

RESEARCH ARTICLE



# Metacognition of frequency, directional association strength, and dispersion of MWEs in first and second language speakers

Yanlu Zhong<sup>1</sup> , Simon Todd<sup>1</sup>, Stefan Th. Gries<sup>1,2</sup>  and Laurel Brehm<sup>1</sup>

<sup>1</sup>Department of Linguistics, University of California, Santa Barbara, USA and <sup>2</sup>Department of English, Justus Liebig University Giessen, Giessen, Germany

**Corresponding author:** Yanlu Zhong; Email: [yanlu\\_zhong@ucsb.edu](mailto:yanlu_zhong@ucsb.edu)

(Received 06 February 2025; Revised 09 September 2025; Accepted 15 October 2025)

## Abstract

Statistical regularities can be acquired from usage. To examine language speakers' statistical metacognition about multiword expressions (MWEs), we collected ratings for frequency, dispersion, and directional association strength of English binomials from L1, advanced and intermediate L2 speakers. Mixed-effects modeling showed all speakers had limited speaker-to-corpus consistency but significant sensitivity to statistical regularities of language, supporting usage-based (Gries & Ellis, 2015) and statistical learning theories (Christiansen, 2019). Their statistical metacognition was also shaped by word-level cues, consistent with dual-route model (Carroll & Conklin, 2014). Despite similarities, frequency metacognition showed the strongest speaker-to-corpus consistency, while dispersion metacognition was the hardest to develop. Advanced L2 speakers showed the greatest speaker-to-corpus consistency and sensitivity, while lower-proficiency speakers relied more on word-level cues in metacognitive judgments, supporting the shallow-structure hypothesis (Clahsen & Felser, 2006). Overall, L1 and L2 speakers develop diverse statistical metacognition, with L2 speakers not necessarily inferior, suggesting that statistical metacognition is not solely shaped by usage-based experience.

**Keywords:** association strength; binomials; dispersion; frequency; L2 speakers; MWEs; statistical metacognition

## Introduction

Statistical regularities—the distributional patterns of linguistic units, which can be measured in various ways, including frequency, association strength, and dispersion—shape language processing (Ellis & Wulff, 2014; Goldberg, 1995; Gries & Ellis, 2015). Statistical learning theory (Christiansen, 2019) suggests that language speakers are highly attuned to statistical distributional patterns in language. Usage-based approaches propose

that these patterns extend from single words to multiword expressions (MWEs). Statistical learning is primarily implicit (Christiansen, 2019), as speakers unconsciously extract statistical regularities from input. Yet, growing evidence suggests an interface between implicit learning and explicit processing (Ellis, 2009). This raises a question: Do language speakers possess statistical metacognition (i.e., explicitly retrievable statistical intuition) that reflects conscious access to implicitly acquired statistical knowledge? We explore L1, advanced and intermediate L2 speakers' statistical metacognition of MWEs.

Previous studies have demonstrated that speakers possess statistical metacognition. In the mid-20th century, Tryk (1968) found L1 speakers' judgments largely aligned with the statistical properties of language. For MWEs, Siyanova-Chanturia and Schmitt (2008) reported that L1 speakers' frequency metacognition closely matched corpus data, while Yi *et al.* (2023) found L1 and L2 speakers' metacognition of frequency and association strength diverged from corpus statistics. Statistical learning also varies individually. From L1 to advanced and intermediate L2 learners, implicit experience tends to decline while reliance on explicit rules and cognitive effort increases (Ellis & Wulff, 2014)—likely leading to group differences in metacognition. However, few studies have examined the multifaceted nature of metacognition beyond frequency across L1, advanced L2, and intermediate L2 speakers of MWEs.

To bridge the gap, the present study investigates L1, advanced and intermediate L2 speakers' ratings of frequency, dispersion, forward association, and backward association for binomials (e.g., salt and pepper)—a key type of MWEs—and compares them to statistical regularities in Corpus of Contemporary American English (COCA; Davies, 2009). Statistical metacognition has been examined from three perspectives: (1) speaker-to-corpus consistency—the consistency between speakers' statistical metacognition and the distributional statistics of language, quantified by the degree of divergence between users' statistical judgments and the statistical regularities of language as measured in a representative corpus; (2) sensitivity—the degree to which language speakers' statistical judgments vary in response to differences in the statistical properties of language, reflecting the responsiveness of their mental representation to distributional patterns; and (3) other contributing factors—how phrasal-level statistical regularities beyond the target type being rated, word-level statistics, and word-form features contribute to language speakers' statistical metacognition.

## Literature review

### *MWE and their statistical regularities*

Language speakers accumulate linguistic constructions of varying granularity that pair form with meaning and function, such as words, MWEs, and sentences (Goldberg, 1995). MWEs are recurrent combinations of words that co-occur more often than would be expected by chance (Siyanova-Chanturia & Martinez, 2015; Yi & Zhong, 2024).<sup>1</sup> They account for nearly half of the occurrences in L1 use (Siyanova-Chanturia, 2015) and show

<sup>1</sup>We adopt a corpus-based definition of MWEs as recurrent combinations of words that co-occur more often than would be expected by chance (Siyanova-Chanturia & Martinez, 2015; Yi & Zhong, 2024), rather than a phraseological approach that treats them as fixed, non-decomposable vocabulary chunks holistically retrieved from the mental lexicon (Wray, 2002). This approach is supported by evidence that MWE components remain accessible (e.g., Boulenger *et al.*, 2012; Yi & Zhong, 2024), suggesting hierarchical processing of MWEs (Carroll & Conklin, 2014).

a processing advantage compared to novel word combinations (Siyanova-Chanturia et al., 2011).

MWE processing is closely tied to chunking processes driven by statistical learning mechanisms (Christiansen, 2019). MWEs are defined by statistical properties rather than specific syntactic or semantic features, making them a broad category that includes collocations (Wolter & Yamashita, 2018), idioms (Carroll & Conklin, 2014), binomials (Morgan & Levy, 2016; Siyanova-Chanturia et al., 2011), and lexical bundles (Tremblay et al., 2011). Among these, binomials have received less attention. Binomials are defined as recurrent combinations of words that occur more often than chance, joined by “and” or “or,” and exhibit a preferred word order (Siyanova-Chanturia et al., 2011). Their uniform structure and semantic simplicity (with opaque binomials excluded from this study) make them comparable across items, serving in this study as both the focus and a window into statistical metacognition across the broad class of MWEs.

MWEs, including binomials, exhibit statistical regularities that reflect the distributional patterns of linguistic units across usage contexts, such as frequency, association strength, and dispersion. Frequency reflects the likelihood of encountering linguistic units. It drives automatized processing (Gries & Ellis, 2015) and is central to language acquisition (Goldberg, 1995). Frequency effects are among the most robust findings in psycholinguistics (e.g., Öksüz et al., 2021; Siyanova-Chanturia et al., 2011). Nevertheless, frequency alone cannot capture the complexity of MWE statistical distribution. Dispersion instead captures how widely a linguistic unit is distributed across macro-level usage contexts (e.g., genres, registers, textual sections) in a language (Gries, 2024; Gries & Ellis, 2015). Contextual learning theory (Martin, 2014) suggests that language acquisition is shaped not only by repeated exposure to linguistic units but also involves acquiring varied macro-level communicative contexts. MWEs with identical frequency may differ in dispersion, with some context-specific and others broadly distributed. Key dispersion metrics include Juillard’s D, Carroll’s D2, and Kullback–Leibler Divergence (KLD, Gries, 2024). Beyond frequency and dispersion, association strength quantifies the statistical relationship between MWE components (Yi et al., 2023). Language speakers are able to track such patterns, as described by associative learning theory (Gries & Ellis, 2015). Pointwise mutual information (PMI), log Dice, and *t*-score capture bidirectional associations but lack directionality (Öksüz et al., 2021; Yi, 2018). Unidirectional measures—such as forward/backward transition probabilities, delta Ps, and KLD—capture directional dependencies (Gries & Ellis, 2015).

### **Statistical learning mechanism and statistical metacognition**

Language speakers possess robust, largely implicit statistical learning mechanisms without the involvement of explicit awareness for encoding distributional patterns (Christiansen, 2019). Language speakers have shown implicit sensitivity to word frequency (Öksüz et al., 2021), whereas evidence for dispersion effects is mixed: Gries (2024) found dispersion more predictive of reaction times than frequency, whereas Baayen et al. (2016) reported the opposite. Frequency effects also extend to MWEs, as evidenced in the processing of binomials (Conklin & Carroll, 2021; Siyanova-Chanturia et al., 2011; Sonbul et al., 2023), collocations (Öksüz et al., 2021), and lexical bundles (Tremblay et al., 2011). For instance, Siyanova-Chanturia et al. (2011) and Morgan and Levy (2016) showed that both L1 and L2 speakers read frequent conventional binomials more fluently than reversed ones. Conklin and Carroll (2021) found novel binomials gain processing advantages after limited exposure in L1 speakers, and Suhad et al. (2023)

showed similar effects for L2 learners. Recent studies also indicated that association strength, using metrics like PMI (Yi *et al.*, 2017) and log Dice (Öksüz *et al.*, 2021), moderated MWE processing.

