Gap prediction is not fully constrained by grammar: Hebrew Maze results

Mandy Cartner & Aya Meltzer-Asscher (Tel Aviv University)

Dependency resolution is a predictive process. In filler-gap dependencies, a filler triggers an active search for its gap, reflected in the filled-gap effect (FGE): increased difficulty when a potential gap position is filled [1]. Likewise, in cataphoric dependencies, a cataphor triggers an active search for an antecedent, reflected in difficulty when a potential antecedent mismatches the cataphor's features [2-3]. We investigate readers' online predictions when these two searches are triggered simultaneously, in sentences headed by a "reflexive-filler", e.g., "Which picture of herself". Grammatically, the two searches triggered by the reflexive-filler are not independent: the antecedent must be hierarchically higher than the gap, and bind the reflexive in its base position. We investigate whether this grammatical constraint informs online predictions, focusing on filled-gap (FG) sentences, e.g. (1).

(1) Which picture of <u>herself</u> did the {queen | king} see **the girl** admiring _?

When the matrix subject matches the reflexive's features ('queen'), a gap would be grammatical in the FG position (object of 'see'); when they mismatch ('king'), it would not. Does this grammatical constraint affect gap prediction, modulating the FGE on 'the girl'? If the parser is grammar-driven, gap prediction should halt until an antecedent is identified, and no FGE should arise after a mismatch (as in Islands [1,4-5]). However, if the parser prioritizes short filler-gap dependencies even when illicit, the FGE should remain.

We report one SPR and two Maze experiments on Hebrew. For each experiment, one brm model was fit to subject RTs, and a separate model was fit to object (FG) RTs.

Exp 1. 60 participants read (via SPR) 24 item sets, manipulating **agreement** (Gender *match*/*mismatch* between the reflexive and subject) and **structure** (*refl-filler/baseline*) (Table 1). <u>Subject</u>: a *mismatching* subject was read slower than a *matching* one following a *refl-filler*, but not in the *baseline*. <u>Object</u>: no slowdown in *refl-filler* compared to *baseline* (no FGE), and no interaction with *agreement* (Figure 1, Table 2).

Exp 2 is a replication of Exp 1 using the G-Maze task. <u>Subject</u>: Replicating the SPR results, a *mismatching* subject following a *refl-filler* caused a slow-down. <u>Object</u>: object RTs were slower in the *refl-filler* conditions than the *baseline* (FGE), both in the *match* condition and, to a lesser extent, in the *mismatch* condition (Figure 2a-b, Table 3).

Exp 3 manipulated the degree of mismatch between the reflexive and the subject, to test whether illicit gap prediction is the result of rational misinterpretation [6]. 90 participants read (via G-Maze) 36 item sets, crossing: **structure** (*refl-filler/baseline*) and **agreement** (*match*/*mismatch*/*2-mismatch)*; in *2-mismatch* the subject mismatches the reflexive in Gen+Num (Table 1). <u>Subject</u>: *mismatching* subjects were read slower in the *refl-filler* structure compared to *baseline*. <u>Object</u>: longer RTs in the *refl-filler* compared to the *baseline* (FGE). This slow-down was smaller, but still reliable, in the *mismatch* condition. There was no evidence for a reliable contrast between *mismatch* and *2-mismatch* (Figure 4, Table 4).

Discussion. Across the three studies, we find a reliable slowdown at a mismatching subject following a fronted reflexive, indicating an active antecedent search [cf. 2-3]. In the Maze studies, we find evidence for an object FGE. Exp 1 (SPR) did not produce a FGE, which we hypothesize is a task effect: SPR is rather passive, and may produce inattentive participants. This is supported by lower comprehension accuracy in the SPR (66%) vs Maze (88%).

In the Maze studies, the FGE was modulated by the presence of an antecedent, such that the FGE was smaller following a mismatching subject. This suggests that gap prediction is sensitive to the requirements of the reflexive-filler. However, the FGE persisted in the *mismatch* condition, which is unpredicted if the parser were grammatically driven: grammatically, no gap should be posited unless an antecedent was identified. Exp 3 was designed to test the hypothesis that a FGE after a mismatch is the result of rationalization over noisy reading [6], i.e., participants misrepresent the reflexive's features, as a mismatch is unlikely, and only thusly predict a gap. In the *2-Mismatch* condition, misrepresentation is less likely, and the FGE should diminish further. This hypothesis was not corroborated, which suggests that the FGE after a mismatch reflects processing pressures overriding grammar, such that gap prediction persists despite ungrammaticality. Interestingly, this is unlike Island structures, which are evidenced to fully block gap prediction [1,4-5].

