

ARTICLE

# Learning speaker-specific linguistic ‘style’ is mediated by deviance from common language use

Nitzan Trainin<sup>1,2</sup>  and Einat Shetreet<sup>1,2</sup>

<sup>1</sup>Department of Linguistics, Tel Aviv University, Tel Aviv, Israel; <sup>2</sup>Sagol School of Neuroscience, Tel Aviv University, Tel Aviv, Israel

**Corresponding author:** Nitzan Trainin; Email: [nitzant@mail.tau.ac.il](mailto:nitzant@mail.tau.ac.il)

(Received 08 December 2022; Revised 22 June 2023; Accepted 29 June 2023)

## Abstract

Different speakers sometimes convey similar meanings differently. This study examined whether listeners could learn to associate a specific linguistic ‘style’ with a certain speaker, with no apparent difference in meaning, and the role of unnatural linguistic choices (or unexpectedness) in such learning. We created an inter-speaker variation in ‘style’ using the weak adjective ordering preferences in Hebrew. Participants were exposed to two different speakers, each producing a different adjective order, consistently. We manipulated the combinations of order pairings, based on their naturalness (with two natural orders, a natural and an unnatural order, and two unnatural orders), and examined participants’ ability to associate a unique order with a specific speaker. In two experiments, using different statistical analyses, we show that listeners can learn speaker-specific language use when it is irrelevant for meaning inferences, when deviance from natural or expected language use is involved. We further discuss whether learning may be facilitated by differences in naturalness or structural form. Our findings suggest that listeners are sensitive to inter-speaker variability in ‘style’, mostly when this ‘style’ is unexpected. This is in line with the predictions of Surprisal theory, and may suggest that surprisal plays a major role in learning speaker-specific language use.

**Keywords:** speaker-specific; reference; surprisal; communication efficiency; learning

## 1. Introduction

Variability between speakers in their language use occurs frequently (Jacewicz et al., 2010; Redi & Shattuck-Hufnagel, 2001; Westbury et al., 1998). For the sake of efficient communication, it seems beneficial to correctly map inter-speaker variation with the corresponding speakers. This mapping, which involves adjusting expectations based on a speaker’s preferences, could facilitate interpretation.

© The Author(s), 2023. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.



Consider the term ‘probably’, for example. It can refer to different ranges of probabilities in the real world. When interpreting a sentence such as “It will probably rain tomorrow”, knowing the range of probabilities ‘probably’ conveys for the specific speaker who uttered this sentence, would help us to assess the likelihood of the event, and hence to decide whether or not to take an umbrella (see Schuster & Degen, 2020 for experimental results on speaker-specificity with ranges of probabilities).

Previous research showed that listeners can detect variability in language use across different speakers. It has been shown that listeners can recognize inter-speaker variability in phonetic characteristics (Creel et al., 2008; Eisner & McQueen, 2005; Kraljic & Samuel, 2007; McLennan & Luce, 2005), in syntactic preferences (Kamide, 2012; although, see Ostrand & Ferreira, 2019), in uncertainty expressions (Schuster & Degen, 2020), in disfluencies (Orena & White, 2015; Yoon et al., 2021), and in referential communication (Grodner & Sedivy, 2011; Metzging & Brennan, 2003; Pogue et al., 2016). Furthermore, previous studies report that listeners can use their knowledge about a specific speaker to apply different processing strategies and derive different meanings (Davies et al., 2017; Khan et al., 2009; Maye et al., 2008; Skoruppa & Peperkamp, 2011; Sobel et al., 2012). In these cases, the motivation to adapt to the language use of specific speakers is clear: such adaptation facilitates the derivation of meaning from a person’s utterance.

Speaker-specificity has been demonstrated also in studies of *lexical entrainment*, a phenomenon describing cases where speakers tend to align with respect to their lexical choices (Brennan & Clark, 1996; Brown-Schmidt, 2009; Garrod & Anderson, 1987; Metzging & Brennan, 2003; Wilkes-Gibbs & Clark, 1992). That is, speakers will use a certain word with one speaker but not with another, due to their previously established references in conversation. For example, participants who engaged in conversation with a confederate who used the uncommon term ‘penny loafer’ for a specific shoe (in a context with more than one shoe) tended to use the same term too, even when it was not required by context. However, when introduced to a new interlocutor/confederate, they revert to using the typical term ‘shoe’ (Brennan & Clark, 1996). These findings suggest that paying attention to speaker-specific language use and mapping inter-speaker variability may be used to establish common ground and facilitate meaning inference in an interactive setting. Moreover, it could assist the prediction of speakers’ utterances and thus make communication more efficient, either by reducing uncertainty as per the utterance’s intended meaning (Bangerter et al., 2020; Brennan & Clark, 1996) or by merely making specific utterances more available in memory (Barr & Keysar, 2005; Kronmüller & Barr, 2007; Pickering & Garrod, 2004, 2006).

### 1.1. Recognizing non-crucial speaker-specific language use

Variation in language use does not always entail critical meaning differences. For example, in discussing the publishing of a scientific work, speaker A might refer to it as ‘a paper’, while speaker B might refer to it as ‘an article’. Yet, misunderstanding is unlikely in this case. Some studies have investigated the recognition of speaker-specific language use in cases of non-crucial variation (Kroczeck & Gunter, 2017; Ostrand & Ferreira, 2019). In these studies, the researchers created inter-speaker

variability at the sentence level that did not substantially alter meaning<sup>1</sup>. Kroczek and Gunter (2017) found that participants learned to associate a specific order with a specific speaker (for SOV versus OSV in German), whereas Ostrand and Ferreira (2019) did not observe any speaker-specific learning or alignment (for various syntactic structures, e.g., prepositional datives versus double object).

Some critical differences between the studies could account for their opposing results. First, the experimental tasks differed substantially. Kroczek and Gunter used comprehension questions throughout the experiment, which required participants to concentrate on the thematic relations within the sentence (i.e., by asking who is the object of the sentence, e.g., “Who was seen?” for the sentences in Footnote 1). These questions were likely to draw attention to the sentence structure and thus highlight the differences between the tested structures, resulting in better identification of the inter-speaker variation. Ostrand and Ferreira used a more implicit manipulation – an interactive picture-matching task, in which participants were required to choose the correct image based on their interlocutor’s description. Because on each trial there was only one image that could have been described by each description, variation in structure was not linked to semantic/contextual variance, and therefore the sentence structure was not likely to influence success.

Second, Kroczek and Gunter contrasted a high-frequency word order (SOV) with a low-frequency word order (OSV), while Ostrand and Ferreira used structures with smaller differences in frequency. Low-frequency structures are probably less expected to be produced because speakers are sensitive to the overall frequencies of linguistic material in their language (e.g., Bod et al., 2003; Bybee & Hopper, 2001; Ellis, 2002; Garnsey et al., 1997; Lieven, 2010; MacDonald, 1994). Unexpected structures, being more salient (Rácz, 2013), may draw more attention to inter-speaker variations. This, in turn, can result in better discrimination between speakers who use them and more conventional speakers.

### 1.2. Factors influencing the recognition of non-crucial speaker-specific language use

As discussed above, it is possible that people recognize and map inter-speaker variation even in cases where such variation is non-meaning-crucial. We have already mentioned facilitation in communication and linguistic unexpectedness as possible explanations for successful speaker-specific recognition in these cases. Below, we discuss these in more detail.

One possibility for mapping non-crucial inter-speaker variation in language could be the facilitation of communicational efficiency (Barr, 2004; Metzging & Brennan, 2003). It could either reduce uncertainty concerning the upcoming utterance when listening to a known speaker or increase a specific utterance’s salience in memory, thereby making it more accessible. Both views come from the literature on *lexical entrainment*. Although the current study concerns the phrasal level (see below), we discuss these ideas (coming from the lexical level) for their relevant aspects of speaker-specificity.

<sup>1</sup>For example, in Kroczek and Gunter’s study, “Today has the [nom.] man the [acc.] friend seen” (SOV) is equivalent to “Today has the [acc.] friend the [nom.] man seen” (OSV). In Ostrand & Ferreira’s study, “A boy is tossing his dad the ball” (double object) is equivalent to “A boy is tossing the ball to his dad” (prepositional dative).

According to one view, it has been suggested that during collaborative tasks, interlocutors interactively create shared linguistic representations, which are sometimes referred to as *conceptual pacts* (Brennan & Clark, 1996). These pacts facilitate reference resolution because they reduce uncertainty in predicting an interlocutor's utterance. This, in turn, can reduce cognitive load and thus be more efficient from the listener's perspective. Moreover, these pacts have been claimed to be speaker-specific, such that using a new, pact-breaking, utterance will be more felicitous with a 'new' interlocutor than with an 'old' one with whom the pact is established (Metzing & Brennan, 2003). This could mean that specific terms are actively stored in memory alongside the person who uttered them.

