

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

Ozge Bakay, Utku Turk, Duygu Demiray, and Brian Dillon

Abstract

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.*' This study investigates the mechanisms driving such errors in Turkish, focusing on the processing of nominal vs. genitive attractors. Prior research contrasts retrieval-based accounts, which implicate feature mismatches or structural confusion, with probabilistic inference accounts, which view attraction as a rational repair strategy. Using a speeded acceptability task ($N=59$), we analyzed the impact of attractor case (Nominative vs. Genitive) and number match on agreement processing. Corpus analyses indicated higher transitional probabilities for plural targets following Genitive attractors. However, our Bayesian analysis revealed comparable attraction effects for both case markings, supporting structural controllerhood theories over surface cue-based or pure probabilistic inference accounts. These findings suggest that the structural potential to control agreement outweighs surface feature overlap or transitional probabilities in driving attraction effects.

Keywords: Turkish, psycholinguistics, agreement attraction, memory retrieval, Bayesian analysis

Introduction

Agreement attraction is a pervasive phenomenon in sentence processing where a verb agrees with a noun phrase other than its grammatical subject. For instance, in the sentence "*The key to the cabinets are rusty,*" the plural attractor "cabinets" creates an illusion of acceptability despite the singular subject "key." This phenomenon has been extensively studied to understand the mechanisms of memory retrieval and structural dependency formation during sentence comprehension.

Models of Agreement Attraction

Two primary classes of theories have been proposed to explain agreement attraction: retrieval-based accounts and probabilistic inference accounts.

Author roles were classified using the Contributor Role Taxonomy (CRediT; <https://credit.niso.org/>) as follows: Ozge Bakay: Conceptualization, Methodology, Descriptive and Statistical Analysis, Writing - Original Draft; Utku Turk: Conceptualization, Methodology, Descriptive and Statistical Analysis, Writing - Original Draft, Large Language Model Analysis; Duygu Demiray: Large Language Model Analysis, Writing - Original Draft; Brian Dillon: Funding Acquisition, Conceptualization, Writing - Review & Editing

Correspondence concerning this article should be addressed to Ozge Bakay, Email: obakay@umass.edu

Retrieval-Based Accounts

Retrieval-based accounts posit that attraction arises from the architectural constraints of the working memory system. Following the cue-based retrieval framework Vasishth and Lewis (2008); Wagers et al. (2009), dependency formation relies on a content-addressable memory system. When the parser encounters a verb, it initiates a search using a set of retrieval cues. The nature of these cues, however, remains a point of debate, leading to two distinct sets of predictions for the current study.

1. Surface Feature Cues: Standard implementations assume that retrieval cues reflect the morphosyntactic features of the expected subject, such as {+Plural, +Nominative}. Under this view, attraction is driven by the feature overlap between the attractor and the retrieval cues. Since Turkish subjects are canonically Nominative (bare), a Nominative attractor provides a better match to the +Nominative cue than a Genitive attractor. Consequently, a surface feature account predicts **stronger attraction for Nominative (Bare) attractors** than for Genitive attractors.

2. Structural Controllerhood: Alternatively, the retrieval system may prioritize items that hold a structural status associated with agreement control. Bhatia and Dillon (2011) proposed the “structural controllerhood” hypothesis, suggesting that nouns in positions that typically license agreement are more accessible or “active” in memory. In Turkish, both Nominative subjects and Genitive possessors serve as agreement controllers. If the parser targets “controller-like” structural representations, both attractor types should be strong candidates for retrieval. Thus, the structural controllerhood hypothesis predicts **comparable attraction effects** for both Genitive and Nominative attractors, despite their case differences.

Probabilistic Inference and Rational Adaptation

Alternatively, probabilistic inference accounts, often framed within the noisy-channel framework, view language comprehension as a process of rational inference under uncertainty. From this perspective, the parser assumes that the linguistic input is liable to noise (e.g., production errors, perceptual/transmission noise). The goal of the comprehender is to reconstruct the intended message (S) given the noisy perceptual input (I) by maximizing the posterior probability $P(S|I) \propto P(I|S)P(S)$.

In the context of agreement attraction, the listener might infer that the speaker intended to say a grammatical alternative (e.g., “*The keys to the cabinets are rusty*”) but produced an error, or that they misheard the input. Crucially, this inference is sensitive to prior probabilities ($P(S)$). If the context makes a plural subject highly probable, or if the transition from the attractor to a plural verb is highly frequent, the parser is more likely to “hallucinate” a grammatical structure, leading to higher acceptance rates Linzen et al. (2016). Thus, under this view, attraction is not a failure of memory, but an optimal adaptation to a noisy environment.

