Evidence for the role of abstract subjecthood in Turkish agreement comprehension

Özge Bakay[1], Utku Türk[2], Duygu Demiray[1], Brian Dillon[1]
University of Massachusetts Amherst, University of Maryland College Park

Agreement attraction is a type of grammaticality illusion where a plural attractor noun makes an ungrammatical sentence appear acceptable, as in "*The key to the cabinets are rusty.*' What drives such errors in agreement comprehension? Retrieval-based accounts attribute these errors to occasional erroneous retrieval of the plural attractor from memory due to either its surface features matching those of the verb like number or case[1-2] or the association of these cues with an abstract notion of agreement controller like subjecthood.[3-6] An alternative account posits that comprehenders rely on probabilistic inferences on the strings they read and perform a repair of the string into a more probable, well-formed alternative, such as assuming a plural target as in *the keys to the cabinets are rusty*.[7-8] Here we directly compared the predictions of these accounts using Turkish possessive constructions featuring GEN(itive) vs. bare attractors. While both attractors can surface as subjects and thus are associated with being an agreement controller in Turkish,[4,9] only bare attractors match the surface features of a tensed verb. Conversely, corpus data show that GEN-marked attractors generate higher probabilities for a grammatical, plural agreement target than bare attractors. The results of our judgment task showed evidence for comparable attraction errors with both attractors, suggesting that <u>abstract structural cues associated with subjecthood, rather than surface cue matching or probabilistic repair, plays a dominant role in driving attraction effects.</u>

EXPERIMENT. A speeded, binary acceptability task ($N_{participant}$=59, N_{item}=24) compared agreement errors with two attractor types in a 2x3 within-subject design, crossing attractor CASE (GEN vs. bare) and AGREEMENT MATCH (<u>target</u>, ***attractor*** or none) (see 1). The verb was always plural marked. Only the sentences with a matching (plural) target were grammatical.

(1) a. TARGET MATCH   b. ATTRACTOR MATCH   c. NO MATCH

Mülteci(-nin) <u>avukat-lar-ı</u>   ****Mülteci-ler-(in)*** <u>avukat-ı</u>   ****Mülteci-(nin)*** <u>avukat-ı</u>
refugee(-GEN) lawyer-PL-POSS   refugee-PL(-GEN) lawyer-POSS   refugee(-GEN) lawyer-POSS

… duruşmada-da   durmadan   bağır-dı-lar.
… at.hearing   non-stop   scream-PST-PL
'The refugee(s)'s lawyer(s) shouted at the hearing non-stop.'

PREDICTIONS. In <u>retrieval-based accounts</u>, agreement attraction is driven by the overlap of its surface features with the verb's retrieval cues or the attractor's abstract association with being an agreement controller. Accounts focused on canonical surface features such as number or case predict higher attraction with bare attractors, as their surface cues are more similar to those of the bare, plural agreement target in the sentence. On the other hand, accounts centered on abstract associations predict comparable attraction effects with both attractors, as both can distributionally surface as subjects, and thus are associated with being an agreement controller. As a tool to index

Fig1. Attention as a correlate of high-level agreement controllerhood abstraction. Y-axis shows difference between Attention from Verb to Attractor vs. Target.

this, we evaluated the predictions of a recent <u>transformer-based model of retrieval</u>:[10-11] We calculated the BERT-internal attention weights toward target vs. attractor in the experimental stimuli that we tested. This analysis (**Fig1**) predicted that both attractors should elicit greater attention relative to the target when they are plural (ATTRACTOR MATCH) compared to when they are singular (TARGET MATCH), aligning with abstract structural accounts of retrieval.
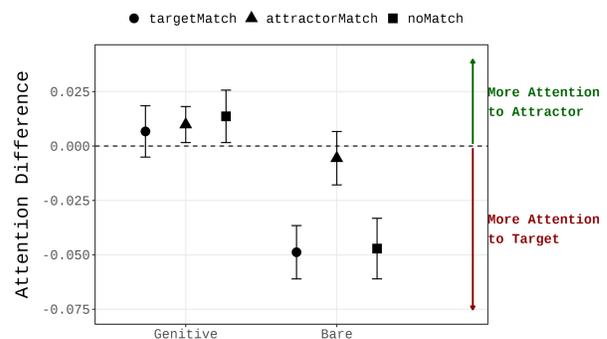
Under <u>probabilistic inference accounts</u>, comprehension involves a repair mechanism where comprehenders evaluate the likelihood of alternative well-formed strings of the perceived input due to potential noise.[12] Attraction effects are thus predicted to be modulated by the frequency of the relevant alternative form of the target; with increased attraction errors when the prior probability of the target being plural is higher in the context of the attractor. To quantify this prediction, we ran **CORPUS ANALYSES**, where we calculated the conditional probability of the number features of the target given the number features of the attractor (Table 1). We found that with GEN-marked attractors, a plural target is more likely with a plural than a singular attractor (24% vs. 11%). With bare attractors, the pattern reverses: a plural target is less likely with a plural than a singular attractor (5% vs. 17%). This predicts attraction with plural GEN-marked attractors but weak(er) or lack of attraction with plural bare attractors.

