

Sensitivity to within-experiment statistics: A case from Turkish agreement attraction

Within-experiment statistics in agreement attraction

Utku Turk^{a,*}

^a*University of Maryland, College Park, Linguistics, Marie Mount Hall, College Park, 20742*

Abstract

Surface level does not affect it, but within-experiment statistics effect the findings.

Keywords: keyword1, keyword2

1. Introduction

Speakers often rely on additional sources of information when processing sentences, including distributional expectations about forms and tasks, as well as the overall composition of an experimental session (e.g., the ratio of fillers to critical items). Recent work has demonstrated that such task-specific factors can substantially modulate reading and judgment behavior (Laurinavichyute and von der Malsburg, 2024; Arehalli and Wittenberg, 2021; Hammerly et al., 2019; Logačev and Vasishth, 2016). One line of research has used the agreement-attraction phenomenon to probe the heuristics that influence sentence processing. Agreement attraction refers to cases in which a verb erroneously agrees with a nearby noun rather than the true subject, giving rise to so-called grammaticality illusions in both production and comprehension (Bock and Miller, 1991; Pearlmutter et al., 1999).

- (1) * The key to the cabinets are rusty.

Agreement errors in sentences like (1) have been treated either as a failure of feature reconciliation or a failure of memory encoding. The former set of accounts explain these errors as a by-product of how number feature of a phrase is calculated in real-time (Bock and Miller, 1991; Eberhard et al., 2005; Hammerly et al., 2019). For example, Eberhard et al. (2005) argue that depending on

*Corresponding author

Email address: `utkuturk@umd.edu` (Utku Turk)

conceptual number, morphophonological number marking, or syntactic dependencies within a phrase, speakers assign a probabilistic number value to phrases. The errors arise when additional plurality features from different sources end up contributing to the final number representation of a phrase. On the other hand, the latter set of accounts claim that the initial representation is not erroneous, but speakers are sometimes unable to correctly retrieve the controller ([Wagers et al., 2009](#); [Dillon et al., 2013](#)). For example, [Wagers et al. \(2009\)](#) argue that the parser normally check the agreement relation by retrieving the relevant chunk in memory using the retrieval cues provided by the agreement probe. In sentences like (1), speakers occasionally retrieve the incorrect element due to the fact that neither nouns fully match the relevant cues.

However, both group of accounts generally are underspecified in terms of how meta-linguistic information should be integrated to the inter-sentential dependency mechanisms. Recently, a growing literature have been testing how different types of additional sources that are independent of the linguistic information affects these errors. Recent experiments show that even small changes in task expectations can alter attraction patterns. For example, [Laurinavichyute and von der Malsburg \(2024\)](#) found that varying the practice structure and task demands (reading vs. judgment) affected reading times at the verb in sentences as in (2). In a self-paced reading task, when participants answered a comprehension question after each trial, reading times at the verb ‘admires’ did not differ between (2a) and (2b). However, when participants were asked to judge grammaticality instead, they spent more time reading the verb ‘admires’ in (2b), suggesting that processing mechanisms can change depending on the expected task.

- (2) a. The singer that the actor openly admires apparently received broad international recognition.
- b. The singers that the actor openly admires apparently received broad international recognition.

A related set of findings came from [Hammerly et al. \(2019\)](#). They challenge long-standing assumption that the agreement errors only surfaced in ungrammatical sentences such as (1), but not in grammatical sentences as in (3). That is, while people are illusioned into thinking ungrammatical sentences are grammatical, they rarely misidentified grammatical sentences as ungrammatical. [Hammerly et al. \(2019\)](#) showed that similar effect surfaced in grammatical sentences, when participants’ a priori expectations about the experiment is altered. They manipulated the instructions and the number of ungrammatical in an experiment so that participants expected to see more ungrammatical sentences than grammatical sentences. With reduced bias towards grammaticality, they found that attraction effects surfaced both in ungrammatical sentences (1) and grammatical sentences (3).

