

Social meaning of vowel variants in Turkish

Muhammed İleri

University of Pennsylvania

ileri@sas.upenn.edu

Abstract

This paper investigates how listeners evaluate variation in the realization of Turkish /e/ in words such as *belki* ‘maybe’, where both [e] and [æ] are attested in the Istanbul variety despite the broader tendency for /e/ to lower in closed syllables with sonorant codas (e.g. *güzæ* ‘beautiful’). While this pattern is productive, it is not fully predictable: some words resist lowering (*elli* ‘fifty’), others show exceptional lowering (the aorist suffix *-mæz*), and some high-frequency items vary (‘maybe’: *belki* *bæ*lki) within and across speakers. We focus on such variable words and ask how listeners evaluate the naturalness of [e] and [æ] in formal and casual sentence frames, and whether these judgments are shaped by lexical expectations. In a matched-guise auditory naturalness judgment experiment, listeners of the Istanbul dialect rated [e] as less natural in casual frames, while [æ] showed no comparable overall difference across styles. A post-hoc norming study showed that stylistic penalties were concentrated in lexically atypical realizations: atypical [e] was penalized in casual frames, and atypical [æ] in formal frames. These findings support an account in which lexical expectations modulate perceptual salience and stylistic congruence.

1 INTRODUCTION

Variation is an essential property of human language, shaped by multiple factors both internal and external to the language system. On one hand, speakers produce variants of linguistic variables at different rates depending on language-internal factors (e.g., word-final position or word category) as well as external factors (e.g., the social groups they belong to or stylistic demands); on the other hand, listeners develop cognitive associations between linguistic variants and the contexts in which they typically occur (Podesva 2008, Campbell-Kibler 2009, Vaughn 2022a).

Listeners can associate different kinds of grammatical objects, ranging from sounds (Trudgill 1972, Phrasao et al. 2014, Vaughn and Kendall 2019, Regan 2022) and syntactic structures (Bender 2005), to semantic objects (Beltrama and Staum Casasanto 2022) with social meaning. One of the most studied sociolinguistic variables, the English /ING/, illustrates a case of (morpho-)phonological variation: the velar variant *-ing* is associated with higher socioeconomic status and intelligence of the speaker, as well as formal styles, whereas the non-velar variant *-in* indexes Southern speech and lower socioeconomic status, among other meanings (Trudgill 1972, Labov 2006, Campbell-Kibler 2009, 2010). Beyond phonological variation, (morpho-)syntactic variables like negative concord or copula deletion in AAE have also been shown to index dialect and register (Bender 2005).

*Many thanks to Meredith Tamminga, David Embick, Marlyse Baptista, Wesley M. Lincoln, my classmates in the second-year seminar, and the audiences at NWAV 53 for insightful and helpful discussions. I am also grateful to the faculty in the Department of Linguistics at Boğaziçi University for sharing the experiment with their students for extra course credit. Last but not least, I thank my Turkish consultants. All errors are my own.

Much of the theorizing in sociolinguistics, and in particular, indexicality, is primarily driven by work on European languages, particularly varieties of English (Labov 1963, Eckert 2008, Campbell-Kibler 2009), even though work on non-European languages is not unattested (e.g., Ngom 2003, Zhang 2005, Gafter and Horesh 2020, Okamoto and Morimoto 2023). In an effort to expand the empirical scope of the field, and validate the generality of the findings, we examine the social meaning of a vowel variation in Turkish: the variation between [e] and [æ]. Although these vowels are mostly in complementary distribution, the choice between them is variable in a specific set of lexical items (e.g., *belki* 'maybe') in Standard Turkish spoken in and around Istanbul, though the inventory of variable words varies across dialects (Göksel and Kerslake 2005, Dadan et al. 2025). The mid variant [e] is generally preferred by professional speakers, such as broadcasters, and among those aspiring to speak "proper" Turkish (e.g., Demirel 2021). Together with native speaker intuitions, this contrast suggests a register-based distinction: [æ] is associated with intimacy and casualness, whereas [e] indexes social distance and formality.

In this paper, we use a novel paradigm combining the matched guise technique with stylistic coherence to investigate the social meaning of [e] and [æ] in Turkish. We present Turkish listeners with recorded sentences that consist of a carrier, which is manipulated for formality, and a target word, which is manipulated for the vowel variant. We ask them to judge the naturalness of the sentences they hear. Our primary hypothesis was that if [e] and [æ] carry stylistic associations, listeners should generally penalize stylistically incongruent pairings (e.g., a formal vowel in a casual frame).

Our results confirm the existence of this stylistic penalty, but suggest that it may be structurally constrained. Specifically, our results indicate that the penalty for stylistic incongruence appears to be sensitive to lexical typicality and phonological markedness. In our data, incongruence was penalized primarily when the variant was lexically dispreferred, with hyper-formal mismatches (using [e] where the language prefers [æ]) receiving the harshest judgments. These findings suggest a hierarchical model of sociolinguistic coherence, where internal lexical and phonological validity may need to be satisfied before stylistic congruency is evaluated.

The remainder of this paper is organized as follows. Section 2 provides background on social meaning and stylistic coherence, followed by a description of the Turkish vowel variation under investigation. Section 3 outlines the experimental methodology, detailing our novel approach combining matched guise technique with stylistic coherence manipulations, along with stimuli creation. Section 4 presents the results of the study, including the post-hoc analysis of lexical constraints. In Section 5, we discuss the implications of these findings for theories of sociolinguistic perception and salience. Section 6 concludes.

2 BACKGROUND

2.1 COVARIATION AND COHERENCE

Linguistic variation does not occur in isolation; instead, variants pattern in constellations of covarying features that form a coherent unit (Pharao et al. 2014, Montgomery and Moore 2018, Buson et al. 2018, Tamminga and Wade 2022). When listeners perceive speech, they encounter multiple features simultaneously across phonological, grammatical, and prosodic levels, and their inferences about social meaning depend on the combined contribution of this constellation rather than on any single feature in isolation (Vaughn 2022a). Speakers and listeners appear to organize this variation around high-level coherent representations—stylistic schemata—that link variants from different linguistic levels to shared social meanings (Podesva 2008, Buson et al. 2018, Vaughn and Kendall 2019).

Evidence for this cognitive organization comes from production studies showing that speakers unconsciously modulate multiple variables in tandem. When non-Southern speakers were asked to produce the -in variant of /ING/, a feature strongly associated with Southern and casual American English, they simultaneously produced shorter /aɪ/ glides, closer mid-front vowel spacing, and longer lax vowels, all aligning with the Southern/casual style, despite receiving no explicit instruction to alter anything but /ING/ (Vaughn and Kendall 2019). This pattern suggests that producing a variant activates a stylistic template that includes multiple features from different linguistic levels. More evidence comes from perception: Buson et al. (2018) find that when listeners hear sociolinguistically mixed utterances (for instance, formal lexical and syntactic material combined with non-standard phonology), they tend to reconstruct them by replacing the inconsistent variant with one that matches the rest of the utterance, a phenomenon they describe as *sociolinguistic restoration*. This restoration effect indicates that high-order coherent representations influence how utterances are reconstructed in memory, implying structured knowledge that links variants across grammatical, phonological, and discourse levels. Taken together, these findings support the idea that producing or hearing one variant activates expectations about other variants, and that these expectations are organized around broader stylistic patterns rather than around isolated variables.