Although statistical learning is primarily implicit, language speakers also develop a form of statistical metacognition—the explicit, conscious, and reflective evaluation of statistical regularities likely acquired intuitively (Siyanova-Chanturia & Spina, 2015; Tryk, 1968; Yi *et al.*, 2023). As suggested by implicit and explicit learning theory (Ellis, 2009), language learning involves both implicit automatized processes that operate without conscious awareness and explicit effortful processes that support metalinguistic reflection (Ellis, 2009). In this study, participants were engaged in explicit rating tasks that required them to consciously evaluate the statistical properties of binomials, in contrast to real-time language processing that typically draws on implicit routines (see Yi *et al.*, 2023, for a review). Such offline, reflective rating tasks are cognitively demanding and may elicit conscious reasoning over intuitively acquired impressions (Ellis, 2009). This contrasts with the real-time, automatic processing of statistical regularities without the involvement of metalinguistic knowledge (Siyanova-Chanturia & Schmitt, 2008; Siyanova-Chanturia & Spina, 2015).

### *Three aspects of statistical metacognition*

Statistical metacognition can be evaluated in terms of speaker-to-corpus consistency, sensitivity, and other contributing factors. Speaker-to-corpus consistency is measured by the absolute standardized difference between language speakers' judgments and the statistical properties of language. Smaller differences indicate greater speaker-to-corpus consistency and a more precise metalinguistic representation of statistical knowledge aligned with language. Sensitivity captures how strongly language speakers' statistical judgments systematically respond to variations in the distributional patterns of linguistic units, reflecting the strength of the statistical learning mechanism. Other contributing factors include phrasal-level statistical regularities beyond the target type being rated (e.g., association strength when rating frequency), as well as word-level statistics (e.g., word frequency and dispersion) and word-form features (e.g., word length, orthographic and phonological neighborhood).

L1 speakers' statistical metacognition of individual words has been examined, partially reflecting the three aspects of metacognition outlined above. Tryk (1968) reported high correlations (>0.7) between L1 word frequency ratings and corpus counts. Schmitt and Dunham (1999) used a relative estimation task where L1 speakers rated word frequency against a reference anchor, reporting generally good absolute frequency judgments of participants and moderate correlations with corpus data (~0.5). In contrast, MWE-level research remains limited. Siyanova-Chanturia and Schmitt (2008) found that L1 speakers could distinguish between high- and medium-frequency English collocations on a six-point scale. Later, Siyanova-Chanturia and Spina (2015) found similar results with Italian collocations rated on a four-point scale, with the clearest separation between highest- and very-low-frequency items. Beyond frequency, Yi *et al.* (2023) found that L1 speakers correctly distinguished collocations across frequency and association strength bins about 40% of the time on a three-point scale, with better differentiation for higher-statistics items. Additionally, low-to-moderate correlations have been reported between L1 speakers' frequency estimates and corpus data for Swedish three-word combinations (Backman, 1978), Italian (Siyanova-Chanturia & Spina, 2015), and English (Siyanova-Chanturia & Schmitt, 2008; Yi *et al.*, 2023) collocations.

Furthermore, metacognitive judgments about one type of statistical regularities (e.g., frequency) were found to be shaped by others (e.g., association strength), and vice versa (Yi et al., 2023). In addition, word-level statistics (e.g., word frequency) and word-form features (e.g., word length) may also shape phrasal statistical metacognition (Siyanova-Chanturia & Spina, 2015), as proposed in the dual-route model (Carrol & Conklin, 2014). For instance, word frequency and dispersion shape expectations about the distributional patterns (Gries & Ellis, 2015) and may serve as cues to phrasal probability. Word length and orthographical/phonological neighborhood—the number of orthographically/phonologically similar words differing by one letter/phoneme (Marian et al., 2012)—reflect surface-level complexity and lexical competition. Larger neighborhoods potentially increase cognitive load in visual and auditory word recognition (Yates et al., 2005). These word-level factors, while often overlooked in MWE studies, may influence statistical metacognition.

### *Differences in statistical metacognition among L1, advanced, and intermediate L2 speakers*

L1 speakers and L2 learners—including advanced learners immersed in target-language environments and intermediate learners in foreign-language contexts—may differ in their statistical learning mechanisms, as reflected in empirical findings. Some studies reported stronger MWE frequency effects for L1 speakers (Siyanova-Chanturia et al., 2011), others for L2 speakers (Yi et al., 2018), and some reported no difference (Öksüz et al., 2021) in online processing of MWEs. For metacognition, Schmitt and Dunham (1999) found L2 speakers' frequency ratings correlated more strongly with corpus data than L1 speakers, while others reported stronger rating-corpus correlations for medium-, high-, and very low-frequency collocations (Siyanova-Chanturia & Schmitt, 2008; Siyanova-Chanturia & Spina, 2015) and association strength (Yi et al., 2023) in L1 speakers.

Differences in statistical metacognition across L1 speakers, advanced L2 learners, and intermediate L2 learners can be explained by differences in acquisition modes (implicit vs. explicit), learning contexts, cognitive resources, developmental trajectories, and processing strategies. First, usage-based (Ellis & Wulff, 2014) and explicit-implicit learning theories (Ellis, 2005, 2009) suggest L1 speakers acquire language implicitly through rich exposure, while intermediate L2 learners rely largely on explicit, rule-based classroom learning. Advanced L2 learners may benefit from a more effective combination of richer statistical knowledge acquired through natural exposure and enhanced metalinguistic reflection fostered by explicit instruction, probably leading to stronger statistical metacognition in rating tasks. Second, contextual learning theory (Martin, 2014) emphasizes the importance of communicative contexts in language development. L1 speakers gain context-rich MWE input through early, varied natural exposure (Adolphs & Durow, 2004). Advanced L2 learners in English-speaking environments, though less immersed than L1 speakers, receive more authentic input than intermediate learners, whose exposure is largely limited to constrained discourse (Batstone, 2002). Third, cognitive cost shapes processing strategies. Under the shallow-structure hypothesis (Clahsen & Felser, 2006), intermediate L2 learners rely on surface-level cues (e.g., word length) due to higher demands, whereas advanced L2 learners and L1 speakers engage in deeper, usage-based processing with less effort. Fourth, developmental timing matters. According to the power law of practice (DeKeyser, 2020), early-stage learning yields greater marginal gains, as each input instance makes up a larger share of total language experience. As a result, L1 speakers—whose extensive

language experience overshadows statistical variation—may show weaker metacognitive sensitivity to statistical regularities than advanced and intermediate L2 learners. Finally, the dual-route model (Carroll & Conklin, 2014) suggests that L1 speakers can process MWEs more holistically, enhancing fluency and reducing cognitive load. Advanced L2 learners may also rely less on word-level processing, especially with extensive natural exposure to MWEs. In contrast, intermediate learners tend to use analytical strategies and focus on constituent words.

Together, these theoretical frameworks suggest that statistical metacognition differs across L1 speakers, advanced L2 learners, and intermediate L2 learners.

## The current study

To summarize, research has largely focused on implicit statistical knowledge, with limited attention to statistical metacognition. In addition, most studies examined individual words, leaving it unclear whether the statistical metacognition extends to binomials—an important MWE type. Third, previous research has not systematically analyzed statistical metacognition in terms of speaker-to-corpus consistency, sensitivity, and contributing factors. Fourth, prior work focused on frequency (Siyanova-Chanturia & Spina, 2015) and undirected association strength (e.g., PMI, Yi *et al.*, 2023), neglecting directional association strength and dispersion. Furthermore, prior studies have relied on relatively coarse rating scales, such as three-point (Yi *et al.*, 2023), four-point (Siyanova-Chanturia & Spina, 2015), and six-point (Siyanova-Chanturia & Schmitt, 2008), limiting the capture of finer-grained, continuous variation in statistical metacognition. Finally, limited studies compare statistical metacognition across proficiency levels, leaving the roles of language experience in statistical metacognition unclear.

To bridge the gaps, we proposed three research questions (RQs):

**RQ<sub>1</sub>:** Speaker-to-corpus consistency of statistical metacognition: (a) To what extent are language speakers' judgments close to the corpus-based statistical regularities of binomials? (b) How do language speaker group (L1 speakers, advanced L2 learners, and intermediate L2 learners), type of statistical regularities (frequency, forward association, backward association, dispersion), and their interactions affect the speaker-to-corpus consistency of statistical metacognition?