- (3) The key to the cabinets is rusty.

- Context and task modulation
- Show that attraction strength varies across tasks and experimental context:
- Task type: self-paced reading with acceptability judgment or comprehension (Laurinavichyute & von der Malsburg 2023).
- Response expectations and bias: Hammerly et al. 2019 Turk 2022, participants’ error patterns change when they expect more ungrammatical items.
- Item composition: Arehalli — manipulating filler ratios mixes alters accuracy patterns.
- Conclude that participants adapt their behavior to the statistical environment of the experiment; attraction is not fixed but context-sensitive.

A related set of findings comes from work by [Slioussar \(2018\)](#), who showed that surface form can sometimes override abstract features in Russian. Exploiting the syncretism between singular genitives and nominative plurals—a pattern absent in plural genitives—she found that participants made more errors and showed faster reading times in sentences like (4b) than in (4a). She argued that, rather than accessing abstract case features, readers relied on surface-level cues that were easier to retrieve.

- (4) a. The singer that the actor openly
 The singer that the actor openly
 ‘The singer that the actor openly’
- b. The singers that the actor openly admires
 The singers that the actor openly admires
 ‘The singers that the actor openly admires’
- Form-based influences. in some languages, surface-form similarity can exacerbate attraction or even drive illusions on its own.
 - Cite [Slioussar \(2018\)](#) and related Russian work: case syncretism between nominative plurals and genitive singulars increased attraction or slowed reading, suggesting reliance on form-based heuristics rather than abstract case features.
 - Cite Chromy and checz data. tell it is not very consistent.

Building on these observations, we utilize Turkish as a testing ground to examine how surface-form overlap influences agreement processing and whether exposure to different kinds of distractors modulates attraction.

- Previous attraction findings in Turkish
- Prior work has reported typical attraction effects with genitive-marked nominal attractors, showing higher acceptance of ungrammatical plural-verb sentences.
- However, no work has directly tested whether verbal plural morphology can induce similar illusions, or how mixing different attractor types within an experiment affects the magnitude of attraction.

Turkish provides an especially informative case because both nominal and verbal plural markers are realized with the same morpheme, *-lAr*, yet only nominal plurals bear the syntactic features required for agreement. This allows us to ask whether participants rely on surface-form similarity or on abstract feature representations when evaluating agreement.

- Morphological properties
- Turkish marks number on both nouns and verbs using the identical plural morpheme *-lAr*.
- Only nominal plurals introduce number features that can agree with the verb; verbal *-lAr* expresses verbal agreement but is not a potential controller.
- Because of this homophony, Turkish allows form-overlap and feature-mismatch to be disentangled experimentally.

In our first experiment, we test whether plural marking on a verbal distractor—which is morphologically identical but syntactically irrelevant—can elicit attraction. In the second experiment, we combine these verbal distractor conditions with standard nominal attractor conditions to assess how their co-occurrence affects participants’ judgments. If attraction effects reflect flexible, context-sensitive processing, the inclusion of verbal distractors should dilute or eliminate the illusion typically observed with nominal attractors.

Together, these experiments extend previous findings on agreement attraction and task sensitivity in two key ways. First, they show that surface-level overlap—even when morphologically identical—does not by itself produce agreement attraction, indicating that participants rely on abstract morphosyntactic features rather than phonological forms. Second, they reveal that participants are not only influenced by the global structure of an experiment (such as the proportion of fillers or grammatical items) but also by the presence of other condition types within the same task. In other words, attraction effects are attenuated when competing, non-attracting conditions are included, suggesting that agreement processing is dynamically tuned to the statistical context of the experiment itself.

1.1. Experimental logic and predictions

- Goal 1: test whether purely form-based overlap (verbal *-lAr*) elicits attraction.
- Prediction: if attraction is driven by form, verbal plural distractors should yield higher “acceptable” rates for ungrammatical plurals.
- Alternative: if attraction depends on abstract features, no effect of verbal *-lAr* should appear.
- Goal 2: test whether the co-occurrence of different attractor types modulates attraction.
- Prediction: if participants adapt to the distribution of conditions, adding verbal distractors (which share the plural form but lack agreement features)

should attenuate or eliminate the nominal-attractor illusion.

