exp. 1&2). Critical regions underlined. Items were counterbalanced for main verb number (1&2) and subject gender (2). Items for exp. 2 had one additional spillover region following the main clause verb (*the extremely | friendly | farmer...*)¹

Number-match/ mismatch	Nadat <u>hij/zij</u> de akker had/hadden ingezaaid, <u>bakte</u> de vriendelijke boer pannenkoeken voor de kinderen. After <u>he/they</u> the field had.SG/PL sown, <u>baked.SG</u> the friendly farmer pancakes for the children. <i>After he/they had sown the field, the friendly farmer baked pancakes for the children.</i>
PRO	Na de field te hebben ingezaaid, <u>bakte</u> de vriendelijke boer... After the field have.INF sown, <u>baked.SG</u> the friendly farmer... <i>After having sown the field, the friendly farmer baked...</i>
Gender-match/ mismatch (only exp. 2)	Nadat <u>hij/zij</u> de vliegtickets had gekocht, schreef <u>Diana</u> meteen de datum van haar/Jans aankomst op. After <u>he/she</u> the airline tickets had.SG bought, wrote.SG <u>Diana</u> immediately the date of her/Jan's arrival up.

Figure 1

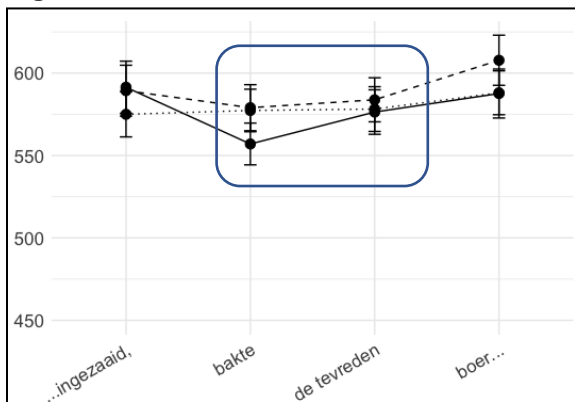


Figure 2

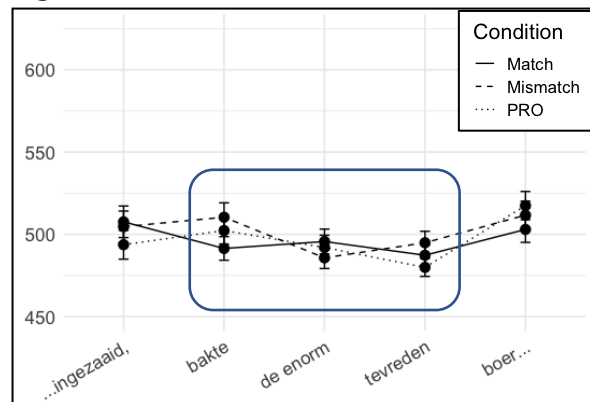


Figure 3

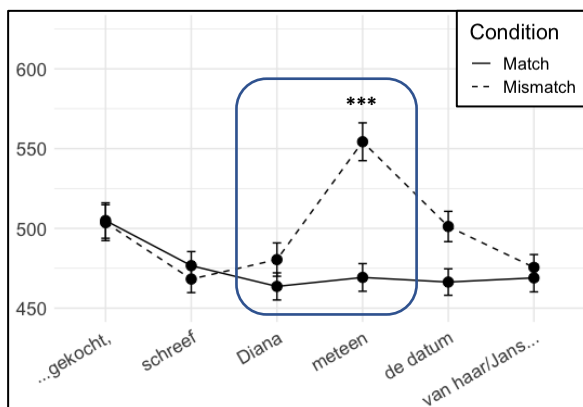


Figure 1: Average RTs + se for experiment 1. Analyzed regions (main verb *bakte* + spillover region) in the boxed area.

Figure 2: Average RTs + se for the number manipulation of experiment 2. Analyzed regions (main verb *bakte* + 2 spillover regions) in the boxed area.

Figure 3: Average RTs + se for the gender manipulation of experiment 2. Analyzed regions (main subject *Diana* + 1 spillover region) in the boxed area.

¹ The Dutch pronoun *zij* is ambiguous between sing-fem, and pl (both masc. and fem.) The number on the auxiliary in the embedded clause (*had/hadden*) disambiguates the pronoun.