Tracking structural cues or relying on probabilistic inference in Turkish agreement?

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

Agreement attraction is a type of illusion where a plural attractor noun makes an ungrammatical sentence appear acceptable, as in "*The key to the cabinets are rusty.' What factors drive such errors in agreement comprehension? Prior research has attributed these errors to various mechanisms, including (i) occasional erroneous retrieval of the plural attractor due to its feature match with the verb[1-2] or (ii) its abstract structural role, such as its likelihood of being an agreement controller;[3-6] or (iii) rational inference to repair the perceived string into a more probable, well-formed alternative.[7-8] Here we directly compare the predictions of these accounts using Turkish possessive constructions featuring GEN(itive) vs. NOM(inative)-marked (bare) attractors. While both case markings are associated with the abstract structural role of being an agreement controller in Turkish,[4,9] GEN-marked attractors generate higher probabilities for a grammatical, plural agreement target than bare attractors. We show evidence for comparable attraction errors across these two case markings, suggesting that the potential to be an agreement controller plays a more dominant role in attraction effects.

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 (target, attractor or none) (see 1). The verb was always plural marked. Only the sentences with a matching (plural) target were grammatical.

PREDICTIONS. Retrieval-based accounts predict that agreement attraction is driven by the attractor's abstract structural role or the overlap of its surface features with the verb's retrieval cues. Accounts centered on association with abstract controllerhood predict comparable attraction effects with both attractors, as both are potential structural controllers in the language. On the other hand, accounts focused on canonical surface features (e.g., number or case) predict higher attraction with bare attractors, as they are more similar to the target in the sentence. As a tool to index this association, we evaluated the predictions of a recent transformer-based model of retrieval: [10-11] We calculated the BERT-internal attention weights directed towards the target vs. attractor. This analysis (**Fig3**) predicted that both attractors should elicit comparable attention, aligning with controller-based accounts. Under probabilistic inference accounts, comprehension involves a repair mechanism where listeners consider 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 (see 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.

RESULTS. For both CASE markings, participants gave more 'yes' responses in ATTRACTOR MATCH (M_{GEN}=34.3, M_{BARE}=28.4) than NO MATCH conditions (M_{GEN}=29.2, M_{BARE}=18.6) (**Fig1**). Our crossed Bayesian GLM model (**Fig2**) 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).

OVERALL, we found similar rates of attraction errors with two differently case marked attractors, which are both controller-like but differ in their co-occurrence probabilities with alternative, grammatical target nouns. This suggests evidence that surface cues and their associations, rather than probabilistic rational inference, is an important determining factor in agreement errors. These findings are also somewhat captured by the predictions of a large language model.

Stimuli. The versions with *–(n)In* show the Genitive conditions, and those without show the Nominative (bare) conditions. Agreement target is underlined; and attractor is bolded.

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.'

Attractor GEN	Target	Counts (out of 1m)	Pr(TargetPL | Att)	Attractor NOM	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.
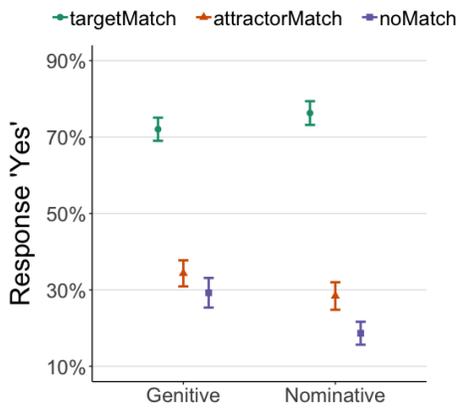

Fig1. Mean percentages of 'Yes' responses in the judgment task. Error bars show by-participant SEs.
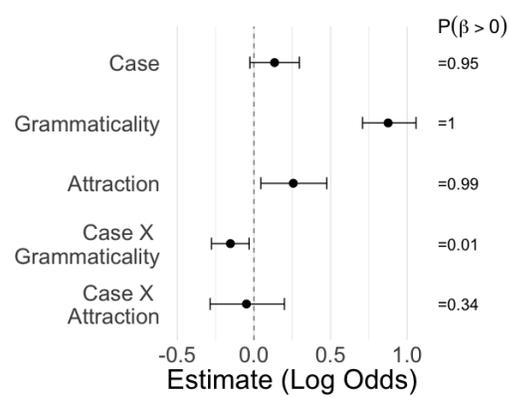

Fig2. Posterior summaries (mean, 95% CrI, and P(β>0)) from the Bayesian GLM model. Case = GEN vs. bare; Grammaticality = targetMatch vs. attractorMatch/noMatch; Attraction = attractorMatch vs. NoMatch.
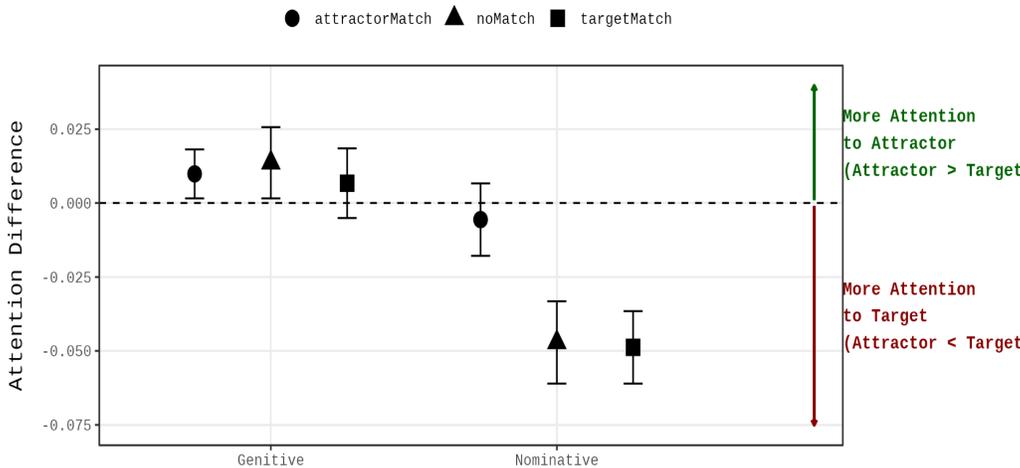

Fig3. Attention as a correlate of high-level controllerhood abstraction. Y-axis shows the difference between Attention from verb to Attractor vs. Target.

References:
[1] Wagers et al., 2009
[2] Lewis & Vasishth, 2005
[3] Lago et al., 2019
[4] Türk & Logacev, 2024
[5] Bhatia & Dillon, 2022
[6] Bleotu & Dillon, 2024
[7] Ryskin, Bergen & Gibson, 2021
[8] Cartner et al., 2025
[9] Göksel & Kerslake, 2005
[10] Ryu & Lewis, 2025
[11] Viota et al., 2019
[12] Levy, 2008
[13] Patson & Husband, 2016
[14] Schuleter et al., 2019