A different view poses that the alignment process is driven by automatic memory processes. Under this view, interlocutors tend to adopt each other's terminology because it is more available in memory and is, therefore, more readily accessed (e.g., Barr & Keysar, 2005; Holtgraves & Barr, 2014; Keysar et al., 1998; Kronmüller & Barr, 2007), making the retrieval process more efficient. According to this view, learning speaker-specific language use is not conversationally motivated. This view further suggests that successful collaboration does not necessarily involve explicitly storing speaker-specific language use, because interlocutors adhere to an egocentric strategy that is based on their memory demands alone (Keysar et al., 1998). However, since speaker-specific effects have been shown to appear in later stages of processing (e.g., Kronmüller & Barr, 2007; Kronmüller & Guerra, 2020), it is possible that speaker-specific language use is somehow stored in these cases, as a by-product of the interaction.

Another possible explanation of successful speaker-specific recognition comes from Surprisal theory (Hale, 2001; Levy, 2008). This theory maintains that listeners use their linguistic experience to derive probabilities of encountering certain utterances and form expectations about the future linguistic input they will come across. That is, the more improbable (surprising) it is for a word to appear in a given context, the higher the cognitive load associated with accessing this word will be. Moreover, surprising events have been shown to be more memorable (Adler, 2008; Upala et al., 2007). Indeed, several studies have linked surprisal to memory and learning (Foster & Keane, 2019; Futrell et al., 2020; Munnich & Ranney, 2019). Among these, learning speaker-specific language use has also been studied in the context of surprisal (Lai et al., 2020). In that study, inter-speaker variation was manipulated using artificial alien languages. The researchers exposed participants to differential linguistic (morphological) input and showed that novel, and therefore unexpected, linguistic material at a later phase led to a better association of a morpheme with a specific social (alien) group. In that case, inter-speaker variation was not crucial for interpretation, but rather – as the authors suggested – learning was driven by the violation of expectations regarding the linguistic input. This process could also be linked to communicative efficiency because using words that are more predictable in a specific context facilitates processing (Futrell et al., 2020). Therefore, mapping inter-speaker variability and storing speaker-specific language use could increase communication efficiency, by means of reducing unexpectedness or surprisal.

### 1.3. The current study

In the current study, we ask whether listeners can differentiate between speakers based on their stylistic preferences, where no apparent critical difference in meaning

is presented in the linguistic choices of the speakers. We further ask whether this mapping of non-crucial speaker-specific language use is mediated by unexpectedness. To examine this, we selected the preferences of adjective ordering in Hebrew. Adjective ordering preferences are considered robust, and according to some studies, their variation may be meaningful (for a detailed discussion see Scontras, 2023). Below we discuss these aspects in relation to Hebrew.

As mentioned above, adjective ordering preferences are considered robust in many languages (Cinque, 1994; Danks & Glucksberg, 1971; Dixon, 1982; Laenzlinger, 2005; Martin, 1969a, 1969b; Martin & Molfese, 1972; Scontras et al., 2017, 2019; Scott, 2002; Svenonius, 2008; Whorf, 1945), with some cross-linguistic variance (Cinque, 1994; Sproat & Shih, 1991). However, a recent study (Trainin & Shetreet, 2021) suggests that adjective ordering preferences in Hebrew (a post-nominal language where adjectives appear after the noun they modify) are not as robust as has been previously claimed (Shlonsky, 2004; for similar results in Spanish, another post-nominal language, see Rosales & Scontras, 2019). In production, naturalness rating, and forced-choice tasks, ordering preferences in Hebrew were shown to be significantly weaker than in English. For example, in three-adjective strings of Size, Color, and Pattern semantic classes, only two different orders were produced in English, but five different orders were produced in Hebrew, none of which produced more than 30% of the time. Furthermore, the differences in naturalness between the least and the most natural orders in English (~4.5 versus ~6.5 respectively) were significantly larger than in Hebrew (~4.5 versus ~5.5). These results may suggest that adjective ordering in Hebrew is more susceptible to inter-speaker variation, especially with three-adjective strings. Furthermore, it is plausible to assume that Hebrew speakers would have less-constraining predictions towards adjective ordering because there are more viable options. Therefore, the weak adjective ordering preferences in Hebrew could allow us to examine stylistic non-crucial meaning preferences.

Some research on English suggests that adjective ordering variation is linked to meaning inference (Danks & Glucksberg, 1971; Danks & Schwenk, 1972, 1974; Scontras, 2023; Whorf, 1945). Specifically, it has been shown that deviation from the common adjective order in English, by placing a certain adjective earlier in the sequence than it usually is, could implicate that some feature of the object is especially worth emphasizing (Martin, 1970; Martin & Ferb, 1973). This is mainly shown in contrastive contexts. For example, in a context with two images of large pencils, differing only in color, it is more felicitous to position the color adjective before the size adjective (i.e., say “the yellow large pencil”, instead of the preferred order of “the large yellow pencil”) than in a non-contrastive context. Because Hebrew does not have a strong preference for a certain order, the displacement may not signal such meaning variation. Furthermore, to avoid such effects in the current study, we did not use contrastive contexts.

Our stimuli were composed of nouns modified by three-adjective strings, which allowed several permutations of orders (unlike two-adjective strings which allow only two permutations). Furthermore, the preferences for three-adjective strings in Hebrew were significantly weaker than for two adjectives (Trainin & Shetreet, 2021), enabling us to treat these preferences as stylistic. We selected adjectives from three semantic classes – Size, Color, and Pattern – for two main reasons. First, they are easy to depict visually, and therefore could be used for our visual task. Second, we had converging evidence from three tasks (Trainin & Shetreet, 2021) for the hierarchy of ordering preferences when using adjectives from these classes, which allowed us to

examine multiple contrasts of ordering preferences. This hierarchy is given in (1). Note that we abbreviate the orders, such that S stands for size, C for color, and P for pattern, and they are ordered in the linear order of Hebrew. For example, the order Noun-Size-Color-Pattern is abbreviated as SCP.

- (1) Hierarchy of adjective order preferences, based on Trainin and Shetreet (2021):

$$\text{SCP} \sim = \text{CSP} > \text{CPS} \sim = \text{PCS} > \text{SPC} \sim = \text{PSC}$$

To test the effects of expectedness, we focus on four different orders, placed at the ends of the hierarchy in (1): SCP and CSP at one end and PSC and SPC, at the other. Trainin and Shetreet (2021) showed that the SCP and CSP orders were produced more often than the SPC and PSC orders (~30% of the times per order versus ~0–2% per order, respectively), and were also considered slightly more natural (~5.5 versus ~4.5 on a scale of 1–7, respectively). Therefore, the SCP and CSP orders will be considered here as more natural and more common than the SPC and PSC orders. As such, they should also be more expected. In other words, the SPC and PSC orders should have higher surprisal values, because listeners should not expect the adjectives to appear in these orders.

Following Ostrand and Ferreira (2019), which used a task that did not draw attention to the language variation, we ensured that language use variation was not crucial for the task, by asking participants to verify whether a description in various adjective orders matched a picture with a single object (see Fig. 1). Because the visual context did not include any competitor which could drive a change in the adjective order, and because we did not ask participants about the manner of descriptions, it is unlikely that this task would make adjective ordering particularly salient. In our study, we used this verification task in an implicit *exposure phase* and then used an explicit recognition test in a *test phase* to examine whether participants mapped inter-speaker variability. In the *exposure phase*, participants heard two different speakers, each producing a different adjective order consistently. We created three conditions contrasting speakers using the two most natural and common orders, the two most unnatural and uncommon orders, and one common order and one uncommon order. In the *test phase*, the participants were asked to recognize which of the orders was produced by which speaker.

We hypothesize that if listeners are sensitive to inter-speaker variability in stylistic non-crucial language use, they should be able to correctly associate a specific adjective order with a specific speaker. If this is the case, participants should be able to map orders to speakers on all conditions. If naturalness or unexpectedness plays a role in successful speaker-specific learning, we may expect participants to have better order-speaker mapping in conditions where one of the speakers produces an unnatural order than in conditions with only natural orders. It is also possible that speakers detect inter-speaker variability only when it is associated with a crucial meaning variation, or needed for efficiency due to task demands. If so, participants should not be able to associate orders with speakers, under all conditions.