2.2 SALIENCE OF COVARYING VARIANTS

While stylistic coherence describes how variants cluster meaningfully, salience determines which features in a cluster actually drive listener perception and social inference (Levon and Fox 2014, Pharao et al. 2014, Montgomery and Moore 2018). Not all features contribute equally to social meaning: some are indexically resilient and so strongly associated with particular personae or varieties that they can override or mask the meanings of other co-occurring features in a given context (Pharao et al. 2014). Levon and Fox (2014) further demonstrate that this salience is not a static property of a variable but is modulated by “attitude strength”; they show that a variable may be robustly stratified in production (like ING in Britain) yet fail to trigger the sociolinguistic monitor (Labov et al. 2011) if the associated social attitudes lack sufficient centrality and commitment for the listener group.

This asymmetric weighting is evident in multiple domains. Pharao et al. (2014) find in Copenhagen Danish that a fronted /s/ is strongly associated with femininity and gayness in “modern Copenhagen” speech, but the same segment loses this interpretation in “street” speech, where prosodic and segmental features associated with an ethnic, tough, streetwise masculinity dominate listeners’ evaluations. Similarly, Vaughn (2022a) shows that listeners’ sensitivity to the internal grammatical constraints of ING disappears when stylistic incongruence is introduced through splicing, suggesting that highly salient stylistic context cues can suppress the social impact of otherwise informative sociolinguistic variables.

Buson et al. (2018) provide further evidence that feature salience modulates how listeners process incongruence. While their repetition task showed that listeners often cognitively “repair” mismatched variants to restore coherence, their judgment task revealed an interesting asymmetry in perception. Listeners were able to detect a non-standard variant in a formal sentence (rating it as less formal), yet they failed to register a standard variant in an informal sentence, rating it comparably to a fully informal utterance. This suggests that marked (non-standard) variants are perceptually salient enough to disrupt a formal frame, whereas unmarked (standard) variants may be overshadowed by a dominant informal style.

2.3 CONTEXT-DEPENDENCE OF SOCIAL MEANING

A critical implication of salience and covariation is that the social meaning of a single variant is not fixed but depends on which other features co-occur, which stylistic template is activated, and what aspects of the signal listeners attend to (Pharao et al. 2014, Montgomery and Moore 2018, Vaughn 2022a). The indexical field of a variant—the set of social personae, stances, and categories it can evoke—is therefore context-dependent rather than invariant (Eckert 2008). At the same time, assessments of stylistic coherence or incoherence are not absolute properties of varieties; they are relative to the social and temporal scale at which patterns are examined (Tamminga and Wade 2022).

Tamminga and Wade (2022) observe that the same variables can show robust covariation at one scale but appear independent at another. Apparent incoherence at a broad community level may mask coherent clustering within narrower subgroups defined by generation, ethnicity, or network, just as broad apparent coherence can decompose into distinct, internally coherent sub-varieties when inspected more locally. They argue that, as a result, whether a set of variants is described as “coherent” or “incoherent” is only meaningful with respect to a specified unit (individuals, subgroups, whole communities) and scope (narrow vs. broad social and temporal frames).

2.4 A HIERARCHICAL MODEL OF FEATURE INTEGRATION IN PERCEPTION

Taken together, these findings support a hierarchical model of how listeners integrate multiple variants to construct social meaning. First, listeners engage in salience-based filtering, attending initially to the most perceptually and ideologically prominent features which establish an overall stylistic frame (Levon and Fox 2014, Montgomery and Moore 2018). Second, the constellation of these high-salience cues activates one or more sociolinguistic schemata that encode expectations about what additional, lower-salience features are likely to be present (Buson et al. 2018, Vaughn and Kendall 2019). Violations of these expectations may be assimilated and effectively “repaired” by the schema, or may themselves become salient if they conflict with entrenched stereotypes (Buson et al. 2018).

Third, when variants from multiple linguistic levels co-occur, the social meanings associated with some levels can override or suppress meanings associated with others, depending on their relative salience, frequency, and accessibility under current attentional demands (Levon and Fox 2014, Vaughn 2022b). Perceived meaning thus emerges not from an additive sum of independent cues but from a gestalt in which certain features are weighted much more heavily than others (Montgomery and Moore 2018). This weighting helps explain why stylistic coherence remains a powerful organizer of perception even in the face of local mismatches: listeners rely on highly salient features to activate stylistic templates and then interpret both present and absent cues in light of those templates (Buson et al. 2018, Vaughn and Kendall 2019). Variants that occur together in enregistered styles (Agha 2003) become cognitively associated (Eckert 2008), such that the presence of one makes others more expected (Vaughn 2022a).

2.5 THE VOWEL VARIATION IN TURKISH

The variation between the mid-vowel [e] and the low-vowel [æ] in Modern Standard Turkish (the Istanbul dialect) results from a complex historical trajectory. Evidence suggests that Old Turkic (8th–13th c.) likely maintained a phonemic contrast between /e/ and /æ/ (Erdal, 2004: 50–52); however, this distinction was largely lost in later periods through a merger towards [æ]. Crucially, this merger was

incomplete: the prestige Istanbul dialect retained conservative [e] variants in specific lexical items (Johanson, 2021). This historical persistence of [e] in the standard variety provides the foundation for both the productive allophonic patterns and the opaque lexical distribution observed today.

In Standard Turkish, the alternation is largely governed by a productive allophonic rule whereby /e/ lowers to [æ] when it appears before a tautosyllabic sonorant (coda /m, n, r, l/), surfacing as [e] elsewhere (Göksel and Kerslake 2005, Gopal 2018, Dadan et al. 2025). This rule is highly productive and sensitive to syllable boundaries. For instance, the first person pronoun *ben* ‘I’ is realized with the open vowel [bæn], but when resyllabified by a suffix (e.g., [bæ.ni] ‘me’), the vowel is realized as [e], demonstrating the rule’s productivity.

At the same time, the system is not fully transparent. There are words in the Standard (Istanbul) variety that do not undergo lowering at all, categorically retaining [e] before a sonorant. A clear case is *benzin* ‘gasoline’ or *elli* ‘fifty’, which are pronounced [ben.zin] and [el.li] in the Istanbul dialect; the lowered forms *[bæn.zin] and *[æll.i] are typically judged as non-standard or simply wrong in this variety. In some regional dialects, however, such as the Samsun dialect, the lowering rule applies without lexical exceptions, resulting in [bæn.zin] and [æll.i] being the preferred forms, whereas [benzin] and [el.li] are dispreferred (though likely acceptable due to exposure to the Standard dialect). Thus, what counts as “regular” or “exceptional” depends on the regional dialect.

Between such categorical cases (like *benzin* in Istanbul vs. Samsun) and fully regular lowering items (like *ben* → [bæn]), there is also a set of words that are variable within the Standard dialect. Items such as *kendi* ‘self’, *belki* ‘maybe’, or *mühendis* ‘engineer’ are produced with both [e] and [æ] across speakers, and even within the same speaker across contexts. In these cases, neither variant sounds outright incorrect to listeners, but they differ subtly in how they pattern with style and register.