**RQ<sub>2</sub>:** Sensitivity of statistical metacognition: (a) To what extent do language speakers' judgments respond to variations in the corpus-based statistical regularities of binomials? (b) Does the sensitivity vary by language speaker group and type of statistical regularities?

**RQ<sub>3</sub>:** Other contributing factors: (a) To what extent do phrasal-level statistical regularities beyond the target type, as well as word-level statistics (e.g., word frequency and dispersion) and word-form features (e.g., word length, orthographic and phonological neighborhood) influence language speakers' statistical metacognition? (b) Do these factors vary in their influence by language speaker group and type of statistical regularities?

## Methodology

### Participants

We recruited 94 participants via Prolific: 34 L1 English speakers (19 males), 30 advanced L2 learners (14 males), and 30 intermediate L2 learners (19 males), all aged 18–35 with a bachelor's degree. Participants were paid \$16/hour.



L1 speakers acquired English from birth. L2 speakers had a non-English L1, began learning English after age three, and received formal instruction. All L1 speakers and advanced L2 learners were residing in the U.S. during participation. To distinguish L2 proficiency levels, extended immersion in an English-speaking environment served as the primary criterion. Advanced L2 learners met the following criteria: (1) length of residence in English-speaking countries  $\geq 8$  years (required), plus at least two of: (2)  $\geq 50$  hours/week English use, (3) age of acquisition  $\leq 10$ , and (4) self-rated proficiency  $\geq 8$  on a 9-point Likert scale across all language skills. Intermediate L2 learners had  $\leq 2$  years of residence (required) and met no more than two of the other criteria.

Mean ages were 28.5 ( $SD = 6.0$ ) for L1 speakers, 28.9 ( $SD = 4.6$ ) for advanced L2 learners, and 27.0 ( $SD = 4.1$ ) for intermediate L2 learners. Advanced L2 learners began acquiring English at 7.3 years old ( $SD = 3.2$ ), lived in the U.S. for 219.5 months ( $SD = 78.0$ ), and used English 82.8 hours/week ( $SD = 35.5$ ) over 212.9 months ( $SD = 212.6$ ), through both formal instruction and daily use. Intermediate L2 learners began acquiring English at 8.2 years old ( $SD = 3.3$ ), had little to no residence in English-speaking environments ( $M = 2.7$  months,  $SD = 6.1$ ), and used English 24.5 hours/week ( $SD = 23.0$ ) over 75.7 months ( $SD = 79.0$ ), primarily in classroom settings. On a 9-point self-assessment scale, advanced L2 learners rated higher in overall proficiency (8.5 vs. 7.2), reading (8.4 vs. 7.9), listening (8.6 vs. 7.7), speaking (8.5 vs. 6.6), and writing (8.1 vs. 6.9; see Table 1).

### Stimuli

As previously noted, binomials are defined as “and”- or “or”-linked recurrent combinations of words that co-occur more often than chance and show a preferred order. We focused on transparent, noun-noun binomials joined by “and” for their structural regularity and semantic simplicity, which allow better control and measurement of statistical properties compared to structurally variable (e.g., lexical bundles) or opaque (e.g., idioms) MWEs. Using regular expressions, we extracted 28,967 “Noun + and + Noun” pairs from COCA (1990–2019; ~950M words, <https://www.corpusdata.org>), a widely used, thematically diverse corpus balanced across eight registers (e.g., spoken, fiction). COCA is considered a reliable proxy for large-scale, broad, balanced English usage patterns, and its pre-labeled text units enable precise dispersion computation. We

**Table 1.** Demographic characteristics of L2 learners

	Advanced L2 speakers		Intermediate L2 speakers	
	Mean	SD	Mean	SD
Age	28.9	4.6	27.0	4.1
Age of acquisition	7.3	3.2	8.2	3.3
Length of residence	219.5	78.0	2.7	6.1
Hours of usage per week	82.8	35.5	24.5	23.0
Length of such rate (hours per week) using English	212.9	212.6	75.7	79.0
Overall proficiency	8.5	0.8	7.2	1.3
Speaking proficiency	8.5	0.8	6.6	1.7
Listening proficiency	8.6	0.7	7.7	1.2
Reading proficiency	8.4	1.0	7.9	1.0
Writing proficiency	8.1	1.1	6.9	1.4

Note: We did not report English test scores, as only 14 L2 learners had taken an English test, with varied types (e.g., IELTS, TOEFL, SAT).

**Table 2.** Contingency table

	Co-occurring element X	Other elements (not X)	Row totals
Element E	a	b	a + b
Other elements (not E)	c	d	c + d
Column totals	a + c	b + d	a + b + c + b + d = n

Note: Cell *a* = E occurs with X; *b* = E occurs without X; *c* = X occurs without E; *d* = neither E nor X occurs.

retained binomials with above-chance co-occurrence (*t*-score > 2)<sup>2</sup> and more than 10 occurrences, yielding 4,813 items. We then removed abbreviations (e.g., *CEO*), redundancies (e.g., *hour and hour*), proper nouns (e.g., *Israeli and Arab*), and semantically opaque items (e.g., *spick and span*). For binomials with both higher-frequency conventional (e.g., *sun and moon*) and lower-frequency reversed (e.g., *moon and sun*) forms, only conventional forms were kept. To ensure consistency, we retained only those with dominant singular forms, resulting in 2,054 binomials.

To enhance representativeness, we grouped all pre-labeled text units in COCA into 2,700 balanced sections. We then applied bootstrap resampling to extract 2,700 sections 1,000 times, generating 1,000 corpora matching the original COCA’s size. For each corpus, we calculated binomial frequency (per million words, log-transformed), forward and backward association strength, and dispersion, then averaged these values across all 1,000 corpora. KLD was used for association strength as it captures directionality and is less correlated with frequency than some of traditional association metrics (e.g., *p*<sub>Fisher-Yates</sub>, *z*; Gries, 2024). Table 2 shows the contingency table (Gries, 2013). Equations (1) and (2) compute forward and backward association. Higher values indicate stronger associations. For the bigram “word1 word2,” forward association quantifies how word2’s distribution following word1 diverges from its overall distribution. Backward association does the same for word1 given word2.

$$D_{KL}(p(X|E)||p(X)) = \frac{a}{a+b} \times \log_2 \frac{a \times n}{(a+b) \times (a+c)} + \frac{b}{a+b} \times \log_2 \frac{b \times n}{(a+b) \times (b+d)} \tag{1}$$

$$D_{KL}(p(E|X)||p(E)) = \frac{a}{a+c} \times \log_2 \frac{a \times n}{(a+c) \times (a+b)} + \frac{c}{a+c} \times \log_2 \frac{c \times n}{(a+c) \times (c+d)} \tag{2}$$

<sup>2</sup>As a subtype of MWEs, binomials were selected based on general corpus-based MWE criteria. The association measure *t*-score, which reduces the impact of low-frequency items, was used to assess whether the constituent words of a binomial co-occur more often than expected by chance. The formula is as follows: *f* (binomial), *f*(*n*<sub>1</sub>), *f*(*n*<sub>2</sub>), and *f*(and) represent the raw corpus frequencies of the binomial, the two nouns, and the conjunction *and*, respectively, and *N* denotes the total number of tokens in the corpus.

$$t - score = \frac{\text{Observed Frequency} - \text{Expected Frequency}}{\sqrt{\text{Observed Frequency}}} = \frac{f(\text{binomial}) - \frac{f(n_1) * f(\text{and}) * f(n_2) * N}{N^3}}{\sqrt{f(\text{binomial})}}$$

We retained binomials with *t*-scores greater than 2 and a frequency greater than 10 in COCA (see Appendix 1 for specific values).



Dispersion was calculated using KLD across COCA's 2,700 predefined sections, each representing a document varying in genre, topic, or register. This segmentation helps explore contextual learning theory (Martin, 2014), which emphasizes the role of micro-level contextual exposure in language acquisition. KLD was chosen to ensure comparability with the KLD-based association strength. Equation (3) shows the dispersion of unit  $a$ , where  $i$  is the number of corpus parts,  $v$  is its frequency in each part,  $f$  is the total frequency, and  $s$  is the percentage of each part size.

$$KLD = \sum_{i=1}^n \frac{v_i}{f} \times \log_2 \left( \frac{v_i}{f} \times \frac{1}{s_i} \right) \quad (3)$$

To balance binomials across low and high bands of frequency, association strength, and dispersion, we split them into two bins per dimension, forming 16 groups (see Appendix 1 for the detailed binomial profiles within each statistical group). From each, 15–20 binomials were selected. Familiarity of each binomial was rated through an online questionnaire by two L1 speakers and five intermediate L2 learners (not in the main study) on a 5-point Likert scale (1 = completely unfamiliar, 5 = very familiar). Binomials were retained only if they received an average rating above 4 from L1 speakers and above 3 from L2 speakers. We also ensured no repeated words across binomials. In total, 300 binomials were selected as stimuli.