## 2. Experiment 1

To examine speaker-specific learning of non-crucial variation, we compared the ability of individuals to associate a speaker with his/her preferred adjective order

under three conditions. To further test the effects of unexpectedness, each condition included two speakers, varying on the naturalness of the orders they produced: (1) the two most natural adjective orders in Hebrew (High frequency – High frequency (HH); see Table 1); (2) one of the natural orders to the most unnatural one (High frequency – Low frequency (HL)); and (3) the two most unnatural orders in Hebrew (Low frequency – Low frequency (LL)). Although no substantial difference in naturalness is introduced in both HH and LL conditions, the LL condition includes a general violation of listeners' expectations as they are extensively exposed to unnatural orders. Therefore, this condition allowed us to examine – provided that an effect of frequency/naturalness does exist – whether it stems from differences in naturalness between speakers or from (at least) one speaker's deviance from common language use (and not relatively to the other speaker(s) in the context).

## 2.1. Methods

### 2.1.1. Participants

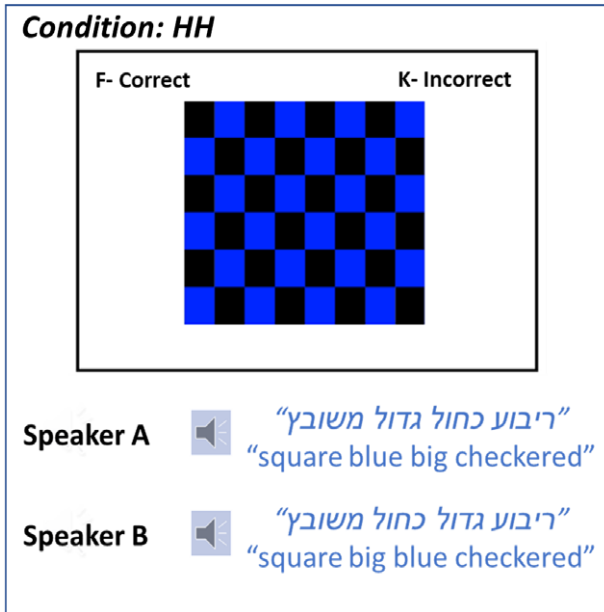
Sixty native Hebrew speakers (18–32 years old,  $M = 24.91$ ,  $SD = 3.22$ ; 63% (38) females), with no language, cognitive, or social impairments, were recruited via social media and were randomly assigned to one of the experimental conditions, with 20 participants in each condition. All participants gave informed consent to take part in the study. The study was approved by Tel Aviv University's ethics committee.

### 2.1.2. Materials and procedure

(a) *Exposure phase*. Ninety-six images of objects, comprised of four possible shapes (square/triangle/circle/star) in different sizes (large/small), in different colors (blue/green/red), and with different patterns (dots/checkers) were used (Fig. 1). We used the same adjective words as in Trainin and Shetreet (2021), for better confidence as per the differences in naturalness between the descriptions in the different adjective orders. Each image was paired with an auditory description including the three features: size, color, and pattern. Half of the descriptions matched the visual image (e.g., 'a big blue checkered square' for Fig. 1), and half did not (because one feature was wrong, e.g., 'a big red checkered square' for Fig. 1).

To make the distinction between the two speakers clear, one of the speakers was a female, and the other one was a male. Thus, half of the descriptions were spoken by a female speaker, presented to the participants by a conventional female name, Naama. The other half was spoken by a male speaker, presented to participants by a conventional male name, Yoav. Each speaker consistently produced the same adjective order throughout the experiment (SCP, CSP, SPC, or PSC, see Table 1), with Speaker A always using a different order than Speaker B.

On each trial, participants heard one of the speakers describing the image and were required to decide whether the description matched the image, by pressing F for correct descriptions or K for incorrect descriptions. For example, for Fig. 1, a matched description, where all the visual features of the image were included and correct in the SCP order, would be *square big blue checkered* (according to the post-nominal ordering in Hebrew), and a mismatched description, with a discrepancy in one of the visual features in the same order, would be *square big blue \*dotted*. The task was aimed to engage the participants in listening to the speakers' utterances, without explicitly addressing the adjective ordering preferences. To ensure that participants



**Figure 1.** An example of one stimulus in the exposure phase. In this example, the experimental condition was HH, where both orders are relatively natural – Speaker A uses the CSP order and Speaker B uses the SCP order. The descriptions were presented auditorily.

**Table 1.** A summary of the experimental conditions in Experiments 1 and 2

Condition	Speaker A	Speaker B
Natural 1 versus Natural 2 (HH)	SCP (Noun-Size-Color-Pattern)	CSP (Noun-Color-Size-Pattern)
Natural 1 versus Unnatural 1 (HL)	SCP (Noun-Size-Color-Pattern)	PSC (Noun-Pattern-Size-Color)
Unnatural 1 versus Unnatural 2 (LL)	PSC (Noun-Pattern-Size-Color)	SPC (Noun-Size-Pattern-Color)

listen to the end of each description, most incorrect descriptions were related to the feature presented by the last adjective in the string. There were four exposure blocks, two for each speaker. Each block consisted of 24 trials of image–audio pairs, all belonging to the same speaker. Blocks were interleaved such that a block presenting Speaker A was followed by a block presenting Speaker B, and so on. The order of blocks was counterbalanced for the identity of the first speaker (Female/Male) and the first order. Participants were instructed to listen to the audio clips and indicate their response using one of the keys (F/K; Fig. 1) as soon as the auditory description ended. No information was explicitly given about the speakers.

(b) *Test phase.* In this phase, participants were presented with written descriptions, without a picture (Fig. 2), such that the descriptions were neither correct nor incorrect, because they were grammatical and did not effectively describe any image. The descriptions were similar to those in the exposure phase, and included either the adjective order used by Speaker A or the one used by Speaker B. Participants were





**Figure 2.** An example of a trial in the test phase (English translation in blue). Naama is a common female name and Yoav is a common male name.

asked to decide which speaker had produced the written description, by pressing F for one speaker, and K for the other speaker (counterbalanced for side across participants). We did not explicitly mention that there is an inter-speaker variation, but simply instructed participants to recognize which speaker used each description. Overall, there were 24 test trials, 12 in each order from the exposure phase.

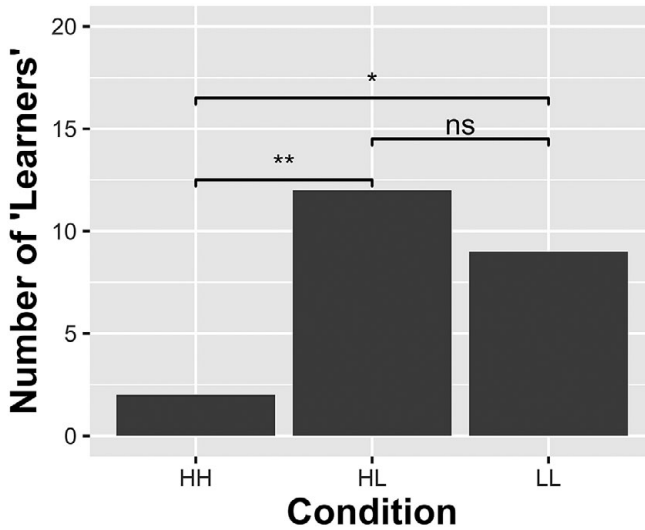
Participants were randomly assigned to one of the three conditions presented in Table 1. In each condition, the identity of the first speaker (Female/Male) and the assignment of a speaker to order was counterbalanced across participants, resulting in 4 lists per condition, with 5 participants in each list (for a total of 20 participants in each condition). The order of presentation in each exposure block and in the *test phase* was fully randomized for each participant. The experiment was carried out by the ‘Pavlovia’ (<https://pavlovia.org>) platform, using version 3.2 of PsychoPy (Peirce et al., 2019), and took approximately 15 minutes to complete.

## 2.2. Results and discussion

We used a binomial test for each participant separately to determine whether participants were able to learn which speaker produced which order. Participants whose accuracy rate in the *test phase* was significantly above chance level (at least 18 correct responses out of 24 trials) were classified as ‘Learners’, and the rest were classified as ‘Non-learners’. In all conditions, some participants were successful in learning speaker-specific preferences. However, the rates of successful learning were different across the conditions, with only 2 successful learners out of 20 in the HH condition contrasting the two natural orders (Fig. 3), 11/20 successful learners in the HL condition contrasting one of the two natural orders with the unnatural order, and 9/20 successful learners in the LL condition contrasting the two unnatural orders.

To verify that the learning rates significantly differed between the conditions, we conducted an equality of proportions test, using the ‘stats’ package (Version 4.2.1; R core team, 2022), on the proportions of successful learners in each condition<sup>2</sup>. This test revealed that both the HL condition and the LL condition yielded more successful learning than the HH condition ( $\chi^2(1) = 7.29, p = 0.007$  and  $\chi^2(1) = 4.51, p = 0.034$ ,

<sup>2</sup>A separate analysis comparing the number of correct classifications across conditions is given in the [Supplementary Materials](#). The overall pattern of results was similar to this analysis, with smaller effect sizes, probably due to the noisier analytic approach.