Prescriptive norms are informative about the social meaning of these variants. Pronunciation guides aimed at broadcasters and television professionals—such as the RTÜK pronunciation guide (Demirel 2021)—explicitly promote conservative [e] realizations in exceptional words that do not obey the lowering rule. For variable items like *mühendis* or *belki*—both of which are attested in everyday Istanbul speech with [e] and [æ]—professionals expected to sound “proper” are encouraged to favor the mid [e] variant (e.g., Günhan 2023). Thus, prescriptivist sources tend to associate [e] with correctness and carefulness.

Taken together, these usage facts suggest that the [e] ~ [æ] variation carries social meaning. The [e] variant is a plausible index of prestige, educatedness, and formality, while [æ] is more closely tied to casualness and to regional or non-standard varieties. At the same time, this indexical potential is constrained by lexical identity. Using [e] in a word that categorically demands [æ] in the standard (e.g., pronouncing *ben* as *[ben]) does not come across as formal—it simply sounds unnatural. These observations suggest that both variants must be acceptable enough for the listener to interpret the choice as stylistic rather than as an error.

The present experiment builds directly on this observation. We focus on a set of target words that (a) follow the /e/ + sonorant pattern, (b) are acceptable with both [e] and [æ] in the Istanbul dialect, and (c) span different sonorant codas and frequency ranges.¹ These are precisely the items where one can plausibly ask whether the choice of [e] vs. [æ] is heard as a stylistic choice rather than a categorical violation.

Against this backdrop, the goal of the study is to test whether listeners are sensitive to a stylistic

¹These frequencies combine both variants. Turkish orthography does not distinguish the two realizations (both written as ‘e’); therefore, corpus counts cannot estimate separate variant frequencies.

contrast between [e] and [æ] in these variable words, and how this contrast interacts with the formality of the surrounding sentence frame. The following section describes the experimental design used to probe this interaction.

3 METHODS

We employed a matched-guise perception experiment to investigate how listeners evaluate the naturalness of sentences when a socially meaningful vowel realization either aligns with or conflicts with the style of the surrounding sentence frame. We crossed two factors in a fully within-subjects 2×2 design. The first factor was the realization of the target vowel in a critical word, with two levels, [e] and [æ]. The second factor was the stylistic character of the sentence frame in which the target word appeared, with a formal frame condition and a casual frame condition. Each participant heard sentences exemplifying all four combinations of vowel realization and frame style (formal frame with [e], formal frame with [æ], casual frame with [e], and casual frame with [æ]).

On every trial in the main experiment, participants heard a single sentence and then rated how natural the utterance sounded using a continuous slider scale ranging from 0 to 99. The working assumption is that sentences in which the stylistic meaning associated with the target vowel realization mismatches the style of the frame will be perceived as less natural than sentences in which vowel and frame are stylistically coherent, in line with work on stylistic coherence (see [Vaughn 2022a](#)).

Before running the experiment, two norming studies were conducted with independent groups of listeners in order to validate the sentence materials. In a naturalness norming study ($N = 20$), listeners rated sentences in which both the formal and casual versions of each frame were paired with a neutral target word, that is, without any one of the critical vowel variants. This study established that both frame variants were acceptable and sounded natural in isolation. In a separate formality norming study ($N = 58$), listeners judged the perceived formality of the same sentence frames, again with neutral target words, to confirm that the formal and casual versions differed reliably in perceived formality. Together, these norming results ensured that the frame manipulation produced a robust formality contrast, and that any naturalness differences in the main experiment could be attributed to the interaction between target vowel realization and frame style rather than to idiosyncratic properties of the materials.

3.1 PARTICIPANTS

112 participants took part in the main perception experiment. Participants were primarily undergraduate students at Boğaziçi University (Istanbul, Türkiye) who received course credit as compensation. Additional participants were recruited via snowball sampling and did not receive compensation. All self-reported Turkish as their native language, with the exception of two participants who reported Arabic and Kabardian (North Caucasian) as their first languages but identified as fluent in Turkish; these participants were included in the final analysis. The norming studies recruited independent samples of listeners: 20 participants for the naturalness norming and 58 participants for the formality norming, recruited under the same criteria as the main experiment.

3.2 MATERIALS

3.2.1 Sentence frames and target words

The experimental stimuli consisted of 40 sentences (10 target words $\times$ 4 conditions), each containing a single critical target word with the relevant vowel ([e] or [æ]). For each sentence, we created two stylistic versions of the surrounding frame, one formal and one casual, differing only in the subject agreement marker on the verb. The casual sentence had the 2SG agreement marker on the verb, whereas the formal sentence had the 2PL agreement marker,² following the conventional association of 2PL with formality in Turkish.

Target words were selected from different frequency ranges in the TSCorpus v2³ (Sak et al. 2008), and from syllables with different sonorant codas (m, n, l, r), since [e] $\sim$ [æ] variation in Standard Turkish occurs before sonorants. Across the ten target items, normalized frequencies ranged from approximately 7 to 1,100 occurrences per million words (mean $\approx$ 179), ensuring the inclusion of both low- and high-frequency items.

In the norming studies, the same formal and casual frame variants were presented with a neutral version of the target word that did not contain either of the critical vowel realizations, in order to assess the naturalness and perceived formality of the frames independently of the vowel manipulation.

3.2.2 Recording and stimulus preparation

The sentences were recorded by a 29-year-old female native speaker of the standard dialect in a sound-attenuated booth, using a high-quality microphone and Audacity. The speaker was instructed to read each sentence at a natural speaking rate and to produce both the [e] and [æ] versions of the target word as naturally as possible, while keeping the rest of the sentence constant.

Each sentence was recorded three times, and the most natural-sounding token was selected by the author, a native speaker of the standard dialect. Sound files were segmented in Praat (Boersma 2001), and all stimuli were normalized to a peak amplitude of 70 dB SPL to ensure comparable loudness across items.

In the main experiment, the 40 critical sentences were not constructed by splicing target words into the exact sound files used in the norming studies. Instead, for each condition we recorded full sentences separately, in the same session, using the same speaker, recording setup, and scripted materials as in norming. For every target word and frame type, the speaker produced a complete sentence with either the [e] or [æ] realization in place, and the most natural-sounding token was selected. This approach prioritizes the ecological validity and stylistic coherence of the full sentence, which is supported by recent work: Vaughn (2022a) found that listeners' sensitivity to subtle linguistic constraints was most robust in naturally produced stimuli, suggesting that natural productions preserve co-varying cues that can be disrupted by splicing. Furthermore, Tamminga (2017) demonstrated that the social evaluation of variation remains consistent across read and conversational speech styles, validating the use of read speech in this context. To ensure the stimuli were valid, all selected recordings were checked by the author, a native speaker of the standard dialect, who confirmed that the frames were natural.

In addition to the 40 critical sentences, we included 42 filler sentences of comparable length. The fillers varied in formality and naturalness so that participants would be encouraged to use the full range

²Göksel and Kerslake (2005: 231) note that 2PL is used when addressing someone on formal terms or of higher status.

³tscorpus.com

of the rating scale. One of the filler sentences was used as a practice item at the beginning of the experiment, allowing participants to familiarize themselves with the trial structure.

3.3 PROCEDURE

The experiment was designed and hosted online on the PCIBex platform⁴ (Zehr and Schwarz 2018). Participants were instructed to complete the task in a quiet environment using a computer and headphones.