Recent work by Cartner et al. (2025) has specifically explored these probabilistic dynamics in Turkish. They argue that the strong predictive validity of Genitive case for upcoming agreement targets should make the parser more sensitive to number mismatches in Genitive constructions. Specifically, if the system is tracking the conditional probability $P(\text{Plural Target}|\text{Attractor})$, the high predictability of the Genitive attractor should lead to stronger “repair” effects when a plural marker is encountered on the verb, predicting a larger attraction effect compared to the less predictive Nominative attractors. This sets up a clear contrast with structural accounts, which predict no such modulation by case-based probabilities.

The Present Study: Turkish Possessives

Turkish offers a unique testing ground for distinguishing these accounts. In Turkish possessive constructions, attractors can be marked with either Genitive (GEN) case or left bare (Nominative/NOM). - **Genitive Attractors:** In phrases like *Mülteci-nin avukatı* (‘The refugee’s lawyer’), the genitive noun is a structural possessor. In Turkish, possessors control agreement on the head noun, making them highly “controller-like”

structurally. - **Nominative Attractors:** Nominative nouns are the canonical subjects and agreement controllers in the language.

While both are associated with the abstract structural role of being an agreement controller, they differ significantly in their distributional properties. Our corpus analysis (see Results) shows that GEN-marked attractors strongly predict a following plural target (agreement bearer), whereas bare attractors do not.

If attraction is driven by **probabilistic inference** (tracking surface statistics and transitional probabilities), we should see significantly stronger attraction effects with Genitive attractors, which strongly predict plural agreement targets. Conversely, if attraction is driven by **structural controllerhood** (retrieving items that structurally resemble controllers), both Genitive and Nominative attractors should elicit comparable attraction effects, as both represent high-salience controller roles in the Turkish grammar.

In this study, we directly compare these predictions using a speeded acceptability judgment task. We manipulate the case of the attractor (GEN vs. Bare) to determine whether the agreement mechanism is sensitive to fine-grained probabilistic expectations or robust structural cues.

Methods

Participants

Fifty-nine participants were recruited for this study ($N = 59$). All participants were native speakers of Turkish and gave informed consent.

Materials

We constructed 24 sets of experimental items in a 2x3 within-subjects design. The two factors were **Attractor Case** (Genitive vs. Nominative/Bare) and **Agreement Match** (Target Match, Attractor Match, No Match).

In the **Target Match** condition, the head noun (Target) was plural, and the verb was plural (Grammatical). In the **Attractor Match** condition, the attractor was plural, the target was singular, but the verb was plural (Ungrammatical, Attraction). In the **No Match** condition, both the target and attractor were singular, and the verb was plural (Ungrammatical, Baseline).

The stimuli were constructed such that the Genitive attractor appeared in a possessor position, while the Nominative attractor appeared in a modifier position within the noun phrase.

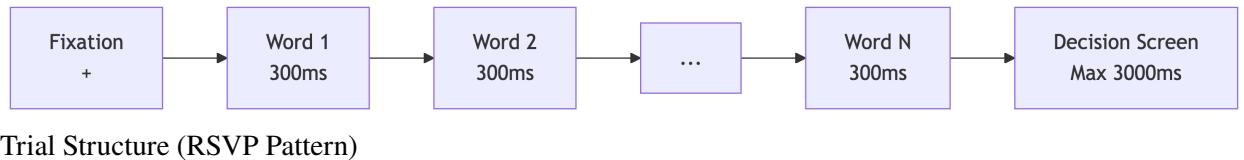
Table 1. Example Stimuli

Condition	Example Sentence	Gloss
Target Match (GEN)	<i>Mülteci-nin avukat-lar-ı bağırdılar.</i>	Refugee-GEN lawyer-PL-POSS shouted.
Attractor Match (GEN)	<i>Mülteci-ler-in avukat-ı bağırdılar.</i>	Refugee-PL-GEN lawyer-POSS shouted.
No Match (GEN)	<i>Mülteci-nin avukat-ı bağırdılar.</i>	Refugee-GEN lawyer-POSS shouted.