**Figure 3.** The number of successful learners in Experiment 1, by condition. HH = both speakers produced the natural orders; HL = one speaker produced a natural order and the other one produced an unnatural one; LL = both speakers produced the unnatural orders. ns = nonsignificant; \* =  $p < 0.05$ ; \*\* =  $p < 0.01$ . As with all plots in this manuscript, this plot was created using the 'ggplot2' package (Version 3.4.2; Wickham et al., 2022). Significance stars and lines were added using the 'ggsignif' package (Version 0.6.4; Ahlmann-Eltze & Patil, 2022)

respectively; Fig. 3). The difference between the HL and the LL conditions was not significant ( $\chi^2(1) = 0.1, p = 0.75$ ).

The comparison between the HH and LL conditions also allowed us to examine the effect of similarity in form (see Table 2). Both orders in each condition show some similarity in form, as they share the same adjective semantic class in the final position (SCP/CSP in HH and SPC/PSC in the LL condition; Table 1). Thus, the higher rates of learners in the LL condition compared with the HH also suggest that similarities in form do not by themselves hinder the ability to associate orders and speakers. Indeed, the LL condition yielded a significantly higher rate of learners. Because the orders were similar to the same extent, it appears that the reduced rates of learners in the HH condition were not due to similarity in form.

These results clearly show that learning speaker-specific 'style' is possible, albeit not in all cases. The most obvious difference between the conditions inducing higher learning rates and the one inducing lower learning rates was the inclusion of an

**Table 2.** The difference between conditions in terms of similarity in form and naturalness, and in terms of whether or not an unnatural order was included. Difference in naturalness is determined by Trainin and Shetreet (2021). Similarity in form means that adjectives from one of the semantic classes are placed in the same position across both orders

Condition	Similarity in naturalness	Similarity in form	Presence of an unnatural order
HH	✓	✓	×
HL	×	×	✓
LL	✓	✓	✓

unnatural order, regardless of how less natural this order is compared with the other order in the context.

Yet, we would like to exercise caution concerning this conclusion because of several important points. First, our sample size was small ( $N = 60$ , with 20 participants per condition). This sample size allowed us to observe large effects, but small-medium effects may have been missed (e.g., the difference between HL and LL). Second, we measured the rates of successful learning using a binomial test for each participant, according to which participants were classified as ‘Learners’ or ‘Non-learners’. Counting the number of learners based on a binomial test is dichotomizing a continuous variable. It could very well be the case that participants who were classified as ‘Non-learners’ did, in fact, detect the inter-speaker variability to some extent, but were ‘punished’ by the chance-level barrier. Finally, our *test phase* included only the two orders that were included in the *exposure phase*. Note that the task in the test phase did draw some attention to the speaker’s language use, as we asked participants to choose which description the speaker used. If so, including only the two orders used might have allowed participants to develop a strategy during the test phase, by guessing a specific order–speaker association and then consistently adhering to it. That is, participants could have succeeded in the test phase without having learned the correct association during the exposure phase. This might also be the factor that led to a few participants showing an opposite pattern in learning, performing significantly below chance level in the *test phase* of Experiment 1 (2 in the HH condition, 3 in the HL condition, and 1 in the LL condition). Importantly, the difficulty to distinguish between the orders (as may be determined by their naturalness) was also likely to modulate the stability of guessing. In other words, the more distinct these orders are from each other, the easier it is to remember and adhere to the guess, such that we expect more successful guessing rates in the HL condition (and possibly in the LL condition) than in the HH condition. To overcome these shortcomings, we conducted a second experiment.

### 3. Experiment 2

Experiment 2 aimed to replicate the findings from Experiment 1, using better methodological and statistical tools to examine the learning of speaker-specific non-crucial language use with adjective orders that deviate from common use. We included several modifications: (a) higher statistical power, with a much larger sample size estimated by an a-priori power analysis; (b) a more sensitive statistical analysis (see below); and (c) an addition of filler stimuli in the *test phase* which consisted of descriptions in an order not included in the *exposure phase*, to prevent response strategy.

In our second, well-powered experiment, we used a different approach to analyze the data – *d'* analysis inspired by the Signal Detection Theory (Swets et al., 1961). In this analysis, a *d'* measure is calculated for each participant. This measure provides an indication of how well a signal was detected by each participant because it is calculated by subtracting the proportion (Z-score) of incorrect positive responses (False Alarms) from the proportion of correct positive responses (Hits). In our case, this signal was the order–speaker association. *d'* measures allow for both the overall detection of learning within each condition and the comparison between conditions. *d'*s that are significantly different from 0 essentially mean that the signal was detected

to some degree. Furthermore, differences between conditions in  $d'$  could indicate that the ability to detect order–speaker association is different under different conditions.

The addition of filler trials allowed us to better mask the experimental manipulation. In these trials, participants were presented with additional adjective orders that were not included in the exposure phase. This addition could have prevented the guessing strategy discussed above in two ways. First, participants might attribute different reasoning to the task in the first place, namely, that they needed to recognize which orders they heard before, and not the order–speaker association. Second, participants might fail in using the guessing strategy without a-priori learning during the exposure phase, because the filler trials could interfere, in trying to assign 3 orders to only 2 speakers, and potentially lead to a breakdown.

### 3.1. Method

#### 3.1.1. Participants

One-hundred and ninety-two native Hebrew speakers (18–45 years old,  $M = 26.19$ ,  $SD = 4.71$ ; 63% (121) females), with no language, cognitive, or social impairments, were recruited via social media and were randomly assigned to one of the experimental conditions, with 64 participants in each condition. To determine the desired sample size for this experiment, we coerced  $d'$  analysis on our data from Experiment 1. This analysis in itself is unreliable because the responses in Experiment 1's *test phase* do not correspond to the type of responses in signal detection experiments. Nevertheless, we arbitrarily defined description assignment to the female speaker as 'Yes', and to the male speaker as 'No'. Accordingly, when participants assigned the correct order to the female speaker, it was considered a Hit (because it was a positive response to a 'true' state of 'Yes'), and when they assigned the order used by the male speaker to the female speaker, it was considered a False Alarm (a positive response for a 'true' state of 'No'). We then calculated the  $d'$  measure using the 'psycho' library (Makowski, 2018) in the R statistical software (Version 4.2.1; R core team, 2022), compared the  $d'$  between all conditions using independent two-sample t-tests (using 'stats' package), and obtained the effect sizes, using the 'effectsize' (Version 0.7.0.5; Ben-Shachar et al., 2020) package. Based on the calculated effect sizes, we estimated the desired sample size with an a-priori analysis using the G\*Power software (Version 3.1.9.4; Faul et al., 2007). We aimed for a sample size that will allow us to detect the smallest effect of interest, which was the difference between the HL and the LL conditions ( $t(37.97) = 1.43$ ; Cohen's  $d = 0.5$ ; Cohen, 1988). This analysis revealed that to obtain 80% power, 64 participants should be included in each group. All the participants gave informed consent to participate in the study. The study was approved by Tel Aviv University's ethics committee.

#### 3.1.2. Materials and procedure

The *exposure phase* was identical to the one used in Experiment 1. The *test phase* was also similar, but with minor modifications, to adapt it to the  $d'$  analysis. Instead of asking participants to assign the presented description on each trial to one of the speakers (Fig. 2), we asked, "Did X say Y?" (see Fig. 4), where X was the name of one of the speakers (Naama or Yoav) and Y was one of the orders used in the exposure phase or a third order used for fillers. Participants were required to answer 'Yes' (the F key) or 'No' (the K key). Furthermore, in addition to the 24 trials in which the description



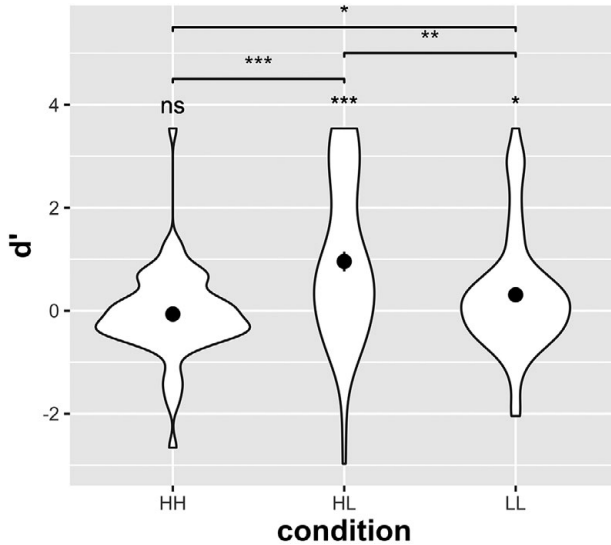
Figure 4. An example of a trial in the test phase of Experiment 2. English translation in blue.

was in one of the orders from the exposure phase, 6 additional filler trials included a third order which was not included in the exposure phase. This order was different for each condition and was selected such that it will be different in form as much as possible from the included orders. For example, for the HH condition, the third order was Noun-Pattern-Color-Size, where the pattern adjectives appear first, as opposed to the included orders (see Table 1), in which the pattern adjectives appeared last. These trials were included to prevent participants from developing a guessing strategy, because rejecting an association between one of the orders and one of the speakers does not entail accepting it for the other speaker. Overall, each participant completed 30 trials in the *test phase*.