On each trial, participants clicked a ► button to listen to the recorded sentence and were allowed to listen up to three times. They then rated the sentence on a 0–99 slider, with 0 corresponding to a highly unnatural (*kulak tırmalayan*; lit. “ear-scratching”) utterance and 99 corresponding to a completely natural (*gayet doğal*; lit. “very natural”) utterance.

The experimental session began with a single practice item to familiarize participants with the interface and task structure. Following the practice trial, the main experiment was divided into two blocks. Each of the 10 target words appeared exactly once in each block (once in one condition in Block 1, and once in a different condition in Block 2), ensuring that participants never heard the same target word in close succession.

Within each block, trials were presented in a pseudo-randomized order designed to space out critical items. In the first block (32 items total), the sequence followed a pattern of two fillers followed by one critical item, ending with two fillers. Between the two blocks, participants completed a brief distractor task consisting of three simple arithmetic problems (addition, subtraction, and multiplication) to minimize recency effects before the second block began. The second block (30 items total) followed a similar pseudo-random structure (sequences of two fillers and one critical item). Across the entire experiment, this yielded a total of 62 items (20 critical trials and 42 fillers), with critical items distributed evenly throughout the session to prevent fatigue effects or strategy formation. At the end, participants completed a demographic survey collecting information about language background, gender, and feedback. The entire session lasted approximately 15 minutes.

4 RESULTS

No participants were excluded based on the pre-established exclusion criteria (e.g., participants with zero standard deviation in their ratings). Prior to analysis, naturalness ratings were centered and scaled (i.e., z-scored) within each participant to control for individual differences in scale usage. The final dataset for analysis consisted of 2,240 observations from 112 participants.

Figure 1 visualizes the naturalness ratings across the four conditions. The raw distributions (left panel) show that ratings are generally high, clustering toward the top of the scale, yet they still align with the predictions of stylistic coherence. In Formal frames, the [e] variant receives higher mean ratings, while in Casual frames, the [æ] variant is slightly preferred. The interaction plot of observed mean z-scores (right panel) clarifies this crossover: the tendency to favor [e] in formal contexts disappears or reverses direction in casual contexts.

To test the statistical significance of these patterns, we fitted a linear mixed-effects model using the `lmerTest` package (Kuznetsova et al. 2017) in R. The dependent variable was the z-scored naturalness rating. The model included fixed effects for Vowel ([e] vs. [æ]), Formality (Formal vs. Casual), and their

⁴farm.pcibex.net

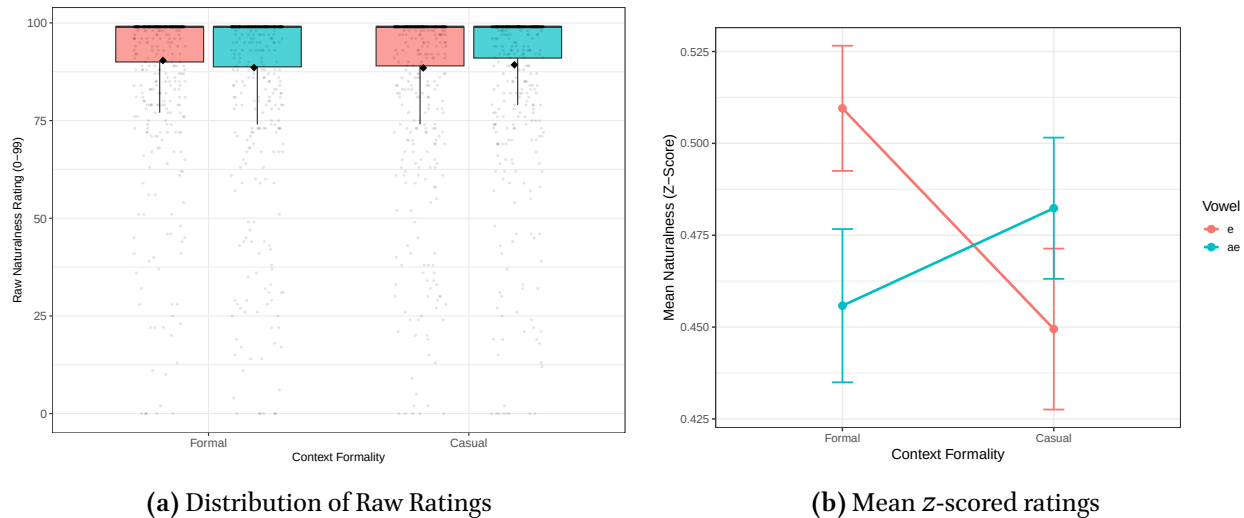


Figure 1: Naturalness ratings by Context Formality and Vowel. (a) Boxplots of raw ratings (0–99) show the overall distribution; black diamond markers indicate mean ratings for each condition. (b) Mean z-scored ratings (with standard error bars) highlight the significant crossover interaction, where the tendency to favor [e] in formal contexts disappears or reverses direction in casual contexts.

interaction, as well as a fixed effect for Block (Block 1 vs. Block 2). Treatment contrasts were used with $\text{VOWEL}=[e]$ and $\text{FORMALITY}=\text{Formal}$ as the reference levels.

We followed a stepwise reduction procedure for the random effects structure (Barr et al. 2013). We began with a maximal structure including all motivated random slopes. When singular fit issues arose, we systematically removed the random slope terms that accounted for the least variance until convergence was achieved. The final model included random intercepts for SUBJECT and TARGET WORD. For random slopes, the model retained the effects of VOWEL and BLOCK by SUBJECT, and the effects of VOWEL and FORMALITY by TARGET WORD. The final model formula was:

$$\text{Naturalness}_z \sim \text{Vowel} \times \text{Formality} + \text{Block} + (1 + \text{Vowel} + \text{Block} \mid \text{Subject}) + (1 + \text{Vowel} + \text{Formality} \mid \text{Target_Word})$$

The fitted model was run through the `check_model` function in the `easystats` package (Lüdtke et al. 2022). Visual inspection of diagnostic plots revealed a negatively skewed distribution of ratings with a ceiling effect. Examination of the residuals showed some deviation from normality and homoscedasticity. However, linear mixed models are robust to violations of distributional assumptions when sample sizes are sufficient (Schielzeth et al. 2020); given that $N = 112$ for our study, the model was kept. Reported p -values were obtained using the Satterthwaite method.

The results of the mixed-effects model on the full sample ($N = 112$) are summarized in Table 1. The main effect of FORMALITY was not statistically significant ($\beta = -0.060$, $p = .113$), nor was the main effect of VOWEL at the reference level ($\beta = -0.056$, $p = .360$). The effect of BLOCK was also non-significant ($\beta = -0.029$, $p = .203$).

However, the interaction between VOWEL and FORMALITY was statistically significant ($\beta = 0.085$, $p = .009$). As illustrated in Figure 1b, this interaction manifests as a crossover: [e] received numerically higher ratings than [æ] in the formal condition, whereas this pattern inverted in the casual condition.

To determine the source of this effect, we conducted post-hoc pairwise comparisons (Tukey adjusted). Despite the significant global interaction, none of the simple pairwise contrasts reached statistical significance in this aggregate model (all $p > .10$). For instance, the rating of [e] did not drop significantly between Formal and Casual contexts ($p = .114$), nor was there a significant preference for either vowel within the Casual context ($p = .63$).