### Procedure

We created questionnaires (see Appendix 2) on Qualtrics and distributed them via Prolific to collect ratings of statistical regularities. Each included demographic questions and four statistical rating blocks, with block order balanced via a Latin square design and binomial order randomized within each block. Detailed instructions and examples were provided for each block. For frequency, participants rated how often a binomial occurred in their language use (1 = a few times, 9 = many times). For dispersion, they rated the number of different usage contexts in which they encountered a binomial (1 = few contexts, 9 = many contexts). For forward association, they assessed how easy it was to predict the second word from the first (1 = very difficult, 9 = very easy). For backward association, they rated how easy it was to predict the first word from the second (1 = very difficult, 9 = very easy). Participants rated all 300 binomials on a 9-point slider (1-decimal precision) across four blocks of statistical regularities. The questionnaire had no time limit and took approximately 100 minutes to complete. We also collected word-level information of binomials. Word frequency (log-transformed per million) and dispersion were extracted from COCA. Word-form features (word length, orthographic and phonological neighborhood) came from CLEARPOND (Marian et al., 2012, see Appendix 1 for rating data, binomial list, and all word-level information).

### Data analyses

We collected data from 34 L1, 30 advanced L2, and 30 intermediate L2 speakers, yielding 28,200 data points. One L1 speaker was excluded for giving identical ratings to over 80% of items. For each binomial under each statistical measure, we excluded ratings beyond 2.5 standard deviations within each language speaker group (458 outliers from L1

speakers, 414 from advanced L2 learners, and 531 from intermediate L2 learners). We then standardized ratings within each participant and corpus data across binomials for each statistical measure. Other phrase- and word-level continuous predictors were also standardized.

To address  $RQ_1$  (speaker-to-corpus consistency), we constructed a linear mixed-effects model with type of statistical regularities (frequency, dispersion, forward and backward association) and speaker group (L1, advanced L2, intermediate L2) as fixed effects. Both variables were treated as categorical and treatment-coded, with their interaction also included as predictors. The dependent variable was the absolute difference between standardized corpus values and standardized ratings (see [Appendix 3](#) for model  $RQ_1$ - $RQ_3$  formulas).

To address  $RQ_2$  (sensitivity) and  $RQ_3$  (other contributing factors), four linear mixed-effects models were built, each with one type of statistical rating (frequency, dispersion, forward and backward association) as the dependent variable. For  $RQ_2$ , one type of statistical regularity from COCA, corresponding to the type of statistical rating used as the dependent variable, was employed as the independent variable. The COCA-based statistical regularities were divided into tertiles (low, medium, and high statistical bins) based on ranked values, with equal item counts per bin, to capture non-linear statistical effects (e.g., thresholds, plateaus) on metacognition observed in prior studies (Yi *et al.*, 2023; Yi & Zhong, 2024). For  $RQ_3$ , additional standardized continuous variables, including other types of phrase-level statistical regularities that are distinct from the dependent variable, word-level statistics (e.g., word frequency and dispersion), and word-form features (e.g., word length, orthographic/phonological neighborhood), were included as independent variables.

Correlation coefficients for all variables were calculated. Four pairs showed high correlation: word1 (i.e., left conjoint) frequency and dispersion ( $r = 0.87$ ), word2 (i.e., right conjoint) frequency and dispersion ( $r = 0.84$ ), word1 orthographic and phonological neighborhoods ( $r = 0.93$ ), and word2 orthographic and phonological neighborhoods ( $r = 0.93$ ). Variance inflation factors for the four variable sets ranged from 4 to 7. To mitigate multicollinearity and model complexity, we applied group-wise Principal Component Analysis (PCA, Tomaschek *et al.*, 2018) to each variable set. PCA combined word1/word2 orthographic and phonological neighborhood into word1/word2 neighborhood ( $PCA_1 = 0.94$  for word1, 0.92 for word2), reflecting lexical neighborhood diversity. Word1/word2 frequency and dispersion were merged into word1/word2 statistics ( $PCA_1 = 0.97$  for both word1 and word2), capturing the likelihood of encountering a linguistic unit while correcting for frequency- or dispersion-only bias.

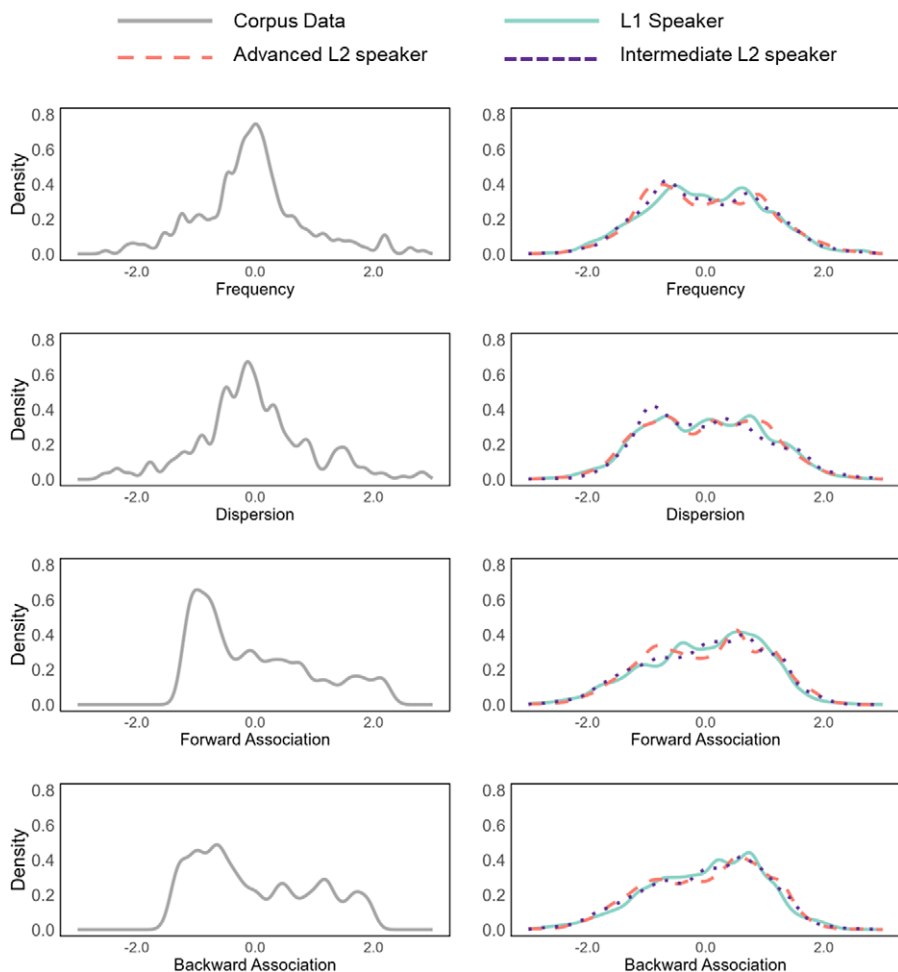
To compare statistical metacognition across speaker groups, language proficiency was treatment-coded into three levels (L1, advanced L2, and intermediate L2), and its interactions with other predictors were included in the models in  $RQ_2$  and  $RQ_3$ . Participants were included as a random effect to account for individual differences.

Model diagnostics showed no significant concerns with convergence, residuals, homoscedasticity, multicollinearity, or outliers (see [Appendix 4](#)). All models were built in R (version 4.5.0; R Core Team, 2025) with lme4 (version 1.1-34; Bates *et al.*, 2015).

## Results

### *Descriptive data analysis*

[Figure 1](#) compares standardized ratings from L1, advanced L2, and intermediate L2 speakers with COCA data across four types of statistical regularities. All groups used the



**Figure 1.** Distributions of standardized COCA data and standardized ratings. The y-axis shows kernel density estimates—smoothed, standardized curves with an area of 1 under each.

full Likert scale, and both ratings and corpus data showed moderate values occurring more frequently than extremes. COCA distributions were highly peaked (Kurtosis > 14) and right-skewed (Skewness > 0), whereas speakers' ratings were more balanced ( $2 < \text{Kurtosis} < 4$ ) and nearly symmetrical ( $-0.5 < \text{Skewness} < 0.1$ , see [Appendix 5](#) for detailed values), indicating fewer low-value binomials in frequency, dispersion, forward and backward association ratings.