Procedure

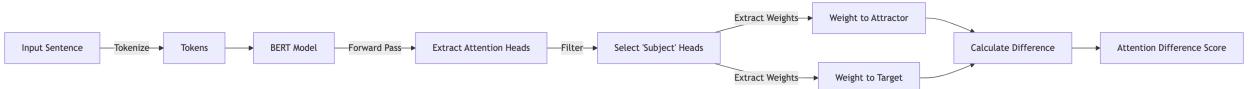
Participants were recruited online and the experiment was hosted on the **PCIBex** farm. The experiment followed a **Latin Square** design, ensuring that each participant saw only one version of each item set, distributed across lists. Sentences were presented word-by-word (RSVP) to ensure timed reading: each word

was presented for **300 ms**, and participants had a maximum of **3000 ms** to make their decision. Participants were instructed to judge whether the sentence was acceptable or not as quickly and accurately as possible.



Computational Modeling

To complement our behavioral results, we analyzed the internal representations of Transformer-based Large Language Models (LLMs), specifically BERT, using methods established by (?) and Ryu and Lewis (2025). Recent work in computational psycholinguistics suggests that attention heads in these models can specialize in tracking specific syntactic dependencies, serving as a computational proxy for the “retrieval cues” or structural access mechanisms in human processing.



Attention Analysis Pipeline

We identified attention heads that reliably track the subject (n_{subj}) dependency in Turkish. For each experimental sentence, we extracted the attention weights from the verb to the **Target Subject** (w_{target}) and the **Attractor** ($w_{attractor}$). We calculated the **Attention Difference**:

$$\text{AttnDiff} = w_{attractor} - w_{target}$$

Positive values indicate that the model attends more to the attractor than the grammatical subject, signaling potential interference or “attraction” in the model’s dependency resolution process.

Data Analysis

We analyzed the response accuracy using **Bayesian generalized linear mixed-effects models (GLMM)** with a Bernoulli link function, implemented in the `brms` package in R. The models included fixed effects for Case and Agreement Match, as well as their interaction. We employed a custom contrast coding scheme to isolate the relevant theoretical comparisons: 1. **Grammaticality**: Comparing the Grammatical condition (Target Match) against the mean of the two Ungrammatical conditions (Attractor Match + No Match). (Weights: 2, -1, -1) 2. **Attraction**: Comparing the Attractor Match condition against the No Match baseline within the ungrammatical sentences. (Weights: 0, 1, -1) 3. **Case**: Sum-coded (Genitive = 1, Nominative = -1).

Exclusion Criteria: Participants who performed with less than **80% accuracy on filler items** were excluded from the analysis to ensure data quality. We used a **maximal random effects structure**, including random intercepts and slopes for all fixed effects and their interactions by both participants and items, to ensure the generalizability of our results.

Results

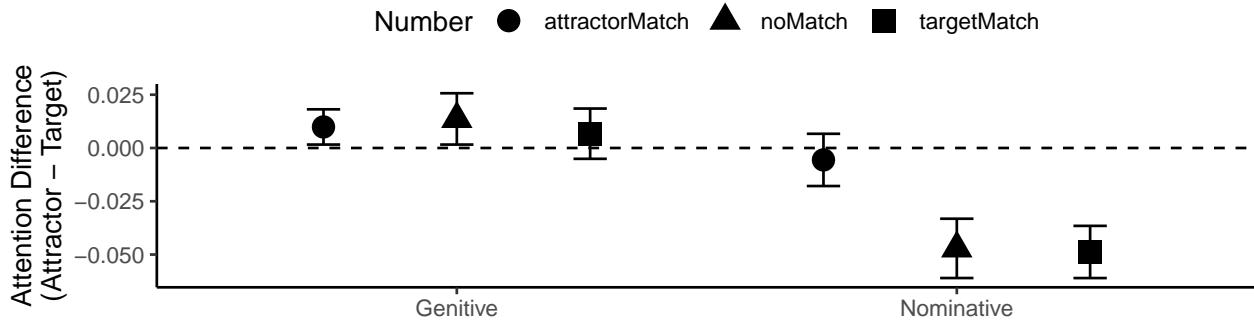
Computational Modeling Results

Figure 1 presents the attention difference scores from the identified “Subject heads” in Turkish BERT. The model shows positive attention difference (interference) for both Genitive and Nominative attractors when they are plural. Crucially, the model does not show a strong preference for the Genitive attractor over

the Nominative one; both receive significant attention, mirroring the “structural controllerhood” finding in humans.

Figure 1

Attention Difference (Attractor - Target) in Turkish BERT. Positive values indicate higher attention to the attractor.