This indicates that the interaction is driven by a change in the relative position of the vowels rather than large absolute changes. While ratings for individual vowels remained stable, the gap between them narrowed enough to eliminate the preference for [e] in the casual context.

However, as detailed in Section 2.5, the distribution of these vowels varies by dialect and word. A pooled analysis may therefore mask more specific patterns. To investigate this, we next model the roles of dialect background and lexical frequency.

Table 1: Fixed effects from the linear mixed-effects model predicting z-scored naturalness ratings (Full Sample, $N = 112$). Reference levels: Vowel=[e], Formality=Formal, Block=Bl.

Effect	Estimate	Std. Error	df	<i>t</i> value	<i>p</i>
(Intercept)	0.526	0.048	13.44	10.91	< .001 ***
Vowel [æ]	−0.056	0.059	10.68	−0.96	.360
Formality [Casual]	−0.060	0.036	14.08	−1.69	.113
Block [B2]	−0.029	0.023	109.19	−1.28	.203
Vowel [æ] × Formality [Casual]	0.085	0.032	1904.68	2.63	.009 **

Significance codes: *** $p < .001$, ** $p < .01$, * $p < .05$, . $p < .1$

4.1 DIALECTAL VARIATION

As noted in Section 2.5, regional dialects differ in their vowel distributions. Our sample ($N = 112$) included a subset of speakers ($N = 24$) identifying with non-standard varieties (e.g., Black Sea, Aegean, Inner Anatolian). Visual inspection of the data (Figure 2) suggested that this heterogeneity might indeed be a confounding factor: speakers of other dialects may evaluate target words based on distinct regional phonologies, potentially rating standard realizations as unnatural or vice versa.

To test this, we re-analyzed the data using only the speakers who identified as Standard Istanbul Turkish ($N = 88$). The model structure remained identical to the full analysis, including random intercepts for Subject and Target Word, and random slopes for Vowel and Block by Subject, and Vowel by Target Word. This targeted analysis revealed a robust Vowel × Formality interaction ($\beta = 0.10$, $p = .008$).

Table 2: Fixed effects from the linear mixed-effects model fit to data from Standard Turkish speakers ($N = 88$). Reference levels: Vowel=[e], Formality=Formal, Block=Bl.

Effect	Estimate	Std. Error	df	<i>t</i> value	<i>p</i>
(Intercept)	0.542	0.052	15.6	10.34	< .001 ***
Vowel [æ]	−0.082	0.064	10.9	−1.29	.225
Formality [Casual]	−0.060	0.027	1560.8	−2.23	.026 *
Block [B2]	−0.055	0.027	86.4	−2.04	.045 *
Vowel [æ] × Formality [Casual]	0.100	0.038	1497.3	2.68	.008 **

Significance codes: *** $p < .001$, ** $p < .01$, * $p < .05$, . $p < .1$

Crucially, post-hoc tests clarified the nature of this interaction. As predicted, Standard Turkish

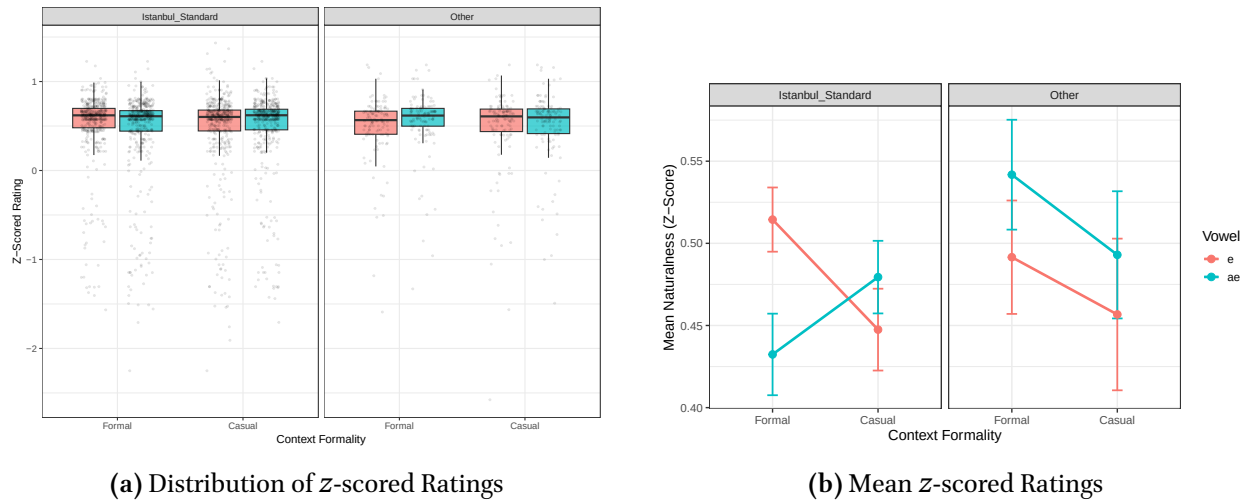


Figure 2: Comparison of Standard Istanbul ($N = 88$) vs. Other Dialects ($N = 24$). The Standard group (left) shows a clear penalty for [e] in casual contexts, whereas the heterogeneous Other group (right) shows no consistent pattern. Ratings are z-scored to correct for ceiling effects in the raw data.

speakers significantly penalized the use of [e] in casual contexts compared to formal ones ($\beta = -0.060$, $p = .026$). In contrast, while the ratings for [æ] numerically increased in the casual context, this increase was not statistically significant ($p = .13$). Furthermore, the difference between [e] and [æ] was not significant in either the casual ($p = .78$) or formal ($p = .225$) contexts.

These results suggest that the crossover effect visualized in Figure 2 is driven primarily by penalizing [e] in casual frames, rather than a significant boost for [æ]. As illustrated in the figure, the Standard Turkish group (left panels) demonstrates this clear stylistic penalty on the [e] variant. In contrast, the pattern in the Other dialect group (right panels) does not align with our predictions, suggesting that the global interaction effect observed in the full sample was driven primarily by the Standard Turkish speakers.

4.2 THE ROLE OF LEXICAL PREFERENCE

While the stylistic constraint was robust among Standard Turkish speakers, the random effects structure of the previous model indicated that the effect of vowel realization was not uniform across items. Specifically, the random slope for Vowel by Target Word had a standard deviation of 0.18 (Variance = 0.033), suggesting that some words showed a much stronger preference for [e] or [æ] than others. To investigate whether this variation was driven by word-specific phonological norms, we conducted a post-hoc norming survey with 8 native speakers of the Standard dialect. Words were categorized based on speaker agreement: words were classified as having a specific LEXICAL PREFERENCE if at least 5 of 8 speakers agreed on the vowel. Words for which the speakers were evenly divided (4–4) between the two vowels were classified as NEUTRAL.

We coded the trials based on LEXICAL TYPICALITY: whether the stimulus vowel was Typical, Atypical, or Neutral for that word.

To statistically test whether lexical preference moderated the stylistic constraint, we fitted a linear mixed-effects model restricted to the Standard Istanbul group ($N = 88$). Neutral words were excluded from the model to focus on the clear contrast between Typical and Atypical items. The model included

fixed effects for Vowel, Formality, and Lexical Typicality (Typical vs. Atypical), along with their three-way interaction.