- Summary: These experiments jointly test whether agreement attraction in Turkish reflects shallow form matching or feature-based computation that is sensitive to the statistical context of the task.

2. Experiment 1: Testing Form-Driven Processing

2.1. Participants

We recruited 80 undergraduate students to participate in the experiment in exchange for course credit. All participants were native Turkish speakers, with an average age of 21 (range: 18 – 31). The experiment was carried out following the principles of the Declaration of Helsinki and the regulations concerning research ethics at Bogazici University. All participants provided informed consent before their participation and their identities were completely anonymised.

2.2. Materials

We used 40 sets of sentences like (5), in which we manipulated (i) the number of the attractor and (ii) the number agreement on the verb. Both plural markings were marked with the suffix -ler/-lar, while the singular number and singular agreement were marked by its absence.

- (5) a. Tut-tuğ-u aşçı mutfak-ta sürekli zıpla-dı.
hire-NMLZ-POSS cook[NOM] kitchen-LOC non.stop jump-PST
'The cook they hired_{sg} jumped_{sg} in the kitchen non-stop.'
- b. *Tut-tuğ-u aşçı mutfak-ta sürekli zıpla-dı-lar.
hire-NMLZ-POSS cook[NOM] kitchen-LOC non.stop jump-PST-PL
'The cook they hired_{sg} jumped_{pl} in the kitchen non-stop.'
- c. Tut-tuk-lar-ı aşçı mutfak-ta sürekli zıpla-dı.
hire-NMLZ-PL-POSS cook[NOM] kitchen-LOC non.stop jump-PST
'The cook they hired_{pl} jumped_{sg} in the kitchen non-stop.'
- d. *Tut-tuk-lar-ı aşçı mutfak-ta sürekli
hire-NMLZ-PL-POSS cook[NOM] kitchen-LOC non.stop
zıpla-dı-lar.
jump-PST-PL
'The cook they hired_{pl} jumped_{pl} in the kitchen non-stop.'

All sentences were adapted by previous studies in Turkish agreement attraction (Lago et al., 2019; Türk and Logačev, 2020). Sentences started with a complex subject NP like 'tuttukları aşçı' 'the cook they hired,' in which the nominalized relative clause functioned as the attractor, and the head noun were bare. Because the plural marking on nominals is not optional and the head noun was singular, absent of -lar, in all conditions, sentences with plural verb agreement were ungrammatical. To inhibit participants from forming a task-related strategy in

which they deemed the sentence ungrammatical upon seeing a plural verb, half of our fillers included plural grammatical verbs, while the other half included singular ungrammatical verbs.