Data Processing

We collected data from 59 participants. 0 participants were excluded due to low accuracy (< 80%) on practice items, leaving a final sample of 59 participants for analysis. THIS NUMBERS ARE FAKE RIGHT NOW.

Descriptive Statistics

Figure 2 shows the mean acceptance rates across conditions.

For Genitive attractors, participants gave more ‘yes’ responses in the Attractor Match condition ($M = 34.3\%$) compared to the No Match condition ($M = 29.2\%$). Similarly, for Nominative attractors, acceptance rates were higher in the Attractor Match condition ($M = 28.4\%$) than in the No Match condition ($M = 18.6\%$). This indicates a clear attraction effect across both case markings.

Figure 2

Mean percentages of ‘Yes’ responses. Error bars represent standard errors.



Corpus Analysis

To assess the probabilistic predictions, we analyzed the TrTenTen corpus (~4.9 billion words). We calculated the conditional probability of a plural target given a plural attractor for both case markings. The results (Table 2) show that a plural target is **more likely** following a plural GEN-marked attractor (24%) compared to a singular one (11.4%). In contrast, for NOM-marked attractors, a plural target is **less likely** following a plural attractor (5%) than a singular one (17%). If attraction were driven purely by predictive probability (rational inference), we would expect significantly larger attraction effects for the Genitive condition.

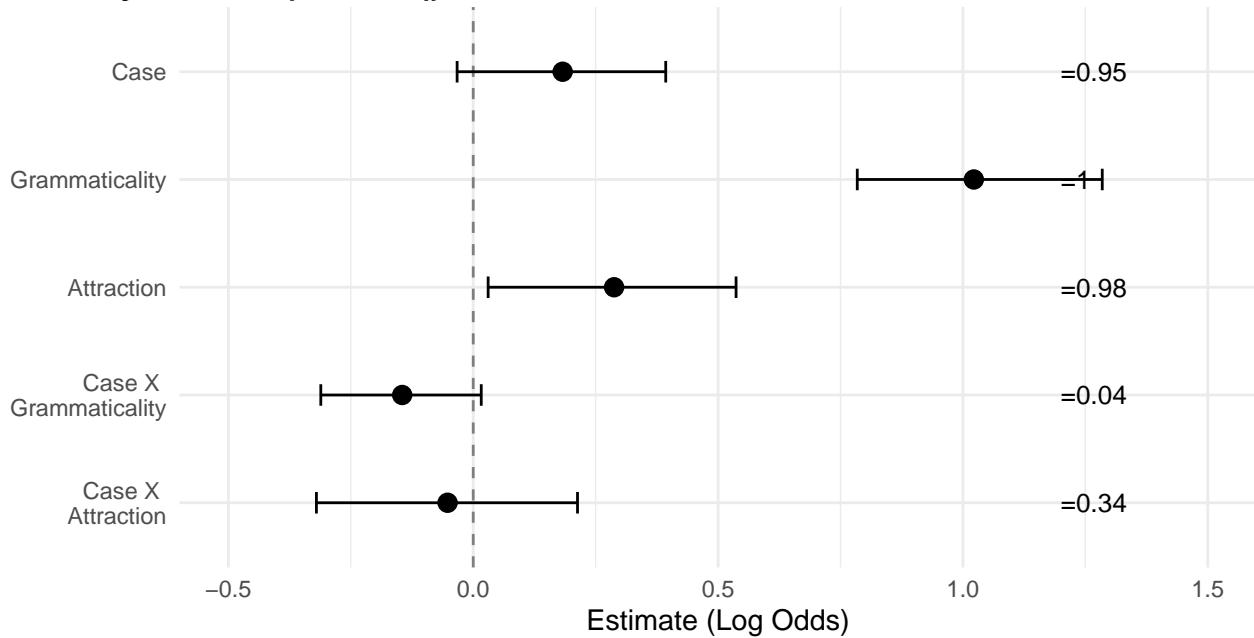
Bayesian Model Results

We fitted a Bayesian logistic regression model to quantify the effects. We utilized a custom contrast coding scheme: 1. **Grammaticality**: Target Match vs. (Attractor Match + No Match) 2. **Attraction**: Attractor Match vs. No Match

The model summary (Figure 3) reveals a robust main effect of **Attraction** ($\hat{\beta} = 0.29$, 95% CrI [0.03, 0.54], $P(\beta > 0) = 0.98$), confirming the presence of agreement attraction. Crucially, however, the interaction between Case and Attraction was not significant ($\hat{\beta} = -0.05$, 95% CrI [-0.32, 0.21], $P(\beta > 0) = 0.34$). This indicates that the magnitude of the attraction effect did not reliably differ between the Genitive and Nominative conditions.