The results, summarized in Table 3, revealed a highly significant three-way interaction between Vowel, Formality, and Lexical Typicality ($\beta = 0.37$, $p < .001$). As visualized in Figure 3, which displays the pattern across all three categories for completeness, the effect is gradient but driven primarily by the Atypical condition.

Table 3: Fixed effects from the lexical interaction model (Standard Istanbul speakers, Neutral words excluded). Reference levels: Vowel=[e], Formality=Formal, Lexical=Typical.

Effect	Estimate	Std. Error	df	<i>t</i> value	<i>p</i>
(Intercept)	0.590	0.066	9.30	9.00	< .001 ***
Vowel [æ]	−0.064	0.117	9.16	−0.55	.598
Formality [Casual]	0.045	0.047	1127.0	0.96	.335
Lexical [Atypical]	−0.144	0.093	6.97	−1.55	.166
Block [B2]	−0.060	0.025	1122.0	−2.41	.016 *
Vowel [æ] × Formality [Casual]	−0.093	0.071	1140.0	−1.31	.190
Vowel [æ] × Lexical [Atypical]	−0.001	0.184	5.82	−0.01	.998
Formality [Casual] × Lexical [Atypical]	−0.227	0.072	1125.0	−3.16	.002 **
Vowel [æ] × Formality [Cas] × Lexical [Atyp]	0.370	0.100	1081.0	3.71	< .001 ***

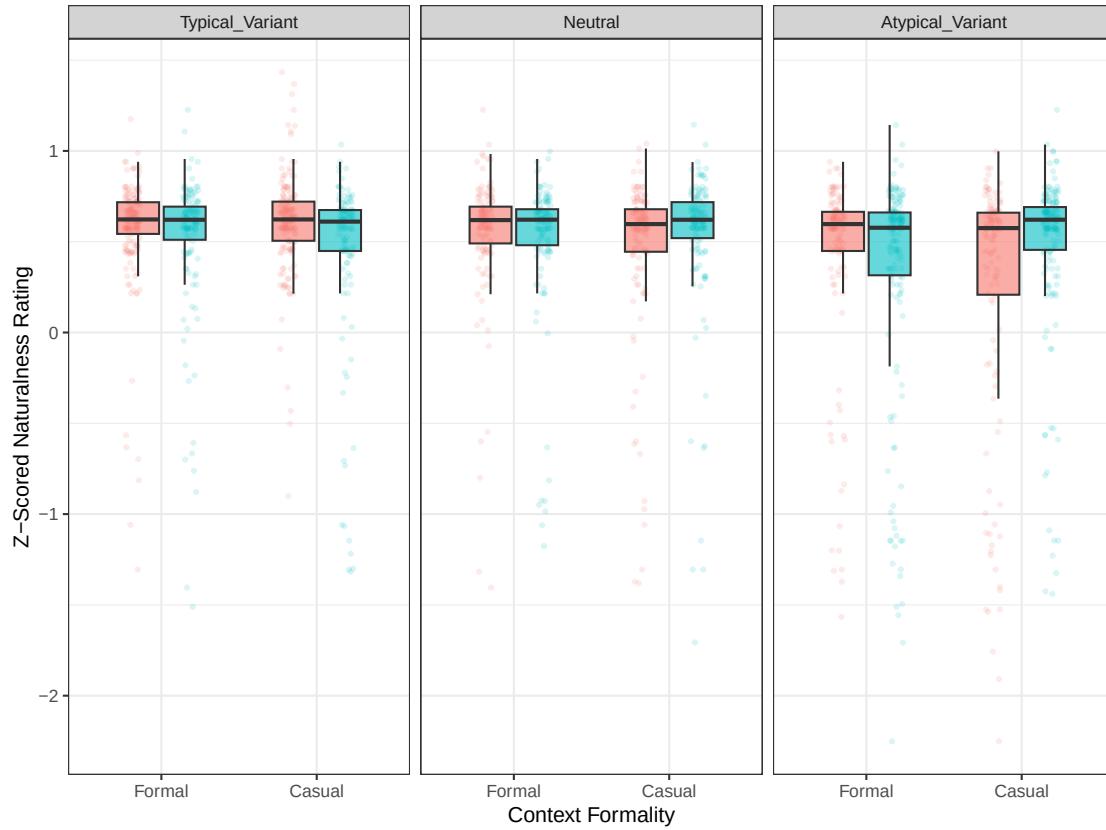
Significance codes: *** $p < .001$, ** $p < .01$, * $p < .05$, . $p < .1$

In the Typical condition (left panel), ratings are relatively stable, indicating that lexical conservativity protects the variant from social penalties. Post-hoc pairwise comparisons confirmed that context had no significant effect on ratings for Typical words ($p > .30$) for either vowel.

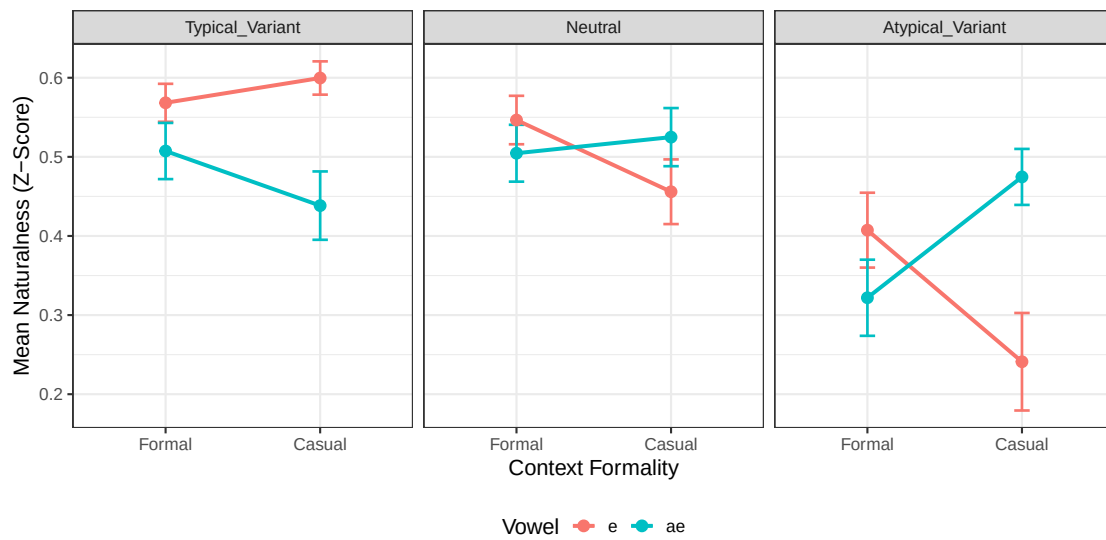
However, in the Atypical condition (right panel), the interaction drives the effect. The most dramatic shift was observed for the atypical [e], which dropped significantly in ratings when used in a casual context (Estimate = 0.18, $p < .001$). A smaller but significant shift was also observed for the atypical [æ], which was rated lower in the formal context compared to the casual one (Estimate = −0.09, $p = .045$). This confirms that the "stylistic penalty" is most severe when a speaker violates both stylistic norms (context) and lexical norms (usage) simultaneously, with the penalty for the formal [e] variant being nearly double the magnitude of the effect for [æ].