We conducted a preliminary analysis comparing participant ratings with COCA data. Binomials were separately ranked by ratings and COCA values and divided into tertiles (low, medium, and high statistical bins) to assess the proportion falling into the same tertile across both groupings. Results showed low-to-moderate match proportions (<50%) between human ratings and COCA data. Ratings were also standardized to compute correlations with standardized COCA data, revealing small-to-moderate significant correlations (<0.5, see [Table 3](#); also see [Appendix 5](#) for binomial distributions

**Table 3.** The proportion of binomials in the same tertile across ratings and COCA data, and their correlation coefficients

	Frequency			Dispersion		
	L1	Advanced L2	Intermediate L2	L1	Advanced L2	Intermediate L2
Proportion	0.407	0.427	0.412	0.358	0.364	0.361
Correlation coefficient	0.252**	0.331**	0.286**	0.094**	0.138**	0.134**

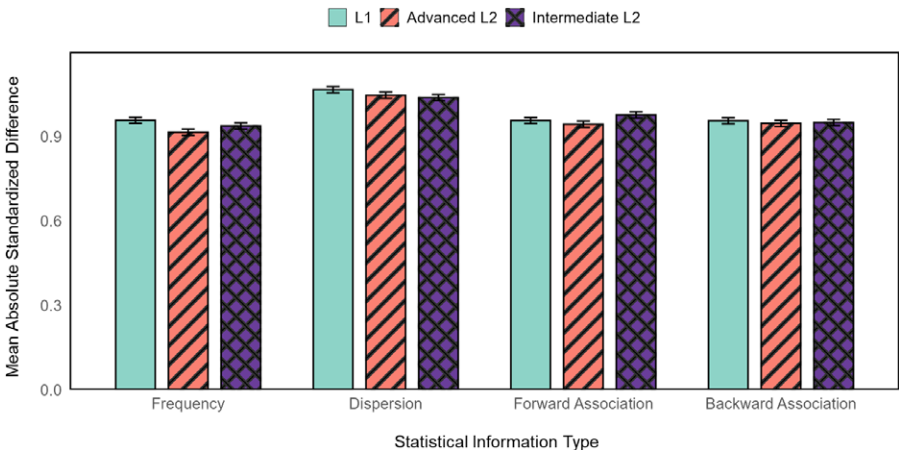
	Forward association			Backward association		
	L1	Advanced L2	Intermediate L2	L1	Advanced L2	Intermediate L2
Proportion	0.385	0.414	0.391	0.408	0.419	0.405
Correlation coefficient	0.177**	0.230**	0.170**	0.159**	0.183**	0.177**

Note: \*\*  $p < 0.01$ .

across COCA data and ratings). The low-to-moderate match proportions and correlations provide preliminary evidence for differences between language speakers’ statistical metacognition and corpus-based patterns.

*Speaker-to-corpus consistency*

To examine speaker-to-corpus consistency across four types of statistical regularities, we used the absolute difference between standardized COCA data and standardized participant ratings as the dependent variables, with type of statistical regularities, speaker group, and their interactions as predictors (see Figure 2 and Table 4). Dispersion ratings diverged more from corpus values than other statistical regularities, and advanced L2 learners showed smaller discrepancies than other speaker groups. Re-leveled results for all reference groups are in Appendix 6.



**Figure 2.** The mean absolute differences between standardized ratings and standardized COCA data across four types of statistical regularities and speaker groups, with error bars representing standard error.

**Table 4.** Results of mixed effects models for the speaker-to-corpus consistency of statistical metacognition

	Estimate	Std	<i>t</i> value	<i>p</i> value
(Intercept)	0.957	0.011	87.983	<.001
Type (ref = Frequency): Dispersion	0.109	0.011	9.517	<.001
Type (ref = Frequency): Forward	−0.001	0.011	−0.101	.919
Type (ref = Frequency): Backward	−0.002	0.011	−0.215	.830
Proficiency (ref = L1): Advanced L2	−0.042	0.016	−2.635	.009
Proficiency (ref = L1): Intermediate L1	−0.020	0.016	−1.268	.206
Type (ref = Frequency): Dispersion × Proficiency (ref = L1): Advanced L2	0.022	0.017	1.295	.195
Type (ref = Frequency): Forward × Proficiency (ref = L1): Advanced L2	0.028	0.017	1.707	.088
Type (ref = Frequency): Backward × Proficiency (ref = L1): Advanced L2	0.033	0.017	1.965	.050
Type (ref = Frequency): Dispersion × Proficiency (ref = L1): Intermediate L2	−0.008	0.017	−0.487	.626
Type (ref = Frequency): Forward × Proficiency (ref = L1): Intermediate L2	0.040	0.017	2.394	.017
Type (ref = Frequency): Backward × Proficiency (ref = L1): Intermediate L2	0.015	0.017	0.890	.374
Marginal R <sup>2</sup> : 0.0784; Conditional R <sup>2</sup> : 0.1053				

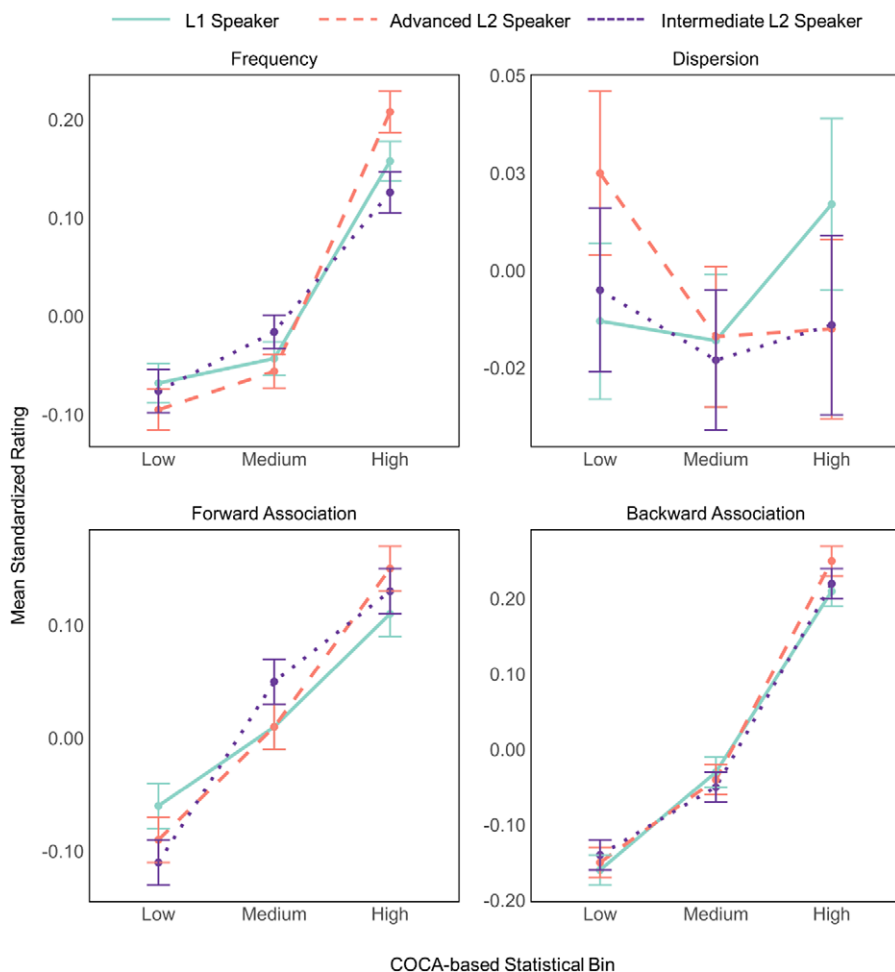
Model intercepts exceeded 0.7 standard deviations across all types of statistical regularities and language speaker groups, indicating large rating-corpus differences and generally low-to-moderate speaker-to-corpus consistency. Advanced L2 learners showed higher speaker-to-corpus consistency than L1 speakers for frequency ( $b = -0.042$ ,  $p = .009$ ) and than intermediate L2 learners for forward association ( $b = -0.033$ ,  $p = .042$ ) ratings. Across statistical measures, dispersion ratings were the least consistent with corpus data for all groups (all  $ps < .001$ ), while frequency ratings showed the highest speaker-to-corpus consistency in advanced L2 learners (all  $ps < .050$ ). For intermediate L2 learners, forward association ratings also showed lower speaker-to-corpus consistency than backward association and frequency (all  $ps < .050$ ). Interaction analyses showed that intermediate and advanced L2 learners had greater gains in speaker-to-corpus consistency for frequency ratings relative to forward ( $p = .017$ ) and backward association ( $p = .050$ ), respectively, compared to L1 speakers. Additionally, intermediate L2 learners had smaller gains in speaker-to-corpus consistency for forward association ratings relative to dispersion than both L1 speakers ( $p = .004$ ) and advanced L2 learners ( $p = .016$ ).