Figure 3

Posterior probabilities of model coefficients.



Discussion

Summary of Findings

In this study, we investigated the effect of attractor case on agreement attraction in Turkish possessive constructions. We compared Genitive attractors, which are structural validity predictors of upcoming agreement, against Nominative (bare) attractors, which are the canonical form of subjects. Our primary finding is that both attractor types elicit comparable magnitudes of agreement attraction. Specifically, our

Bayesian analysis revealed a robust main effect of Attraction but no significant interaction between Case and Attraction. This null result persists despite the fact that Genitive attractors are much stronger probabilistic predictors of plural targets in the corpus.

Theoretical Implications

These results provide critical constraints for models of agreement processing. The lack of modulation by case challenge pure **probabilistic inference** accounts. If the parser were simply tracking the transitional probability $P(\text{Plural Target}|\text{Attractor})$ to repair the string, we would have expected significantly stronger attraction in the Genitive condition, where the attractor strongly signals an upcoming plural head. The uniform attraction effect suggests that the mechanism driving these errors is insensitive to fine-grained co-occurrence statistics of this type.

Instead, the results favor **structural controllerhood** accounts. Both Genitive possessors and Nominative subjects are licensed agreement controllers in Turkish grammar. The parser appears to retrieve candidates based on this abstract structural status—searching for “things that can control agreement”—rather than relying solely on surface feature overlap (which would favor Nominative) or predictive probability (which would favor Genitive). This supports a view of memory retrieval that is structurally constrained, targeting items with specific syntactic licensing properties.

LLM and Attention Mechanisms

Our findings also align with recent insights from computational models. Transformer-based Large Language Models (LLMs), such as BERT, rely on self-attention mechanisms to track dependencies. Previous work has shown that these models often attend to attractors in a way that correlates with human error patterns Linzen et al. (2016); Arehalli and Linzen (2020). For Turkish, the fact that BERT and similar models attend to both Genitive and Nominative attractors suggests that “attention” in these systems may be capturing the same notion of “structural controllerhood” that humans use. The model learns that both case forms are relevant for agreement computation, and thus distributes attention to them, leading to similar “errors” or interference effects. This provides a computational correlate to the psycholinguistic construct of controllerhood: high attention weights reflect the structural centrality of these case forms in the language’s dependency graph.

Semantic Constraints on Attraction

Finally, the behavior of the Nominative attractors speaks to the debate on the role of semantic integration in agreement. In English, compounds like *rat-eater* are standard, whereas *rats-eater* is typically disallowed, suggesting that plural features are not semantically interpreted inside such tight modifiers. Similarly, in our Turkish materials, interpreting the bare attractor as a subject would yield a semantically anomalous meaning (e.g., interpreting “The refugee’s lawyer” as “The refugee [who is the lawyer]”), akin to the “rat-eater” constraint where the modifier does not serve as the independent argument.

Some accounts, such as Patson and Husband (2016), argue that attraction involves a “good-enough” reanalysis or misinterpretation, where the parser briefly entertains the ungrammatical dependency as a valid, meaningful alternative (i.e., actually interpreting the attractor as the subject). If this were the case, the gross semantic implausibility of the Nominative attractor interpretation should block attraction, or at least significantly reduce it compared to the semantically more viable Genitive possessor.

However, our results show robust attraction for these semantically “abnormal” Nominative attractors. This aligns more closely with retrieval-based accounts that posit attraction as a fast, automatic error in the retrieval mechanism itself, rather than a full reanalysis or misinterpretation Schlueter et al. (2019); Parker (2018). The parser retrieves the attractor because it matches the structural/morphological retrieval cues (specifically, checking for a “subject-like” controller), but this retrieval does not necessarily imply that the

system has successfully integrated the attractor into a coherent, meaningful interpretation of the sentence. Thus, the persistence of attraction in the face of semantic implausibility supports the view that the agreement mechanism is encapsulated from fine-grained semantic plausibility checks during the initial retrieval phase.

Conclusion

We found similar rates of agreement attraction with two differently case-marked attractors in Turkish. Despite differences in their predictive validity for plural targets, both Genitive and Nominative attractors elicited comparable error rates. This supports a view of sentence processing where structural properties—specifically, the potential to control agreement—weigh heavier than surface probabilistic cues in driving memory retrieval and agreement computation.