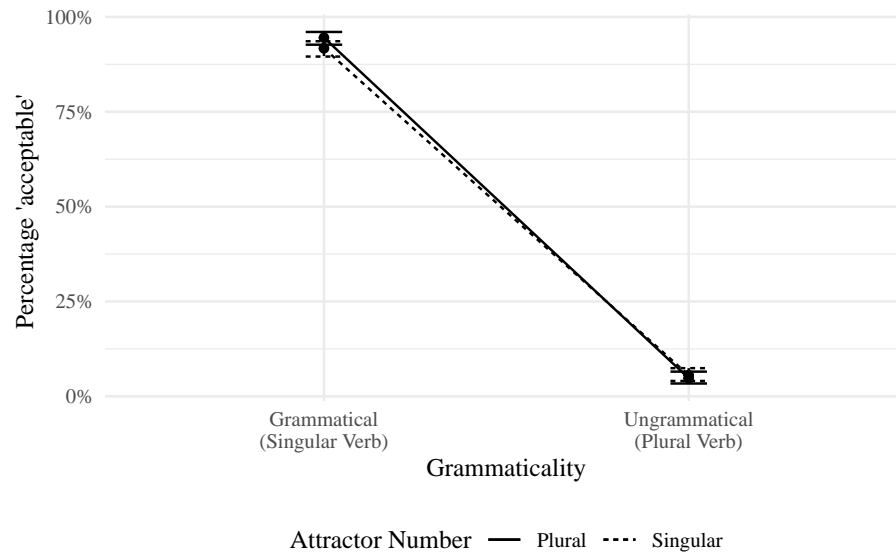
2.3. Procedures

The experiment was run online, using the web-based platform Ibex Farm (Drummond, 2013). Each experimental session took approximately 25 minutes to complete. Participants provided demographic information and gave informed consent to participate in the experiment. They then proceeded to read the instructions and were given nine practice trials before the experiment began.

Each trial began with a blank screen for 600 ms, followed by a word-by-word RSVP presentation of the sentence in the center of the screen, followed by a prompt to indicate their acceptability judgment. Sentences were presented word-by-word in the center of the screen in 30 pt font size, at a rate of 400 ms per word. Participants saw a blank screen for 100 ms between each word, and to see the next item, they needed to press the space key. Participants were asked to press the key P to indicate that a sentence is acceptable and Q to indicate that the sentence is unacceptable. They were instructed to provide judgments as quickly as possible. During the practice, but not during the experiment, a warning message in red font appeared if they did not respond within 5,000 ms.

Participants saw 40 experimental and 40 filler sentences. Experimental sentences were distributed among four different lists according to a Latin-square design. Every participant saw one version of the experiment with a specific list and one item per condition.

2.4. Analysis and Results



- Goal: determine if surface plural forms (verbal -lAr) elicit illusory agreement.
- Participants: 80 Turkish speakers (Boğaziçi undergraduates).
- Design: 2×2 (Grammaticality \times Attractor Number).
- Materials: relative-clause verbs as attractors; same surface morphology as nominal plurals.
- Procedure: speeded acceptability judgments, 1500 ms deadline.
- Analysis: Bayesian probit GLM (brms); random intercepts/slopes by subject/item.
- Results:
 - High filler accuracy ($> .9$).
 - No difference in ungrammatical sentences between plural vs singular attractors.
 - Posterior coefficients near 0; 95 % CIs within ROPE.
- Discussion:
 - No evidence for form-driven guessing.
 - Participants rely on abstract number features, not phonological similarity.

3. Experiment 2: Testing Within-Experiment Statistical Sensitivity

3.1. Participants

We recruited 118 undergraduate students to participate in the experiment in exchange for course credit. All participants were native Turkish speakers, with an average age of 20 (range: 18 – 32). The experiment was carried out following the principles of the Declaration of Helsinki and the regulations concerning research ethics at Bogazici University. All participants provided informed consent before their participation and their identities were completely anonymised.

3.2. Materials

The same materials were used with Exp1. We added items from TurkLogacev2024 as an additional condition.

3.3. Procedures

The same procedure with Experiment 1 was used.

3.4. Analysis and Results

- Goal: test whether attraction changes when both attractor types occur in one experiment.
- Participants: 95 Turkish speakers.
- Design: $2 \times 2 \times 2$ (Grammaticality \times Attractor Number \times Attractor Type [nominal vs verbal]).
- Procedure & analysis: same as Experiment 1.
- Results:
 - Attraction replicated for nominal attractors ($\Delta = 0.07$).

- Verbal attractors again showed null effect.
- Global decline in yes-responses relative to earlier studies → participants became more conservative.
- Discussion:
 - Exposure to verbal conditions reduced attraction magnitude overall.
 - Indicates participants adapt to statistical properties of the task.
 - Aligns with learning-based cue-weighting accounts (Haskell et al. 2010).