Fixed effects explained 7.8% of the variance, increasing to 10.5% with random intercepts, suggesting contributions from both fixed effects and individual variation.

### Sensitivity

Figure 3 shows that all speaker groups exhibited positive sensitivity to the statistical regularities of binomials, except for dispersion, which displayed weaker and less consistent patterns with larger error bars.

As shown in Table 5, all speaker groups rated high-frequency binomials higher than low- and mid-frequency ones (all  $ps < .001$ ), and intermediate L2 speakers also rated mid-frequency binomials higher than low-frequency ones ( $b = -0.060$ ,  $p = .025$ ).



**Figure 3.** Mean standardized ratings across COCA-based bins (low, medium, high) for four types of statistical regularities and language speaker groups. Error bars represent standard error.

Interaction analyses showed that advanced L2 learners had larger rating increases for high-frequency binomials over mid- ( $p = .001$ ) and low-frequency ones ( $p = .003$ ) than intermediate L2 learners. All language speakers showed positive sensitivity in their ratings to forward and backward association across all association strength levels (all  $p$ s  $< .050$ ). Interaction analyses indicated that intermediate L2 learners showed greater rating increases for medium-over low-forward association binomials compared to L1 speakers ( $p = .019$ ).

### *Influence of other contributing factors*

Figure 4 shows that phrase-level statistical regularities facilitated ratings, word-level statistics aided frequency and dispersion ratings, and word-form features generally had inhibitory effects.

**Table 5.** Mixed-effects model results for metacognition (i.e., ratings) of frequency, dispersion, forward, and backward association.

	Dependent variable				Dependent variable				Dependent variable				Dependent variable			
	Frequency				Dispersion				Forward association				Backward association			
	b	Std	t	p	b	Std	t	p	b	Std	t	p	b	Std	t	p
(Intercept)	−0.07	.02	−3.33	<.01	−0.01	.02	−0.68	.50	−0.06	.02	−2.63	.01	−0.16	.02	−7.44	<.01
statistical bin (ref = low)	0.03	.03	0.98	.33	0.00	.03	−0.18	.86	0.07	.03	2.58	.01	0.14	.03	5.13	<.01
medium																
statistical bin (ref = low)	0.23	.03	7.13	<.01	0.03	.03	0.95	.34	0.17	.04	4.50	<.01	0.37	.04	10.00	<.01
high																
proficiency (ref = L1)	−0.03	.03	−0.91	.36	0.04	.03	1.36	.17	−0.03	.03	−0.91	.36	0.01	.03	0.39	.69
Advanced L2																
proficiency (ref = L1)	−0.01	.03	−0.27	.79	0.01	.03	0.31	.76	−0.06	.03	−1.76	.08	0.03	.03	0.89	.37
Intermediate L2																
frequency_coca	/	/	/	/	0.08	.02	4.86	<.01	0.18	.02	9.14	<.01	0.17	.02	8.58	<.01
dispersion_coca	0.05	.01	3.50	<.01	/	/	/	/	−0.02	.02	−1.08	.28	−0.05	.02	−3.19	<.01
forward_coca	0.06	.02	3.86	<.01	0.03	.02	1.82	.07	/	/	/	/	0.06	.01	4.32	<.01
backward_coca	0.03	.02	1.74	.08	−0.03	.02	−1.62	.10	0.09	.02	4.81	<.01	/	/	/	/
word1 length	−0.14	.02	−8.36	<.01	−0.05	.02	−3.04	<.01	−0.13	.02	−7.66	<.01	−0.08	.02	−4.97	<.01
word1 neighborhood	−0.08	.01	−7.66	<.01	−0.05	.01	−4.28	<.01	−0.06	.01	−6.00	<.01	−0.04	.01	−4.08	<.01
word2 neighborhood	−0.02	.01	−2.30	.02	−0.04	.01	−4.27	<.01	−0.03	.01	−2.93	<.01	−0.02	.01	−2.26	.02
word2 length	−0.04	.01	−2.40	.02	0.03	.02	2.20	.03	−0.07	.01	−4.75	<.01	−0.08	.01	−5.84	<.01
word1 statistics	0.04	.01	3.79	<.01	0.05	.01	4.75	<.01	−0.01	.01	−0.78	.43	−0.01	.01	−1.15	.25
word2 statistics	0.02	.01	1.30	.20	0.07	.01	5.58	<.01	−0.01	.01	−1.13	.26	0.03	.01	2.56	.01
statistical bin (ref = low)	0.01	.04	0.38	.70	−0.04	.04	−1.01	.31	0.03	.04	0.73	.47	−0.02	.04	−0.58	.56
medium: proficiency																
(ref = L1)																
Advanced L2																
statistical bin (ref = low)	0.08	.05	1.68	.09	−0.07	.05	−1.52	.13	0.06	.06	1.10	.27	0.03	.05	0.49	.62
high: proficiency (ref = L1)																
Advanced L2																
statistical bin (ref = low)	0.04	.04	0.95	.34	−0.01	.04	−0.36	.72	0.09	.04	2.35	.02	−0.05	.04	−1.36	.17
medium: proficiency																
(ref = L1) Intermediate L2																

(Continued)



Table 5. (Continued)

	Dependent variable				Dependent variable				Dependent variable				Dependent variable			
	Frequency				Dispersion				Forward association				Backward association			
	b	Std	t	p	b	Std	t	p	b	Std	t	p	b	Std	t	p
statistical bin (ref = low) high: proficiency (ref = L1) Intermediate L2	−0.02	.05	−0.52	.60	−0.04	.05	−0.85	.40	0.07	.06	1.26	.21	−0.01	.05	−0.23	.82
proficiency (ref = L1) Advanced L2: frequency_coca	/	/	/	/	0.08	.02	3.07	<.01	0.04	.03	1.54	.12	0.03	.03	1.02	.31
proficiency (ref = L1) Intermediate L2: frequency_coca	/	/	/	/	0.04	.02	1.49	.14	−0.03	.03	−0.92	.36	0.00	.03	0.13	.90
proficiency (ref = L1) Intermediate L2: dispersion_coca	0.06	.02	2.82	<.01	/	/	/	/	0.03	.02	1.14	.25	0.04	.02	1.73	.08
proficiency (ref = L1) Advanced L2: forward_coca	0.01	.02	0.24	.81	0.00	.02	−0.15	.88	/	/	/	/	0.01	.02	0.34	.73
proficiency (ref = L1) Intermediate L2: forward_coca	−0.04	.02	−1.88	.06	−0.04	.02	−1.84	.07	/	/	/	/	0.01	.02	0.40	.69
proficiency (ref = L1) Advanced L2: backward_coca	0.01	.03	0.26	.79	−0.01	.03	−0.26	.79	0.03	.03	1.28	.20	/	/	/	/
proficiency (ref = L1) Intermediate L2: backward_coca	−0.01	.03	−0.40	.69	0.03	.03	1.13	.26	0.05	.03	1.85	.06	/	/	/	/
proficiency (ref = L1) Advanced L2: word1 length	−0.06	.02	−2.32	.02	−0.04	.02	−1.54	.12	−0.05	.02	−1.95	.05	−0.08	.02	−3.38	<.01

(Continued)

Table 5. (Continued)