Table 1. Simplified stimuli from Experiments 1-3. Gender was counterbalanced across sets.

structure	agreement	Example sentence, originally in Hebrew
reflexive-filler	match	*Which video of himself **the-detective.M** filmed.M the-suspect.M delete?*
	mismatch	*Which video of himself **the-detective.F** filmed.F the-suspect.M delete?*
	2-mismatch Exp3	*Which video of himself **the-detective.FPL** filmed.FPL the-suspect.M delete?*
baseline	match	*When **the-detective.M** filmed.M the-suspect.M delete video of himself?*
	mismatch	*When **the-detective.F** filmed.F the-suspect.M delete video of himself?*
	2-mismatch Exp3	*When **the-detective.FPL** filmed.FPL the-suspect.M delete video of himself?*

Figure 1. Subj & Obj log RTs Exp 1

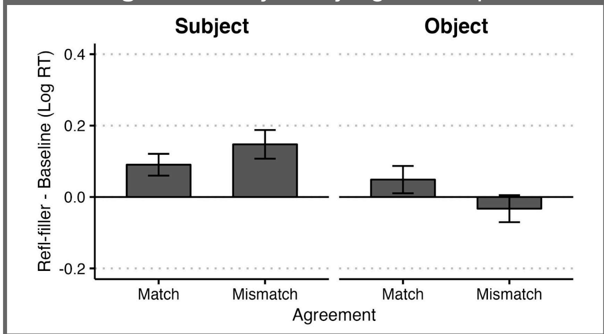


Figure 2a. Subj & Obj log RTs Exp 2

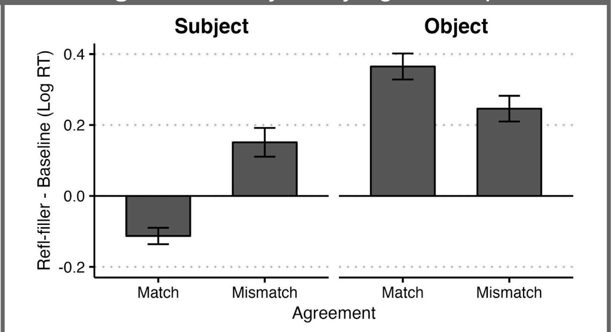


Figure 3. Subj & Obj log RTs Exp 3

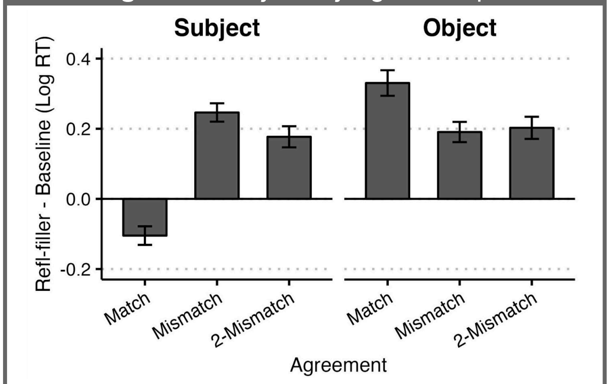
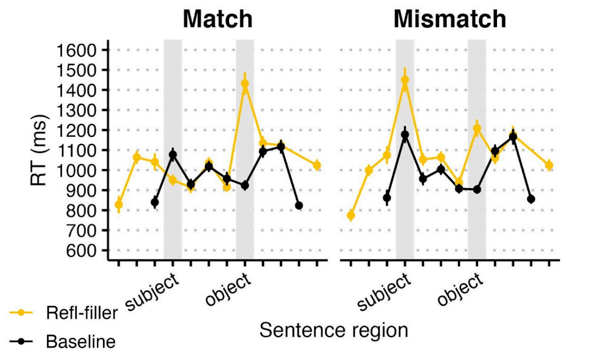


Figure 2b. Word-by-word RTs Exp 2



	Subject region					Object region			
Table 2. Experiment 1 brm models									
factor	**Est**	**SE**	**l-95%**	**u-95%**	**factor**	**Est**	**SE**	**l-95%**	**u-95%**
structure	0.04	0.01	0.01	0.06	structure	0.00	0.01	-0.03	0.03
gender	0.01	0.01	-0.01	0.04	gender	0.01	0.01	-0.02	0.03
stru:gen	0.03	0.01	0.00	0.05	stru:gen	-0.00	0.01	-0.03	0.02
Table 3. Experiment 2 brm models									
factor	**Est**	**SE**	**l-95%**	**u-95%**	**factor**	**Est**	**SE**	**l-95%**	**u-95%**
structure	0.01	0.02	-0.02	0.05	structure	-0.15	0.02	-0.18	-0.12
gender	0.10	0.02	0.07	0.13	gender	-0.04	0.01	-0.06	-0.01
stru:gen	0.07	0.01	0.04	0.10	stru:gen	0.03	0.01	0.01	0.06
Table 4. Experiment 3 brm models									
factor	**Est**	**SE**	**l-95%**	**u-95%**	**factor**	**Est**	**SE**	**l-95%**	**u-95%**
structure	-0.05	0.01	-0.08	-0.03	structure	-0.11	0.01	-0.014	-0.09
mat/1mis	0.06	0.01	0.05	0.07	mat/1mis	-0.03	0.01	-0.04	-0.01
1mis/2mis	0.02	0.01	-0.00	0.04	1mis/2mis	-0.00	0.01	-0.02	0.02
stru:mat/1mis	-0.05	0.01	-0.06	-0.04	stru:mat/1mis	0.02	0.01	0.01	0.04
stru:1mis/2mis	0.02	0.01	0.00	0.0	stru:1mis/2mis	-0.00	0.01	-0.02	0.02

References: [1] Stowe 1986, *Lang Cogn Process*; [2] Kazanina et al. 2007, *JML*; [3] Giskes & Kush 2021, *Mem Cognit.*; [4] Traxler & Pickering 1992, *JML*; [5] Keshev & Meltzer-Asscher 2017, *Language.*; [6] Levy 2008, *EMNLP*.