### 3.2. Results and discussion

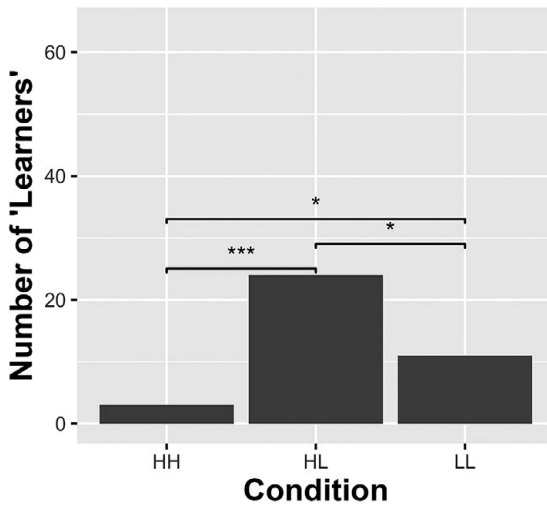
We have conducted two types of analysis with our  $d'$  measures. First, to assess the overall success in order–speaker association, we conducted one-sample t-tests in which we compared the  $d'$  in each condition to 0. Second, to assess the different learning rates across conditions, we performed a series of two-sample unpaired t-tests, comparing the  $d'$ s of the different conditions, HH, HL, and LL. The  $d'$  for each participant was calculated using the experimental items only (i.e., excluding fillers).

One-sample t-tests revealed that  $d'$ s in the HL ( $M = 0.96$ ,  $SD = 1.53$ ;  $t(63) = 5.00$ ,  $p < 0.001$ ) and in the LL ( $M = 0.31$ ,  $SD = 1.12$ ;  $t(63) = 2.19$ ,  $p = 0.03$ ), but not in the HH condition ( $M = -0.07$ ,  $SD = 0.85$ ;  $t(63) = -0.63$ ,  $p = 0.53$ ), were significantly larger than 0. This suggests that in the HL and LL conditions, participants could – to some degree – associate a speaker with their preferred orders correctly, but in the HH condition, they could not reliably do so. Pairwise comparisons revealed that the differences in  $d'$  between all conditions were significant (see Fig. 5), with  $d'$ s in the HL being higher than in the HH condition ( $t(98.27) = 4.68$ ,  $p < 0.001$ ) and in the LL condition ( $t(115.92) = 2.73$ ,  $p = 0.007$ ), and higher in the LL condition than in the HH condition ( $t(116.83) = 2.13$ ,  $p = 0.036$ ). This suggests that speaker-specific learning is facilitated both by differences in naturalness (or form; see General Discussion) and by the mere existence of deviation from common language use in the linguistic environment. Learning was virtually absent when both conditions were relatively and comparably natural (HH), slightly better when both orders were comparably and relatively unnatural (LL), and considerably better when one order was natural and one was not (HL).



**Figure 5.**  $d'$  by condition in the main analysis of Experiment 2. Dots and lines within each violin represent the group mean and Standard Error, respectively; The stars above each violin represent the comparison of each group's  $d'$  to a  $t$  distribution with 0 as its mean. HH = both speakers produced the natural orders; HL = one speaker produced a natural order and the other one produced an unnatural one; LL = both speakers produced the unnatural orders. ns = nonsignificant; \* =  $p < 0.05$ ; \*\* =  $p < 0.01$ ; \*\*\* =  $p < 0.001$ .

To be able to compare the results of Experiments 1 and 2, we further conducted the same analysis as in Experiment 1, counting the number of 'Learners' based on passing the threshold of chance-level performance as calculated by a binomial test (see Fig. 6). We then conducted a Chi-square test of equality of proportions and found that the



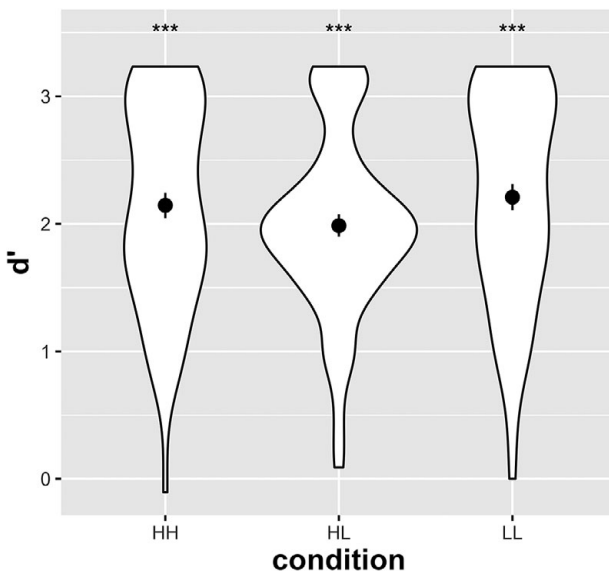
**Figure 6.** The number of 'Learners' by condition in Experiment 2. HH = both speakers produced the natural orders; HL = one speaker produced a natural order and the other one produced an unnatural one; LL = both speakers produced the unnatural orders.

number of learners significantly differed between all conditions, such that the proportion of 'Learners' in the HL condition (23/64) was significantly higher than the proportion of learners in the HH condition (3/64;  $\chi^2(1) = 17.42, p < 0.001$ ; Fig. 6) and the proportion of learners in the LL condition (11/64;  $\chi^2(1) = 4.85; p = 0.026$ ). Furthermore, the proportion of learners in the LL condition was significantly higher than in the HH condition ( $\chi^2(1) = 3.93, p = 0.047$ ).

### 3.3. An exploratory investigation of speaker-independent learning

The inclusion of fillers in this experiment also allowed us to investigate speaker-independent learning. If participants could correctly reject the association of filler orders to both speakers, it would mean that they generalized the overall linguistic environment properly. Thus, to examine speaker-independent learning, we calculated the  $d'$  measures for the filler trials (consisting of an order which was not used by any of the speakers) alone.

Similarly to what was done to investigate speaker-specific learning, we conducted a one-sample t-test for the fillers in each condition, compared to 0. For the filler trials,  $d'$ 's in all the conditions were significantly larger than 0, as determined by a one-sample t-test for each condition separately (HH:  $M = 2.15, SD = 0.79, t(63) = 21.6, p < 0.001$ ; HL:  $M = 1.99, SD = 0.69, t(63) = 22.96, p < 0.001$ ; and LL:  $M = 2.21, SD = 0.82; t(63) = 21.62, p < 0.001$ ; see Fig. 7). This result indicates that participants correctly recognized that the order which was not included in the exposure phase was indeed not uttered by any of the speakers. This further suggests that participants properly recognized the overall linguistic environment in all conditions.



**Figure 7.**  $d'$  by condition in the exploratory analysis (including fillers). Dots and lines within each violin represent the group mean and Standard Error, respectively; The stars above each violin represent the comparison of each group's  $d'$  to a t distribution with 0 as its mean. HH = both speakers produced the natural orders; HL = one speaker produced a natural order and the other one produced an unnatural one; LL = both speakers produced the unnatural orders. \*\*\* =  $p < 0.001$ .

Taken together, these results suggest that learning speaker-specific adjective ordering preferences is – as shown in Experiment 1 – extremely hard when both speakers produce natural orders. In the HH condition,  $d'$  was not significantly different from 0, and significantly smaller than in the HL and the LL conditions. In the other two conditions,  $d'$  was significantly larger than 0, indicating that at the group level, these conditions yield learning, to some extent. However, the  $d'$  in the HL condition was also significantly larger than in the LL condition, suggesting that learning is facilitated by increasing differences between the orders that listeners are exposed to (in form and in naturalness, as we discuss in the General Discussion).