Attractor GEN	Target	Counts (out of 1m)	Pr(TargetPL | Att)	Attractor BARE	Target	Counts (out of 1m)	Pr(TargetPL | Att)
SG	SG	4901.8	11.4%	SG	SG	18407	17%
	PL	631.5			PL	3842.4	
PL	SG	1684.7	**24%**	PL	SG	655	5%
	PL	540.3			PL	37	

Table 1. Counts for each combination of number features in [attractor, target] tokens in the TrTenTen corpus (~4.9 billion words). Rational inference predictions are highlighted. TargetPL = plural target, Att = attractor.

RESULTS. For both CASE markings, participants gave more 'yes' (=acceptable) responses to ungrammatical sentences in ATTRACTOR MATCH (M_{GEN}=34.3%, M_{BARE}=28.4%) than in NO MATCH conditions (M_{GEN}=29.2%, M_{BARE}=18.6%) (**Fig2**). Our crossed Bayesian GLM model (**Fig3**) verified this: There was strong evidence for an overall effect in the ATTRACTOR vs. NO MATCH comparison (β = 0.26 [.05,.48], P(β>0) = 0.99), but this did not interact with CASE (β = -0.05 [-.29,.20], P(β>0) = 0.34). A nested model further showed evidence for ATTRACTOR vs. NO MATCH difference with GEN-marked (β = 0.20 [-.10,.51], P(β>0) = 0.91) and bare attractors (β = 0.31 [-.01,.64], P(β>0) = 0.97) separately.
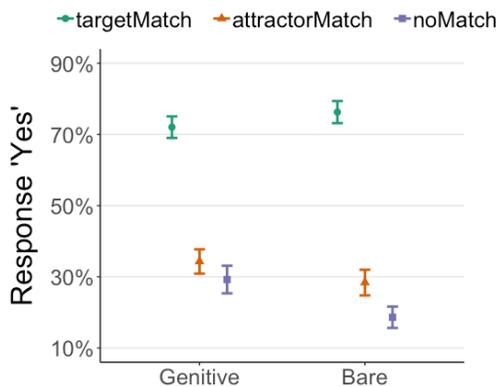

Fig2. Mean percentages of 'Yes' responses in the judgment task. Error bars show by-participant SEs.

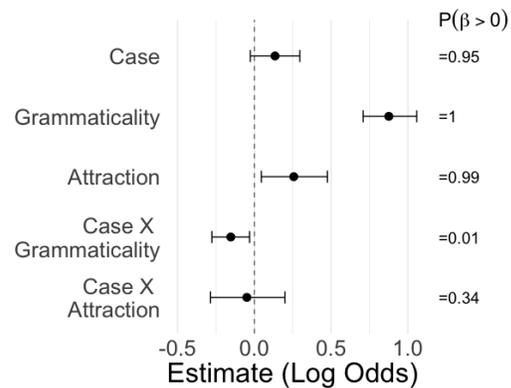**Fig3.** Posterior summaries (mean, 95% CrI, and P(β>0)) from the crossed Bayesian GLM model. Case = GEN vs. bare; Grammaticality = targetMatch vs. attractorMatch/noMatch; Attraction = attractorMatch vs. NoMatch.

OVERALL, we found similar rates of attraction errors with two different attractors. While both attractors are associated with subjecthood, they differ in their surface cue overlap with the verb and their co-occurrence probabilities with alternative grammatical (plural) targets. This suggests evidence that abstract structural cues associated with being an agreement controller is an important determining factor in agreement errors. These findings also align with the predictions of a large language model.

[1] Wagers et al., 2009 [2] Lewis & Vasishth, 2005 [3] Lago et al., 2019 [4] Türk & Logačev, 2024 [5] Bhatia & Dillon, 2022 [6] Belotu & Dillon, 2024 [7] Ryskin et al., 2021 [8] Cartner et al., 2025 [9] Göksel & Kerslake, 2005 [10] Ryu & Lewis, 2025 [11] Viota et al., 2019