5 DISCUSSION

The present study investigated whether the [e] ~ [æ] variation in Standard Turkish carries stylistic meaning and whether this meaning is modulated by the register of the carrier phrase. While we initially sought to determine if the variants functioned as simple markers of formality or intimacy, our results revealed a more complex architecture of sociolinguistic perception. We found that listeners are indeed sensitive to stylistic coherence, yet this sensitivity appears to be constrained by another factor, lexical typicality. The results are compatible with a hierarchical view of perception where internal linguistic expectations serve as a primary filter, protecting stylistically incongruent variants from social penalty.



(a) Distribution of z-scored Ratings



(b) Mean z-scored Ratings

Figure 3: Interaction of Context and Vowel by Lexical Typicality (Standard Istanbul speakers, $N = 88$). The crossover effect is minimal for Typical words (left), moderate for Neutral words (middle), and severe for Atypical words (right).

5.1 THE LEXICAL FILTER ON STYLISTIC COHERENCE

Current models of sociolinguistic perception emphasize stylistic schemata or constellations, where variants from different levels are expected to covary in a coherent fashion (Vaughn and Kendall 2019, Montgomery and Moore 2018). Under a strong interpretation of stylistic coherence, any mismatch between a variant and its context—such as a formal vowel in a casual sentence—should trigger a processing cost or social penalty, as the listener detects a violation of the active stylistic template.

Our results suggest that this coherence constraint is not absolute, but is instead mediated by lexical representation. We found that stylistic incongruence triggered a significant penalty only when the variant was also lexically atypical. When a target word was produced with its lexically preferred vowel (e.g., a word that typically takes [e] produced with [e]), listeners rated it as natural regardless of the stylistic frame. The potential conflict between the formal vowel and the casual frame was effectively neutralized by the phonological expectedness of the word.

This pattern indicates that the stylistic value of the [e] ~ [æ] variation is not accessed in every instance, but is instead constrained by lexical expectations. This finding aligns with predictive processing accounts of sociolinguistic salience, which define salience as a function of surprisal—the extent to which a signal diverges from prior expectations (Jaeger and Weatherholtz 2016). When a vowel matches the listener's lexical prior, the signal is processed with high fluency. Because the variant generates minimal prediction error at the lexical level, it may not trigger the attentional resources required to evaluate stylistic congruency (Zarcone et al. 2016).

A surprisal based account also parallels Vaughn (2022b), who shows that listeners' social evaluations of variable ING are modulated by a word's typical grammatical function across the lexicon, even when its local syntactic use is controlled. In both that study and ours, cumulative, word-specific experience exerts a strong influence on social processing. Our results indicate that such lexical expectations can be strong enough not just to amplify stylistic effects, but to modulate them: when lexical expectations are satisfied, the salience of stylistic incongruence is undermined.

5.2 ASYMMETRIC SALIENCE AND PHONOLOGICAL MARKEDNESS

When penalties did occur (in the Atypical condition), they were strongly asymmetric: the use of an atypical formal variant ([e]) in a casual frame received a penalty nearly double the magnitude of the reverse mismatch (atypical [æ] in a formal frame).

We propose that this asymmetry is driven by the phonological status of the variants. As detailed in Section 2.5, the use of [e] before sonorants is an exception to the dominant allophonic rule of lowering. Because these forms do not obey the prevalent phonological distribution, they are phonologically marked.

This interpretation reflects a broader trend in the “third wave” of variation studies, which suggests that the variant with the richest social meaning tends to be the one that is linguistically marked (Beltrama and Staum Casasanto 2022). While Beltrama and Staum Casasanto demonstrate this for semantic intensifiers, they note that the principle holds across linguistic domains. For example, Podesva (2011) argues that rising intonation in declaratives acquires specific social indexicality precisely because it is the infrequent (marked) realization. Similarly, Callier (2013) finds that vocal creak triggers stronger social evaluations in syntactic positions where it is least expected.

In our data, the [e] variant represents the marked, exceptional form. Its phonological markedness—stemming from its violation of the otherwise robust surface generalization of lowering before sonorant

codas—generates higher surprisal (Jaeger and Weatherholtz 2016), thereby drawing the attentional resources necessary to encode social meaning (Zarcone et al. 2016). Consequently, when a speaker produces an atypical [e], they are producing a highly salient signal that listeners readily index as “formal”. In contrast, the use of [æ]—the unmarked, default form—fails to generate the same degree of surprisal or social salience, resulting in a significantly weaker penalty when used incongruently.

5.3 LIMITATIONS AND FUTURE DIRECTIONS

While our findings provide strong evidence for the interaction of lexical and stylistic constraints, our design introduces a limitation regarding the time-course of processing. In our materials, the primary stylistic manipulation (the agreement morphology) appeared at the right edge of the sentence, after the target word. This leaves open the possibility that the apparent dominance of lexical information is partly a consequence of linear order. Because the target vowel precedes the clearest stylistic cue, listeners may anchor their interpretation on the lexical item first, making them less likely to revise their judgment when the conflicting stylistic cue arrives later. Future work can directly address this by placing a stylistic cue (e.g., a formal second-person pronoun or address term) before the target word. If the lexical protection we observed is robust, then typical vowel words should remain immune to penalty even when an incongruent frame is cued. If, however, the effect is driven by temporal primacy, a pre-established incongruent frame might heighten the surprisal of a typical vowel, causing listeners to penalize it even if it is lexically typical.

A further limitation concerns our post-hoc estimation of “lexical bias,” which relied on aggregate norms from a small group ($N = 8$). While this approach captures broad community tendencies, it masks idiolectal variation; an individual listener’s lexical representation may diverge from the group average. Consequently, a variant classified as “high surprisal” based on the group norm might actually be the expected form for a specific listener. To test the hypothesis that social monitoring is triggered by subjective surprisal, future studies should adopt a within-subjects design that measures individual baselines. By administering a post-experiment forced-choice task to the same participants who provided the social ratings, we can derive listener-specific preference scores. Incorporating these individual priors into the model would provide a more precise test of whether social weight is determined by the variant’s divergence from the listener’s own internal expectations.

Finally, the method employed here—measuring naturalness in context—warrants a complementary investigation into the perception of these vowels in isolation. Our results suggest that the social meaning of [e] is highly context-dependent. Asking listeners to rate isolated tokens of the same word with [e] versus [æ] for perceived formality would test the metalinguistic accessibility of this contrast. If listeners cannot reliably differentiate the formality of isolated tokens (which was our experience, and which led us to design this experimental paradigm), it would confirm that the strong penalty observed in our study arises only when a salient, high-surprisal vowel interacts with a conflicting context.

6 CONCLUSION

In sum, our results indicate that the [e] ~ [æ] variation carries robust stylistic meaning in Standard Turkish, but that this meaning is not an invariant property of the vowel. Instead, social evaluation appears to be modulated by lexical expectations. Stylistic incongruence triggers a penalty primarily when it involves a lexically dispreferred variant, suggesting that the sociolinguistic monitor is activated not by

the variant itself, but by the spike in surprisal caused by a lexical violation. This finding highlights the necessity of integrating lexical-level constraints into models of stylistic processing.