Another point worth discussing is the substantial difference in learning rates between Experiment 1 and Experiment 2. One possible explanation is in the task. In Experiment 2 participants were required to answer 'Yes' or 'No' in the *test phase* and not to assign a description to a speaker in a forced-choice task like in Experiment 1. This means that in Experiment 2, participants might have been less aware of the manipulation, possibly thinking they were asked to recognize which orders they had heard overall, and not to recognize the inter-speaker variation. This was not the case in Experiment 1, where each description had to be assigned to a speaker, making the task more explicit and the manipulation more salient. Moreover, the contrast between the two orders in Experiment 1 may have deemed the differences in orders even more salient (for the role of contrast in task success, see Shetreet & Novogrodsky, 2019). Another explanation concerns the inclusion of a third order in the *test phase*, which was not included in the *exposure phase*. This perhaps did in fact prevent participants from developing a successful guessing strategy. In Experiment 1, all the descriptions in the *test phase* consisted of the orders used by the speakers from the *exposure phase*. Thus, to be defined as a successful learner, it was possible to notice the manipulation in the *test phase*, pick a guess, and then be consistent. In Experiment 2, the filler trials could have confused participants who would have gone down that road, because trying to assign 3 orders to 2 speakers with no memory of the order–speaker association might increase uncertainty and make the guessing behavior less successful.

#### 4. General discussion

In this study, we examined, in two experiments, the ability to learn speaker-specific stylistic non-meaning-crucial language use, and the role of unexpectedness in such learning. We used the weak adjective ordering preferences in Hebrew (Trainin & Shetreet, 2021), and compared speaker-specific learning rates across three adjective order pairs: (1) two natural orders in Hebrew (HH); (2) one natural order and one unnatural order (HL); and (3) two unnatural orders (LL). Our results show successful learning (in all combinations in Experiment 1 and in conditions 2 and 3 in Experiment 2). Notably, lower rates were observed when the two natural adjective orders were used (the HH condition). This indicates that listeners can learn speaker-specific language use, even when such use does not alter the interpretation of utterances, and furthermore, that this learning is modulated by unexpectedness.

One could argue that the learning observed in the current study is not simply 'stylistic' because it is possible that – in real-life referential communication –



differences in adjective ordering entail different meanings (Danks & Glucksberg, 1971; Danks & Schwenk, 1972, 1974; Scontras, 2023; Whorf, 1945). As discussed in the introduction section, this is unlikely in the context of the current study. First, Hebrew has weak adjective ordering preferences, such that using a certain adjective order to emphasize a certain property of an object (Martin, 1970; Martin & Ferb, 1973) may not be perceived as deviant. Second, our study used single (non-contrastive) images with banal and recurring sizes, colors, and patterns. Under these circumstances, participants were not expected to anticipate any meaning modification as is possibly conveyed by different adjective orders. Furthermore, if the difference in meaning was the driving force behind learning, we would expect it to operate under all conditions. Taken together, these suggest the difference between conditions could not be solely attributed to meaning differences.

The speaker-specific mapping of non-crucial language use has been demonstrated before, at the lexical level (e.g., Brennan & Clark, 1996; Metzger & Brennan, 2003) and the sentence level (Kroczek & Gunter, 2017). In our study, both speakers used the same words (in different orders) throughout the experiment. Therefore, distinguishing between their 'styles' requires going beyond the lexical level. Because adjective ordering does not appear to be subject to structural constraints in Hebrew (Trainin & Shetreet, 2021), this learning appears to occur within the phrase level in this study. Importantly, we show that such learning (that extends beyond the lexical level) occurs even without drawing explicit attention to the form of the utterance (cf. Kroczek & Gunter, 2017).

#### 4.1. Differential learning rates

One critical finding of our study is that learning rates of speaker-specific 'stylistic' language use were not similar in all cases. Specifically, learning rates were significantly lower when listeners were exposed to two speakers who uttered the most natural orders (the HH condition). Why would this be the case? We consider two explanatory factors: similarities in form and naturalness. The reader should note that these two factors are intertwined in our data because of the limited permutations of adjective ordering using three specific semantic classes. Unfortunately, data from Trainin and Shetreet (2021) on the naturalness of adjective ordering does not offer a possible comparison that could completely tease these factors apart<sup>3</sup>.

We first note that we cannot completely rule out the possibility that similarities in form play a role in the ability to learn speaker-specific adjective ordering preferences. Differential learning rates were observed for the condition with the two orders of different degrees of naturalness (HL) and the condition with two unnatural orders (LL) in Experiment 2, with the former showing higher learning rates. These two orders differ in both form and naturalness, such that both factors can explain the differences in learning rates (see Footnote 3).

<sup>3</sup>Ideally, we would have liked to compare a condition in which the two orders were similar in form but substantially different in naturalness. However, the pair of orders that are similar in form (SCP and CSP in the HH condition, PSC and SPC in the LL condition, and the unexamined PCS and CPS) are also similar in naturalness, according to Trainin and Shetreet. That leaves this limitation of our study unresolved.

Yet, our findings suggest that similarities in form could not be the only explanation for the difference in learning rates. Notably, similarities in form in the semantic class in the last position occurred in two tested conditions: the HH condition comparing two natural orders (SCP & CSP) and the LL condition comparing two unnatural orders (PSC & SPC; see Table 2). Despite these similarities, these two conditions yielded differential learning rates with the condition comparing two unnatural orders showing higher learning rates than the condition with two natural orders. Thus, at the very least, our findings suggest that the presence of at least one speaker who deviates from common language use facilitates speaker-specific learning (present in both the HL and the LL conditions).

Therefore, it is likely that learning speaker-specific 'style' is facilitated when the style is unexpected or perceived as peculiar. Unexpected and peculiar speakers may draw more attention to themselves because the language use of such speakers might be considered somewhat infelicitous (e.g., violating the Grice Maxim of manner; Grice, 1975). Listeners may try to understand why the speaker communicated in such an infelicitous manner, to better understand the communicative situation, and to better predict his/her linguistic behavior in future exchanges.

The differential learning rates observed across conditions are compatible with the results of studies of non-crucial speaker-specific variation at the sentence level, despite their contrasting results. Kroczeck and Gunter (2017) used the common SOV order and the uncommon OSV order and showed evidence that listeners could learn speaker-specific preferences. Similarly to that study, we observed speaker-specific learning mainly in conditions that included an uncommon and unnatural adjective order. Ostrand and Ferreira (2019), on the other hand, contrasted natural structures in English (e.g., double object versus prepositional datives) and did not observe speaker-specific learning. Indeed, the condition in our study that included two natural and common adjective orders yielded virtually no learning. This suggests that the difference between the findings of these two studies could be related to the naturalness (or in other words, the expectedness status) of the linguistic material in the exposure phase.

#### 4.2. Motivations for learning non-crucial speaker-specific language use

It is suggested that learning speaker-specific language use is motivated by the need for efficient communication. We assume that this is the case also for learning speaker-specific *stylistic* preferences, as tested in the current study. Because adjective ordering preferences in Hebrew are relatively weak, it is likely that Hebrew speakers do not have clear a-priori expectations towards them. Identifying speaker-specific preferences and detecting a consistent adjective order per speaker could speed up the processing of the referential phrase and therefore aid object identification. For example, learning that a specific speaker always produces the SCP order would allow the prediction of the class of adjective that appears at a given point in the phrase, thereby facilitating the processing of the phrase (i.e., expect a size adjective in the first position, and so on).

It is important to note that the task we used did not demand communicative efficiency, because it was not interactive, and also because there was only one image on each trial, and therefore no visual search was needed for object identification. This could mean that learning speaker-specific language use is not necessarily only used

for enhancing efficiency within task, but could rather be related to a more general mechanism, by which learning speaker-specific language use might reduce uncertainty in future interactions.

A second motivation that we offered for learning speaker-specific language use comes from Surprisal theory (Hale, 2001; Levy, 2008). According to this theory, processing difficulty depends on the probability of a certain linguistic input appearing in a specific context. Specifically, lower surprisal (i.e., higher probability) is linked to easier processing. Surprisal was found to be relevant for learning speaker-specific language use, where a better association of certain speakers to novel linguistic materials was observed when the stimuli had higher surprisal values (Lai et al., 2020). The unnatural orders we used here were virtually never produced (0 ~ 2% of the time, Trainin & Shetreet, 2021), given the same visual stimuli as used in this study. Thus, given this visual context, descriptions using these adjective orders were highly unexpected, and therefore had high surprisal values. These high surprisal values were present in the HL and LL conditions, which exhibited higher rates of learning, in line with the findings of Lai et al. (2020). In other words, conditions including at least one unexpected order led to better learning than the condition including only expected orders. Thus, our findings suggest that surprisal-driven learning can occur when speakers differ in stylistic non-crucial language use.