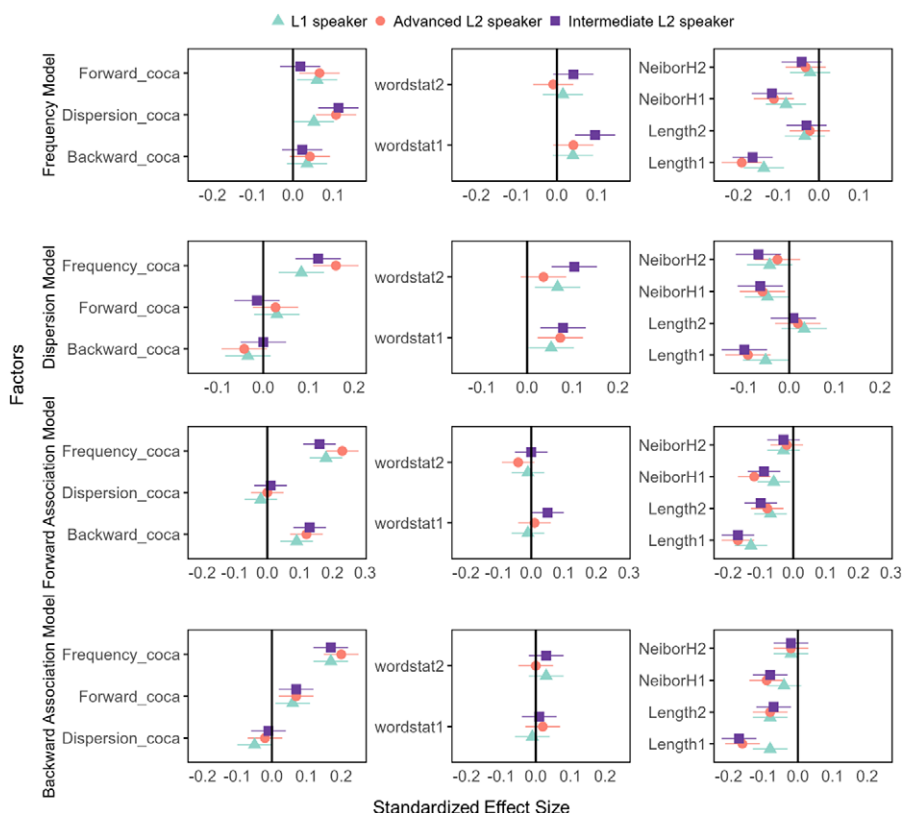
	Dependent variable				Dependent variable				Dependent variable				Dependent variable			
	Frequency				Dispersion				Forward association				Backward association			
	b	Std	t	p	b	Std	t	p	b	Std	t	p	b	Std	t	p
proficiency (ref = L1) Intermediate L2: word1 length	−0.03	.02	−1.16	.25	−0.05	.02	−1.87	.06	−0.04	.02	−1.70	.09	−0.09	.02	−3.60	<.01
proficiency (ref = L1) Advanced L2: word1 neighborhood	−0.03	.02	−1.94	.05	−0.01	.02	−0.68	.49	−0.05	.02	−3.40	<.01	−0.05	.02	−3.13	<.01
proficiency (ref = L1) Intermediate L2: word1 neighborhood	−0.03	.02	−2.24	.03	−0.02	.02	−1.01	.31	−0.03	.02	−1.99	.05	−0.04	.02	−2.79	.01
proficiency (ref = L1) Advanced L2: word2 neighborhood	−0.01	.01	−0.76	.45	0.02	.01	1.13	.26	0.01	.01	0.76	.45	0.01	.01	0.39	.70
proficiency (ref = L1) Intermediate L2: word2 neighborhood	−0.02	.01	−1.44	.15	−0.02	.01	−1.71	.09	0.00	.01	0.01	.99	0.00	.01	0.33	.74
proficiency (ref = L1) Advanced L2: word2 length	0.01	.02	0.59	.56	−0.01	.02	−0.67	.50	−0.01	.02	−0.71	.48	0.00	.02	−0.03	.98
proficiency (ref = L1) Intermediate L2: word2 length	0.00	.02	0.21	.83	−0.02	.02	−1.10	.27	−0.03	.02	−1.31	.19	0.01	.02	0.67	.51
proficiency (ref = L1) Advanced L2: word1 statistics	0.00	.02	0.04	.96	0.02	.02	1.21	.23	0.02	.02	1.01	.31	−0.01	.02	−0.46	.65
proficiency (ref = L1) Intermediate L2: word1 statistics	0.05	.02	3.58	<.01	0.03	.02	1.57	.12	0.06	.02	3.03	<.01	0.02	.02	1.34	.18

(Continued)

Table 5. (Continued)

	Dependent variable				Dependent variable				Dependent variable				Dependent variable			
	Frequency				Dispersion				Forward association				Backward association			
	b	Std	t	p	b	Std	t	p	b	Std	t	p	b	Std	t	p
proficiency (ref = L1) Advanced L2: word2 statistics	−0.02	.02	−1.42	.15	−0.03	.02	−1.75	.08	−0.03	.02	−1.73	.08	−0.03	.02	−1.59	.11
proficiency (ref = L1) Intermediate L2: word2 statistics	0.03	.02	1.52	.13	0.04	.02	2.14	.03	0.02	.02	1.02	.31	−0.01	.02	−0.36	.72
Marginal R <sup>2</sup>		0.1176				0.0950					0.1482				0.1797	
Conditional R <sup>2</sup>		0.1180				0.0950					0.1483				0.1797	

Note: The statistical bins in Table 5 reflect frequency, dispersion, forward association, and backward association, corresponding to their respective metacognitive dependent variables.



**Figure 4.** Forest plot of regression estimates for phrasal- and word-level factors. The x-axis shows standardized effect sizes from Table 5, where positive and negative values indicate facilitative and inhibitory effects on statistical metacognition, respectively.

We first examined phrasal-level statistical factors (see Table 5). Across speaker groups, one type of statistical regularity (e.g., frequency) generally served as a heuristic cue facilitating judgments of others (e.g., association strength, all  $ps < .001$ ), except dispersion, which negatively affected L1 speakers' backward association ratings ( $b = -0.053, p = .001$ ). Interaction analysis revealed that advanced L2 learners' ratings of one type of statistical regularity were more positively influenced by other statistical regularities than were those of intermediate L2 learners and L1 speakers. Specifically, dispersion had a greater positive effect on frequency ratings in both L2 groups than in L1 speakers (all  $ps < .050$ ), and forward association influenced frequency ratings more in advanced than intermediate L2 learners ( $b = 0.048, p = .039$ ). Advanced L2 learners' ratings also showed stronger frequency effects on dispersion ratings than L1 speakers ( $p = .002$ ) and on forward association ratings than intermediate L2 learners ( $p = .016$ ).

Word-level statistics generally facilitated statistical ratings in both L1 and L2 speakers (all  $ps < .050$ ), except that word2 statistics inhibited forward association ratings of advanced L2 learners ( $b = -0.042, p < .001$ ). Interaction analyses showed that intermediate L2 learners' statistical ratings were more positively influenced by word1/word2 statistics than those of L1 speakers and advanced L2 learners

(all  $ps < .050$ ), except that no group differences were found in the moderating effects of word1/word2 statistics on backward association ratings and word1 statistics on dispersion ratings.

Word-form features, namely word1/word2 length and neighborhood, generally inhibited statistical ratings across all speaker groups (all  $ps < .050$ ), except that word2 length positively affected L1 speakers' dispersion ratings ( $b = 0.033$ ,  $p = .028$ ). Interaction analyses showed stronger inhibitory effects in lower-proficiency learners: Word1 length and word1 neighborhood more strongly affected frequency ratings in advanced ( $p = .020$ ) and intermediate L2 learners ( $p = .025$ ), respectively, than in L1 speakers. Word2 neighborhood had a stronger negative effect on dispersion ratings in intermediate than in advanced L2 learners ( $p = .005$ ). Additionally, both L2 groups showed stronger inhibitory effects of word1 neighborhood on forward/backward association ratings and of word1 length on backward association ratings compared to L1 speakers (all  $ps < .050$ ).

Fixed effects explained 9.5–17.9% of variance across measures, with minimal random variance, indicating consistent participant sensitivity to each type of statistical regularity.

## Discussion

We examined three aspects of statistical metacognition among L1, advanced L2, and intermediate L2 speakers. Using an offline rating task with a more precise 9-point slider than prior studies (Siyanova-Chanturia & Schmitt, 2008; Yi et al., 2023), we explored statistical metacognition with greater granularity—how language speakers draw on metalinguistic awareness to consciously access and explicitly judge statistical knowledge that is largely acquired implicitly (Christiansen, 2019). Our findings bridge implicit learning and metalinguistic knowledge, suggesting that language competence involves both automatic processing and reflective judgment. Binomials—a syntactically and semantically regular MWE subtype—were used to examine phrase-level statistical metacognition and extend prior work on collocations (Yi et al., 2023) and three-word combinations (Backman, 1978).