REFERENCES

- Agha, Asif. 2003. The social life of cultural value. *Language & Communication* 23:231–273.
- Barr, Dale J., Roger Levy, Christoph Scheepers, and Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68:255–278.
- Beltrama, Andrea, and Laura Staum Casasanto. 2022. The social meaning of semantic properties. In *Social meaning and linguistic variation: Theorizing the third wave*, ed. Lauren Hall-Lew, Emma Moore, and Robert J. Podesva, 126–151. Cambridge University Press.
- Bender, Emily M. 2005. On the boundaries of linguistic competence: Matched-guise experiments as evidence of knowledge of grammar. *Lingua* 115:1579–1598.
- Boersma, Paul. 2001. Praat, a system for doing phonetics by computer. *Glott International* 5:341–345.
- Buson, Laurence, Aurélie Nardy, Dominique Muller, and Jean-Pierre Chevrot. 2018. The sociolinguistic repetition task: A new paradigm for exploring the cognitive coherence of language varieties. *Topics in Cognitive Science* 10:803–817.
- Callier, Patrick. 2013. Linguistic context and the social meaning of voice quality variation. Doctoral Dissertation, Georgetown University, Washington, DC.
- Campbell-Kibler, Kathryn. 2009. The nature of sociolinguistic perception. *Language Variation and Change* 21:135–156.
- Campbell-Kibler, Kathryn. 2010. The sociolinguistic variant as a carrier of social meaning. *Language Variation and Change* 22:423–441.
- Dadan, M. C., U. Erken, and Stefano Canalis. 2025. /e/ lowering in Turkish. In *Proceedings of the Workshop on Turkic and Languages in Contact with Turkic*, volume 9.
- Demirel, Serhat. 2021. *Türkçe Telaffuz Sözlüğü*. İstanbul: Radyo ve Televizyon Üst Kurulu.
- Eckert, Penelope. 2008. Variation and the indexical field. *Journal of Sociolinguistics* 12:453–476.
- Gafter, Roey J., and Uri Horesh. 2020. Two languages, one variable? pharyngeal realizations among Arabic–Hebrew bilinguals. *Journal of Sociolinguistics* 24:369–387.
- Göksel, Aslı, and Celia Kerslake. 2005. *Turkish: A comprehensive grammar*. London: Routledge.
- Gopal, Deepthi. 2018. Similarity and representations in sonorant phonology. Ph.d. thesis, University of Manchester, School of Arts, Languages and Cultures.
- Günhan, Tolga. 2023. Tolga günhan [video]. <https://youtu.be/Fy8PZEhmztE?si=SEW5F9GhTWYo93Zw>. Accessed: 2024-10-20.
- Jaeger, T. Florian, and Kodi Weatherholtz. 2016. What the heck is salience? how predictive language processing contributes to sociolinguistic perception. *Frontiers in Psychology* 7:1115.
- Kuznetsova, Alexandra, Per Bruun Brockhoff, and Rune Haubo Bojesen Christensen. 2017. lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software* 82:1–26.

- Labov, William. 1963. The social motivation of a sound change. *WORD* 19:273–309.
- Labov, William. 2006. *The social stratification of english in new york city*. Cambridge: Cambridge University Press, 2 edition.
- Labov, William, Sharon Ash, Maya Ravindranath, Tracey Weldon, Maciej Baranowski, and Naomi Nagy. 2011. Properties of the sociolinguistic monitor. *Journal of Sociolinguistics* 15:431–463.
- Levon, Erez, and Sue Fox. 2014. Social salience and the sociolinguistic monitor: A case study of ING and TH-fronting in Britain. *Journal of English Linguistics* 42:185–217.
- Lüdecke, Daniel, Mattan S. Ben-Shachar, Indrajeet Patil, Brenton M. Wiernik, Etienne Bacher, Rémi Thériault, and Dominique Makowski. 2022. *easystats: Framework for easy statistical modeling, visualization, and reporting*. CRAN. *R package*.
- Montgomery, Chris, and Emma Moore. 2018. Evaluating S-cilly voices: The effects of salience, stereotypes, and co-present language variables on real-time reactions to regional speech. *Language in Society* 47:227–255.
- Ngom, Fallou. 2003. The social status of Arabic, French, and English in the Senegalese speech community. *Language Variation and Change* 15:351–368.
- Okamoto, Shigeko, and Miyako Morimoto. 2023. Gender norms and styling in japanese conversation: A multilevel analysis. *Journal of Sociolinguistics* 27:42–65.
- Pharao, Nicolai, Marie Maegaard, Janus Spindler Møller, and Tore Kristiansen. 2014. Indexical meanings of [s] among Copenhagen youth: Social perception of a phonetic variant in different prosodic contexts. *Language in Society* 43:1–31.
- Podesva, Robert J. 2008. Three sources of stylistic meaning. *Texas Linguistic Forum* 51:134–143.
- Podesva, Robert J. 2011. Salience and the social meaning of declarative contours: Three case studies of gay professionals. *Journal of English Linguistics* 39:233–264.
- Regan, Brendan. 2022. The social meaning of a merger: The evaluation of an andalusian spanish consonant merger (ceceo). *Language in Society* 51:481–510.
- Sak, Haşim, Tunga Güngör, and Murat Saraçlar. 2008. Turkish language resources: Morphological parser, morphological disambiguator and web corpus. In *Advances in natural language processing*, ed. Bengt Nordström and Aarne Ranta, volume 5221 of *Lecture Notes in Computer Science*, 417–427. Berlin, Heidelberg: Springer.
- Schieltzeth, Holger, Niels J. Dingemanse, Shinichi Nakagawa, David F. Westneat, H. Allegue, Céline Teplitzky, Denis Réale, N. A. Dochtermann, László Z. Garamszegi, and Yimen G. Araya-Ajoy. 2020. Robustness of linear mixed-effects models to violations of distributional assumptions. *Methods in Ecology and Evolution* 11:1141–1152.
- Tamminga, Meredith. 2017. Matched guise effects can be robust to speech style. *The Journal of the Acoustical Society of America* 142:EL18.
- Tamminga, Meredith, and Lacey Wade. 2022. Coherence across social and temporal scales. In *Linguistic variation: Structure and context*, ed. Gregory R. Guy, H. Ahlers, and Meredith Tamminga, 40–62. New York: Routledge.
- Trudgill, Peter. 1972. Sex, covert prestige and linguistic change in the urban british english of norwich.

- Language in Society* 1:179–195.
- Vaughn, Charlotte. 2022a. The role of internal constraints and stylistic congruence on a variant's social impact. *Language Variation and Change* 34:331–354.
- Vaughn, Charlotte. 2022b. What carries greater social weight, a linguistic variant's local use or its typical use? In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 44.
- Vaughn, Charlotte, and Tyler Kendall. 2019. Stylistically coherent variants: Cognitive representation of social meaning. *Variantes estilisticamente coerentes: Representação cognitiva de significados sociais* 27:1787–1830.
- Zarcone, Alessandra, Marten van Schijndel, Jorrig Vogels, and Vera Demberg. 2016. Salience and attention in surprisal-based accounts of language processing. *Frontiers in Psychology* 7:844.
- Zehr, Jeremy, and F. Schwarz. 2018. PennController for Internet Based Experiments (IBEX).
- Zhang, Qing. 2005. A chinese yuppie in beijing: Phonological variation and the construction of a new professional identity. *Language in Society* 34:431–466.