#### 4.3. Implications for speaker-specificity in lexical entrainment

Previous research has shown indications that listeners can track, implicitly (Pickering & Garrod, 2004) or explicitly (Brennan & Clark, 1996; Hanna et al., 2003), speaker-specific stylistic preferences at the lexical level. As observed in studies on *lexical entrainment* (Brennan & Clark, 1996; Clark & Wilkes-Gibbs, 1986; Metzing & Brennan, 2003; Wilkes-Gibbs & Clark, 1992), interlocutors adopt the same terminology for entities in the conversation, thus establishing common ground for communication. Our study examined the learning process itself, allowing us to examine whether inter-speaker variation is actively stored, or whether this information is transient and occurs only within-task, to assist in fulfilling the task demands. Given that participants in our study successfully recognized the association between speakers and their language use following the exposure phase, it is reasonable to conclude that the language–speaker association is not solely the result of an automatic memory mechanism (Pickering & Garrod, 2004, 2006), but rather that individuals actively map – under some conditions – inter-speaker stylistic variability. Importantly, in our experiment, participants learned the association of language use with specific speakers without interactively forming conceptual pacts, meaning that language–speaker mapping is not necessarily the result of an interactive process by which interlocutors create shared representation jointly.

**Supplementary material.** The supplementary material for this article can be found at <http://doi.org/10.1017/langcog.2023.32>.

**Acknowledgments.** This work was funded by the Israeli Science Foundation (ISF 1824/17) and by the Alon fellowship. The authors would also like to thank Prof. Sarah Brown-Schmidt for her helpful comments on the data analysis and Shiri Hornick and Omri Kimchi-Feldhorn for contributing their voices for the auditory stimuli.

**Data availability.** Data and code are available on this link: <https://osf.io/sru8d/>.

**Competing interest.** The authors declare none.

## References

- Adler, J. E. (2008). Surprise. *Educational Theory*, 58(2), 149–173.
- Ahlmann-Eltze, C., & Patil, I. (2022). ggsignif: Significance Brackets for “ggplot2” (0.6.4). <https://CRAN.R-project.org/package=ggsignif>
- Bangerter, A., Mayor, E., & Knutsen, D. (2020). Lexical entrainment without conceptual pacts? Revisiting the matching task. *Journal of Memory and Language*, 114, 104129. <https://doi.org/10.1016/j.jml.2020.104129>
- Barr, D. J. (2004). Establishing conventional communication systems: Is common knowledge necessary? *Cognitive Science*, 28(6), 937–962. [https://doi.org/10.1207/s15516709cog2806\\_3](https://doi.org/10.1207/s15516709cog2806_3)
- Barr, D. J., & Keysar, B. (2005). Making sense of how we make sense: The paradox of egocentrism in language use. In Herbert L. Colston & Albert N. Katz (eds.), *Figurative language comprehension: Social and cultural influences* (pp. 21–43). Mahwah, N. J.: Lawrence Erlbaum.
- Ben-Shachar, M., Lüdtke, D., & Makowski, D. (2020). effectsize: Estimation of effect size indices and standardized parameters. *Journal of Open Source Software*, 5(56), 2815. <https://doi.org/10.21105/joss.02815>
- Bod, R., Hay, J., & Jannedy, S. (2003). *Probabilistic linguistics*. Cambridge, MA: MIT Press.
- Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22, 1482–1493. <https://doi.org/10.1037/0278-7393.22.6.1482>
- Brown-Schmidt, S. (2009). Partner-specific interpretation of maintained referential precedents during interactive dialog. *Journal of memory and language*, 61(2), 171–190.
- Bybee, J., & Hopper, P. (2001). *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins.
- Cinque, G. (1994). On the evidence for partial N-movement in the romance DP. In G. Cinque, J. Koster, J. Y. Pollock, L. Rizzi, & R. Zanuttini (Eds.), *Paths towards universal grammar* (pp. 85–110). Washington, DC: Georgetown University Press.
- Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, 22(1), 1–39.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: Routledge. <https://doi.org/10.4324/9780203771587>
- Creel, S. C., Aslin, R. N., & Tanenhaus, M. K. (2008). Heeding the voice of experience: The role of talker variation in lexical access. *Cognition*, 106(2), 633–664.
- Danks, J. H., & Glucksberg, S. (1971). Psychological scaling of adjective orders. *Journal of Verbal Learning and Verbal Behavior*, 10(1), 63–67. [https://doi.org/10.1016/S0022-5371\(71\)80095-6](https://doi.org/10.1016/S0022-5371(71)80095-6)
- Danks, J. H., & Schwenk, M. A. (1972). Prenominal adjective order and communication context. *Journal of Verbal Learning and Verbal Behavior*, 11(2), 183–187. [https://doi.org/10.1016/S0022-5371\(72\)80075-6](https://doi.org/10.1016/S0022-5371(72)80075-6)
- Danks, J. H., & Schwenk, M. A. (1974). Comprehension of prenominal adjective orders. *Memory & Cognition*, 2(1), 34–38. <https://doi.org/10.3758/BF03197488>
- Davies, C., Porretta, V., Koleva, K., & Klepousniotou, E. (2017). Do speaker-specific cues influence ambiguous word interpretation? In *Poster presented at the 23rd Architectures and Mechanisms of Language Processing Conference, Lancaster, UK*.
- Dixon, R. M. (1982). *Where have all the adjectives gone? and other essays in semantics and syntax*, vol. 107. Mouton, Berlin: de Gruyter.
- Eisner, F., & McQueen, J. M. (2005). The specificity of perceptual learning in speech processing. *Perception & Psychophysics*, 67(2), 224–238.
- Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24(2), 143–188.
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G\*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191.
- Foster, M. I., & Keane, M. T. (2019). The role of surprise in learning: Different surprising outcomes affect memorability differentially. *Topics in Cognitive Science*, 11(1), 75–87.
- Futrell, R., Gibson, E., & Levy, R. P. (2020). Lossy-context surprisal: An information-theoretic model of memory effects in sentence processing. *Cognitive Science*, 44(3), e12814. <https://doi.org/10.1111/cogs.12814>

- Garnsey, S. M., Pearlmutter, N. J., Myers, E., & Lotocky, M. A. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language*, 37(1), 58–93.
- Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, 27(2), 181–218.
- Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Speech Acts* (pp. 41–58). New York: Academic Press.
- Grodner, D., & Sedivy, J. (2011). The effect of speaker specific information on pragmatic inferences. In N. Pearlmutter and E. Gibson (Eds.), *The processing and acquisition of reference*. Cambridge, MA: MIT Press.
- Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics* (pp. 159–166). Pittsburgh, PA: ACM.
- Hanna, J. E., Tanenhaus, M. K., & Trueswell, J. C. (2003). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language*, 49(1), 43–61. [https://doi.org/10.1016/s0749-596x\(03\)00022-6](https://doi.org/10.1016/s0749-596x(03)00022-6)
- Holtgraves, T. M., & Barr, D. J. (2014). Perspective taking and its impostors in language use. In T. M. Holtgraves (Ed.), *The Oxford handbook of language and social psychology*. Oxford: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199838639.013.035>
- Jacewicz, E., Fox, R. A., & Wei, L. (2010). Between-speaker and within-speaker variation in speech tempo of American English. *Journal of the Acoustical Society of America*, 128(2), 839–850. <https://doi.org/10.1121/1.3459842>
- Kamide, Y. (2012). Learning individual talkers' structural preferences. *Cognition*, 124(1), 66–71.
- Keysar, B., Barr, D. J., & Horton, W. S. (1998). The Egocentric basis of language use: Insights from a processing approach. *Current Directions in Psychological Science*, 7(2), 46–49. <https://doi.org/10.1111/1467-8721.ep13175613>
- Khan, M., Hartshorne, J. K., & Snedeker, J. C. (2009). *Can listeners modulate rapid pragmatic inferences based on knowledge about the speaker? In Poster presented at 15th Annual conference on architectures and mechanisms for language processing, Barcelona, Spain.*
- Kraljic, T., & Samuel, A. G. (2007). Perceptual adjustments to multiple speakers. *Journal of Memory and Language*, 56(1), 1–15.
- Kroczek, L. O., & Gunter, T. C. (2017). Communicative predictions can overrule linguistic priors. *Scientific Reports*, 7(1), 1–9.
- Kronmüller, E., & Barr, D. J. (2007). Perspective-free pragmatics: Broken precedents and the recovery-from-preemption hypothesis☆. *Journal of Memory and Language*, 56(3), 436–455. <https://doi.org/10.1016/j.jml.2006.05.002>
- Kronmüller, E., & Guerra, E. (2020). Processing speaker-specific information in two stages during the interpretation of referential precedents. *Frontiers in Psychology*, 11, 552368. <https://doi.org/10.3389/fpsyg.2020.552368>
- Laenzlinger, C. (2005). French adjective ordering: perspectives on DP-internal movement types. *Lingua*, 115(5), 645–689. <https://doi.org/10.1016/j.lingua.2003.11.003>
- Lai, W., Rácz, P., & Roberts, G. (2020). Experience with a linguistic variant affects the acquisition of its sociolinguistic meaning: An alien-language-learning experiment. *Cognitive Science*, 44(4), e12832. <https://doi.org/10.1111/cogs.12832>
- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106, 1126–1177.
- Lieven, E. (2010). Input and first language acquisition: Evaluating the role of frequency. *Lingua*, 120(11), 2546–2556.
- MacDonald, M. C. (1994). Probabilistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes*, 9(2), 157–201.
- Makowski, D. (2018). The Psycho Package: An efficient and publishing-oriented workflow for psychological science. *Journal of Open Source Software*, 3(22), 470. R package, <https://github.com/neuropsychology/psycho.R>.
- Martin, J. E. (1969a). Semantic determinants of preferred adjective order. *Journal of Verbal Learning and Verbal Behavior*, 8(6), 697–704.