4. General Discussion

- Synthesis:
 - No evidence for surface-form matching; effects are feature-based.
 - Attraction magnitude changes with condition distribution → adaptive tuning.
- Interpretation:
 - Supports an adaptive parser sensitive to within-experiment statistics.
 - Challenges “shallow” or “good-enough” accounts that attribute attraction to phonological overlap.
- Broader implication:
 - Agreement processing is flexible and probabilistic; illusions arise from learned cue validity.
- Limitations:
 - Syntactic depth asymmetry (verbal attractors more embedded).
 - Need future designs equating structure (e.g., embedded-object attractors).
- Conclusion:
 - Turkish attraction effects arise from abstract feature retrieval not surface level shallow form-matching.
 - The evaluation of abstract features are modulated by distributional learning within the experiment.

References

- Arehalli, S., Wittenberg, E., 2021. Experimental filler design influences error correction rates in a word restoration paradigm. *Linguistics Vanguard* 7, 20200052. URL: <https://doi.org/10.1515/lingvan-2020-0052>, doi:<http://dx.doi.org/doi:10.1515/lingvan-2020-0052>.
- Bock, K., Miller, C.A., 1991. Broken agreement. *Cognitive Psychology* 23, 45–93. doi:[http://dx.doi.org/10.1016/0010-0285\(91\)90003-7](http://dx.doi.org/10.1016/0010-0285(91)90003-7).
- Dillon, B., Mishler, A., Sloggett, S., Phillips, C., 2013. Contrasting intrusion profiles for agreement and anaphora: Experimental and modeling evidence.

- Journal of Memory and Language 69, 85–103.
doi:<http://dx.doi.org/10.1016/j.jml.2013.04.003>.
- Drummond, A., 2013. *Ibex farm*. <https://spellout.net/ibexfarm>.
- Eberhard, K.M., Cutting, J.C., Bock, K., 2005. Making syntax of sense: Number agreement in sentence production. *Psychological review* 112, 531–559. doi:<http://dx.doi.org/10.1037/0033-295X.112.3.531>.
- Hammerly, C., Staub, A., Dillon, B., 2019. The grammaticality asymmetry in agreement attraction reflects response bias: Experimental and modeling evidence. *Cognitive Psychology* 110, 70–104.
doi:<http://dx.doi.org/10.1016/j.cogpsych.2019.01.001>.
- Lago, S., Gračanin-Yukse, M., Şafak, D.F., Demir, O., Kırkıcı, B., Felser, C., 2019. Straight from the horse’s mouth: Agreement attraction effects with Turkish possessors. *Linguistic Approaches to Bilingualism* 9, 398–426.
doi:<http://dx.doi.org/10.1075/lab.17019.lag>.
- Laurinavichyute, A., von der Malsburg, T., 2024. Agreement attraction in grammatical sentences and the role of the task. *Journal of Memory and Language* 137, 104525. URL:
<https://www.sciencedirect.com/science/article/pii/S0749596X24000287>,
doi:<http://dx.doi.org/https://doi.org/10.1016/j.jml.2024.104525>.
- Logačev, P., Vasishth, S., 2016. A multiple-channel model of task-dependent ambiguity resolution in sentence comprehension. *Cognitive Science* 40, 266–298.
- Pearlmutter, N.J., Garnsey, S.M., Bock, K., 1999. Agreement processes in sentence comprehension. *Journal of Memory and Language* 41, 427–456.
doi:<http://dx.doi.org/10.1006/jmla.1999.2653>.
- Slioussar, N., 2018. Forms and features: The role of syncretism in number agreement attraction. *Journal of Memory and Language* 101, 51–63.
doi:<http://dx.doi.org/10.1016/j.jml.2018.03.006>.
- Türk, U., Logačev, P., 2020. The role of shallow processing in agreement attraction. doi:<http://dx.doi.org/10.17605/osf.io/X9PV7>.
- Wagers, M.W., Lau, E.F., Phillips, C., 2009. Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language* 61, 206–237. doi:<http://dx.doi.org/10.1016/j.jml.2009.04.002>.