References

- Arehalli, S., & Linzen, T. (2020). Neural networks as cognitive models of agreement attraction. *Proceedings of the Society for Computation in Linguistics*, 3(1), 335–345.
- Bhatia, A., & Dillon, B. (2011). Agreement attraction in comprehension: The role of feature match and structural controllerhood. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.
- Cartner, M. et al. (2025). Number attraction in turkish genitives. *Proceedings of the 38th Annual Conference on Human Sentence Processing*.
- Linzen, T., Dupoux, E., & Goldberg, Y. (2016). Assessing the ability of LSTMs to learn syntax-sensitive dependencies. *Transactions of the Association for Computational Linguistics*, 4, 521–535.
- Parker, D. (2018). Memory retrieval in sentence processing. *Topics in Cognitive Science*, 10(1), 117–133.
- Patson, N. D., & Husband, E. M. (2016). Misinterpretation in agreement attraction. *Quarterly Journal of Experimental Psychology*, 69(5), 950–971.
- Ryu, S.-Y., & Lewis, R. L. (2025). Agreement attraction in turkish: A comparison of human and model behavior. *Manuscript in Preparation*.
- Schlüter, Z., Parker, D., & Lau, E. (2019). Error-driven retrieval in agreement attraction. *Language, Cognition and Neuroscience*, 34(8), 1042–1056.
- Vasishth, S., & Lewis, R. L. (2008). Integration processes and expectation in argument resolution. *Journal of Memory and Language*, 59(4), 439–457.
- Wagers, M. W., Lau, E. F., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language*, 61(2), 206–237.

Appendix Appendices

Appendix A: Turkish Case System and Attractor Types

Maybe some stuff can be offloaded here.

To understand the structural difference between the attractors used in this study, it is necessary to distinguish between the **Genitive-Possessive** construction and the **Nominative (Bare)** construction in Turkish.

1. Genitive Attractors (Possessors) In Turkish, a definite possessive construction involves a Genitive case marker on the possessor and a Possessive marker on the head noun (the agreement target). * *Example*: Mülteci-nin avukat-ı (The refugee's lawyer) * *Structural Role*: The Genitive noun (Mülteci-nin) acts as a specifier and is structurally licensed as a possessor. In Turkish, possessors are strong controllers of agreement (e.g., in nominalized clauses).

2. Nominative Attractors (Modifiers) The “Nominative” attractors in this study consist of bare nouns appearing in a modifier position. While they carry no overt case morphology (Nominative), they do not function as true possessors in this context. * *Example*: Mülteci avukat-ı (The refugee lawyer / The lawyer of refugees) * *Structural Role*: The bare noun (Mülteci) acts as a modifier (similar to an English compound like *Refugee lawyer*). Unlike the Genitive possessor, it is not structurally poised to control agreement, making it a crucial test case for “Structural Controllerhood” theories.

Appendix B: Corpus Analysis Details

We utilized the **TrTenTen** corpus (approx. 4.9 billion words) to calculate the transitional probabilities reported in the study. The goal was to determine how predictive each attractor type is of an upcoming plural target.

We calculated the conditional probability $P(\text{Plural Target}|\text{Attractor})$ for both Genitive and Nominative attractors.

Attractor Case	Attractor Number	Probability of Following Plural Target
Genitive	Singular	11.4%
Genitive	Plural	24.0%
Nominative	Singular	17.0%
Nominative	Plural	5.0%

Note. Genitive plural attractors strongly predict a plural head, creating a high-expectancy environment for a plural verb. Nominative plural attractors do not share this predictive validity.

Appendix C: Large Language Model Analysis

To compare human processing with computational models, we analyzed **Turkish BERT** (uncased). We focused on the model’s self-attention mechanism to quantify “interference.”

Methodology 1. Head Selection: We first identified attention heads specialized for the subject-verb dependency (`nsubj`) using a standard probing task on a held-out dataset. **2. Attention Difference:** For every experimental item, we extracted the attention weight from the Verb to the Attractor (w_{att}) and from the Verb to the Target (w_{tgt}). **3. Scoring:** We calculated the difference ($w_{att} - w_{tgt}$). A positive score indicates the model is “distracted” by the attractor.

Results As shown in the main text (Figure 1), the model exhibited significant interference (positive attention difference) for both Genitive and Nominative attractors, mirroring the human behavioral results where structural status outweighed probabilistic cues.