- Martin, J. E. (1969b). Some competence-process relationships in noun phrases with prenominal and postnominal adjectives. *Journal of Verbal Learning and Verbal Behavior*, 8(4), 471–480.
- Martin, J. E. (1970). Adjective order and juncture. *Journal of Verbal Learning and Verbal Behavior*, 9(4), 379–383. [https://doi.org/10.1016/S0022-5371\(70\)80076-7](https://doi.org/10.1016/S0022-5371(70)80076-7)
- Martin, J. E., & Ferb, T. E. (1973). Contextual factors in preferred adjective ordering. *Lingua*, 32(1), 75–81. [https://doi.org/10.1016/0024-3841\(73\)90005-3](https://doi.org/10.1016/0024-3841(73)90005-3)
- Martin, J. E., & Molfese, D. L. (1972). Preferred adjective ordering in very young children. *Journal of Memory and Language*, 11(3), 287–292.
- Maye, J., Aslin, R. N., & Tanenhaus, M. K. (2008). The weckud wetch of the wast: Lexical adaptation to a novel accent. *Cognitive Science*, 32(3), 543–562.
- McLennan, C. T., & Luce, P. A. (2005). Examining the time course of indexical specificity effects in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(2), 306.
- Metzing, C., & Brennan, S. E. (2003). When conceptual pacts are broken: Partner-specific effects on the comprehension of referring expressions. *Journal of Memory and Language*, 49(2), 201–213. [https://doi.org/10.1016/S0749-596X\(03\)00028-7](https://doi.org/10.1016/S0749-596X(03)00028-7)
- Munnich, E., & Ranney, M. A. (2019). Learning from surprise: Harnessing a metacognitive surprise signal to build and adapt belief networks. *Topics in Cognitive Science*, 11(1), 164–177.
- Orena, A. J., & White, K. S. (2015). I forget what that's called! children's online processing of disfluencies depends on speaker knowledge. *Child Development*, 86(6), 1701–1709.
- Ostrand, R., & Ferreira, V. S. (2019). Repeat after us: Syntactic alignment is not partner-specific. *Journal of Memory and Language*, 108, 104037. <https://doi.org/10.1016/j.jml.2019.104037>
- Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, 51(1), 195–203.
- Pickering, M. J., & Garrod, S. (2004). The interactive-alignment model: Developments and refinements. *Behavioral and Brain Sciences*, 27(2), 212–225. <https://doi.org/10.1017/S0140525X04450055>
- Pickering, M. J., & Garrod, S. (2006). Alignment as the basis for successful communication. *Research on Language and Computation*, 4(2), 203–228. <https://doi.org/10.1007/s11168-006-9004-0>
- Pogue, A., Kurumada, C., & Tanenhaus, M. K. (2016). Talker-specific generalization of pragmatic inferences based on under- and over-informative prenominal adjective use. *Frontiers in Psychology*, 6, 2035.
- R Core Team (2022). R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria*. <https://www.R-project.org/>.
- Rącz, P. (2013). *Saliency in sociolinguistics: A quantitative approach*. Berlin: De Gruyter. <https://doi.org/10.1515/9783110305395>.
- Redi, L., & Shattuck-Hufnagel, S. (2001). Variation in the realization of glottalization in normal speakers. *Journal of Phonetics*, 29(4), 407–429. <https://doi.org/10.1006/jpho.2001.0145>
- Rosales, C. M., & Scontras, G. (2019). On the role of conjunction in adjective ordering preferences. *Proceedings of the Linguistic Society of America*, 4(1), 32. <https://doi.org/10.3765/plsa.v4i1.4524>
- Schuster, S., & Degen, J. (2020). I know what you're probably going to say: Listener adaptation to variable use of uncertainty expressions. *Cognition*, 203, 104285.
- Scontras, G. (2023). Adjective ordering across languages. *Annual Review of Linguistics*, 9(1), 357–376. <https://doi.org/10.1146/annurev-linguistics-030521-041835>
- Scontras, G., Degen, J., & Goodman, N. D. (2017). Subjectivity predicts adjective ordering preferences. *Open Mind*, 1(1), 53–66.
- Scontras, G., Degen, J., & Goodman, N. D. (2019). On the grammatical source of adjective ordering preferences. *Semantics and Pragmatics*, 12(7), 1–21.
- Scott, G. J. (2002). Stacked adjectival modification and the structure of nominal phrases. *Functional Structure in DP and IP: The Cartography of Syntactic Structures*, 1, 91–120.
- Shetreet, E., & Novogrodsky, R. (2019). Differential patterns of children's knowledge of quantifier meaning revealed under different tasks. *Frontiers in Communication*, 4, 45.
- Shlonsky, U. (2004). The form of Semitic noun phrases. *Lingua*, 114(12), 1465–1526.
- Skoruppa, K., & Peperkamp, S. (2011). Adaptation to novel accents: Feature-based learning of context-sensitive phonological regularities. *Cognitive Science*, 35(2), 348–366.
- Sobel, D. M., Sedivy, J., Buchanan, D. W., & Hennessy, R. (2012). Speaker reliability in preschoolers' inferences about the meanings of novel words. *Journal of Child Language*, 39(1), 90–104.

- Sproat, R., & Shih, C. (1991). The cross-linguistic distribution of adjective ordering restrictions. In C. Georgopoulos, & R. Ishihara (Eds.), *Interdisciplinary approaches to language: Essays in honor of S. Y. Kuroda* (pp. 565–593). Dordrecht: Springer.
- Svenonius, P. (2008). The position of adjectives and other phrasal modifiers in the decomposition of DP. In L. McNally, C. Kennedy (Eds.), *Adjectives and adverbs: Syntax, semantics (and discourse)* (pp. 16–42). New York: Oxford University Press.
- Swets, J. A., Tanner Jr., W. P., & Birdsall, T. G. (1961). Decision processes in perception. *Psychological Review*, 68(5), 301–340.
- Trainin, N., & Shetreet, E. (2021). It's a dotted blue big star: On adjective ordering in a post-nominal language. *Language, Cognition and Neuroscience*, 36(3), 320–341.
- Upala, M. A., Gonce, L. O., Tweney, R. D., & Slone, D. J. (2007). Contextualizing counterintuitiveness: How context affects comprehension and memorability of counterintuitive concepts. *Cognitive Science*, 31(3), 415–439.
- Westbury, J. R., Hashi, M., & Lindstrom, M. J. (1998). Differences among speakers in lingual articulation for American English /ɹ/. *Speech Communication*, 26(3), 203–226. [https://doi.org/10.1016/S0167-6393\(98\)00058-2](https://doi.org/10.1016/S0167-6393(98)00058-2)
- Whorf, B. L. (1945). Grammatical Categories. *Language*, 21(1), 1–11. <https://doi.org/10.2307/410199>
- Wickham, H., Chang, W., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C., Woo, K., Yutani, H., Dunnington, D., & RStudio (2022). ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics (3.4.0). <https://CRAN.R-project.org/package=ggplot2>.
- Wilkes-Gibbs, D., & Clark, H. H. (1992). Coordinating beliefs in conversation. *Journal of Memory and Language*, 31(2), 183–194. [https://doi.org/10.1016/0749-596x\(92\)90010-u](https://doi.org/10.1016/0749-596x(92)90010-u)
- Yoon, S. O., Jin, K. S., Brown-Schmidt, S., & Fisher, C. L. (2021). What's new to you? Preschoolers' partner-specific online processing of disfluency. *Frontiers in psychology*, 11, 612601. <https://doi.org/10.3389/fpsyg.2020.612601>

---

**Cite this article:** Trainin, N., & Shetreet, E. (2024). Learning speaker-specific linguistic 'style' is mediated by deviance from common language use *Language and Cognition* 16: 197–219. <https://doi.org/10.1017/langcog.2023.32>