All speaker groups showed low-to-moderate speaker-to-corpus consistency, reflecting shared challenges in developing metacognition aligned with language use. While high speaker-to-corpus consistency may not be critical for implicit competence, it may support conscious processes such as pattern detection and error correction (Ellis, 2009). Nonetheless, all groups exhibited significant metacognitive sensitivity, suggesting shared statistical learning mechanisms for detecting phrase-level patterns in both L1 and L2 speakers, consistent with usage-based theories (Ellis, 2009; Goldberg, 1995; Gries & Ellis, 2015). This sensitivity supports the interface between implicitly acquired knowledge and its accessibility via explicit, task-driven processing (Ellis, 2005). Statistical metacognition was also shaped by word-level information across all groups, reflecting the co-acquisition of multiple information sources as proposed by the dual-route model (Carroll & Conklin, 2014). However, metacognition of statistical regularities varied by type: speaker-to-corpus consistency was highest for frequency, moderate for association strength, and absent for dispersion—highlighting frequency's central role in statistical learning and the difficulty of explicitly accessing abstract, context-dependent cues like dispersion. These findings reflect the complexity of multi-layered statistical learning, engaging distinct cognitive pathways for processing occurrence-based, associative, and contextual cues (Yi et al., 2023). Statistical metacognition also

varied across groups with varying language backgrounds—L1 speakers, advanced L2 learners with English immersion, and intermediate L2 learners with limited natural input. A gradient shift emerged across these groups: from implicit to explicit learning (Ellis, 2009), rich to limited MWE exposure (Adolphs & Durow, 2004), lower to higher cognitive demands (Clahsen & Felser, 2006), less to more efficient practice (DeKeyser, 2020), and holistic to analytical MWE processing (Carroll & Conklin, 2014). In general, L2 speakers' statistical metacognition did not necessarily lag behind that of L1 speakers. Advanced L2 learners, in particular, demonstrated greater speaker-to-corpus consistency and sensitivity than both L1 speakers and intermediate L2 learners, and more effectively used one type of statistical regularity to inform metacognition of another. This suggests that statistical metacognition benefits from integrating rich natural exposure with explicit learning (e.g., rule-based instruction) to support conscious access to implicitly acquired knowledge. In contrast, lower-proficiency learners likely rely more on analytical, word-level cues when forming binomial metacognition, consistent with the dual-route model (Carroll & Conklin, 2014) and the shallow structure hypothesis (Clahsen & Felser, 2006).

A detailed discussion of speaker-to-corpus consistency, sensitivity, and other contributing factors across speaker groups and statistical regularity types is provided below.

### *Speaker-to-corpus consistency of statistical metacognition*

Statistical metacognition showed low-to-moderate speaker-to-corpus consistency, varying by speaker group and regularity type. Advanced L2 speakers generally demonstrated the highest consistency. Notably, their frequency metacognition aligned more closely with actual language distributions than that of L1 speakers—that is, their explicit, conscious frequency judgments better reflected underlying statistical patterns. L1 speakers may overestimate low-frequency binomials if they nevertheless have high familiarity due to their contextual salience (e.g., “*trials and tribulations*”; Siyanova-Chanturia & Pellicer-Sánchez, 2018). In contrast, L2 learners may outperform in explicit, reflective rating tasks due to greater engagement with explicit learning processes and strategies (Ellis, 2009). This suggests that speaker-to-corpus consistency in statistical knowledge is closely tied to explicit learning, though also influenced by real-world usage (Siyanova-Chanturia et al., 2011). Second, advanced L2 learners also showed greater consistency with corpus statistics than intermediate learners in forward association metacognition, consistent with Yi et al. (2023). Associative learning applies more to later stages of L2 acquisition (Ellis & Wulff, 2014). Advanced learners' stronger performance may reflect a more effective integration of usage-based associative learning and explicit instruction, facilitating both acquisition and conscious access to MWE-related association knowledge.

Speaker-to-corpus consistency also varied by statistical regularity: frequency showed the highest consistency across groups, followed by association strength, with dispersion lowest. Prior studies have shown that speakers can distinguish MWEs by frequency (Backman, 1978; Siyanova-Chanturia & Schmitt, 2008; Siyanova-Chanturia & Spina, 2015). Usage-based theories consider frequency the most easily acquired feature supporting broader statistical knowledge, making its high consistency unsurprising. Only one study has examined metacognitive alignment with association strength, reporting low-to-moderate match rates between rating- and corpus-based tertiles for both L1 and L2 speakers (Yi et al., 2023). Associative learning theory (Gries & Ellis, 2015) posits that speakers track co-occurrence beyond frequency. Our findings

reveal finer patterns: while association strength can be explicitly accessed, strong consistency, especially for directional associations, is harder to achieve than for frequency. In contrast, dispersion remains underexplored. As a measure of how linguistic units are distributed across discourse contexts, dispersion captures broader contextual patterns typically acquired implicitly (Batstone, 2002), making explicit encoding and metacognitive access more demanding. These findings suggest that statistical metacognition develops unevenly: speaker-to-corpus consistency is highest for frequency (occurrence-based learning), moderate for association strength (associative learning), and lowest for dispersion (contextual learning).

Interaction analysis showed that lower-proficiency speakers had a smaller advantage (or greater disadvantage) in speaker-to-corpus consistency for association strength than for other statistical regularities. Association strength reflects the configurational patterns of MWEs (Siyanova-Chanturia *et al.*, 2011). This may be because acquiring such information relies more heavily on implicit learning (Ellis, 2009), rich MWE exposure (Adolphs & Durow, 2004), and holistic processing (Carrol & Conklin, 2014). These demands may place lower-proficiency learners, who are limited in these areas, at a greater disadvantage in speaker-to-corpus consistency for association strength than for other statistical measures.

### *Sensitivity of statistical metacognition*

Our results confirm speakers' sensitivity in statistical metacognition, although the effect was not strong. First, all groups showed positive sensitivity to frequency, consistent with robust frequency effects in MWE processing (e.g., Öksüz *et al.*, 2021; Wolter & Yamashita, 2018), particularly for binomials (Conklin & Carrol, 2021; Siyanova-Chanturia *et al.*, 2011; Suhad *et al.*, 2023). Extending prior work, our findings suggest that greater usage-based exposure supports the development of statistical metacognition. Similar to Siyanova-Chanturia and Spina (2015), the sensitivity of frequency metacognition mainly reflected in rating differences between high- and medium/low-frequency items, rather than across the full continuum. Further interaction analyses showed stronger frequency sensitivity in advanced than intermediate L2 speakers, suggesting that speakers' explicit access to frequency variation may be strengthened by a more integrated mix of relatively rich implicit exposure and explicit learning experiences (Ellis & Wulff, 2014).

All speaker groups showed a continuum of positive sensitivity to both forward and backward association. Previous research on MWE processing (e.g., Öksüz *et al.*, 2021; Yi *et al.*, 2018) and rating tasks (Yi *et al.*, 2023) demonstrated L1 and L2 sensitivity to undirected association (e.g., MI, log Dice). Our findings extend this by confirming sensitivity to directional indices, indicating that metacognitive judgments of phrasal formulaicity—rooted in associative learning—are responsive to statistical patterns in language. Interaction analyses further revealed that intermediate L2 learners exhibited stronger forward association sensitivity than L1 speakers, possibly because extensive implicit experience in L1 users compresses statistical distinctions in the mental lexicon, as predicted by the power law of practice (DeKeyser, 2020).

Regarding dispersion, neither L1 nor L2 speakers appeared to explicitly access contextual distributional information of binomials. While context provides semantic cues in processing (Camblin *et al.*, 2007), learners may struggle to consciously apply the concept of "contextual distributional evenness" in explicit rating tasks. This difficulty likely stems from its implicit nature, as proposed by contextual learning



theory (Martin, 2014), or from the greater complexity of tracking macro-level distributions compared to frequency or association strength. Rather than ignoring contextual distributional features, language users may rely on them implicitly for informing language acquisition.

Together, our findings show that speakers across proficiency levels can track statistical variation with metalinguistic awareness, supported by underlying statistical learning mechanisms. Richer usage-based input and explicit instruction further enhance this metacognitive sensitivity. However, sensitivity varies by type: frequency requires threshold exposure, association strength spans a broader continuum, and contextual distribution remains largely inaccessible to explicit encoding—highlighting a distinction between using context for learning and explicitly representing contextual distributional patterns.

### *The influence of other contributing factors*

First, one type of statistical regularity positively influenced the metacognition of another in both L1 and L2 speakers. Consistent with Yi et al. (2023), this suggests that mental representations of frequency (occurrence), association strength (formulaicity), and dispersion (distributional evenness) are interconnected despite being conceptually distinct. An exception was that dispersion negatively predicted backward association metacognition in L1 speakers, possibly because limited contextual variability strengthens word-to-word predictability, whereas broader contextual distribution introduces competing cues that weaken backward links. Interaction analyses further showed that cross-regularity effects were strongest in advanced L2 learners, moderate in intermediate L2 learners, and weakest in L1 speakers. This suggests that the combination of relatively rich explicit learning and sufficient implicit exposure may enhance the integration of different types of statistical knowledge in the mental lexicon.

Word-level statistics—capturing linguistic exposure and mental entrenchment via frequency and dispersion—can be activated to facilitate phrase-level metacognition in both L1 and L2 speakers, consistent with the dual-route model (Carroll & Conklin, 2014). However, prior findings are mixed: Siyanova-Chanturia and Spina (2015) found that word frequency facilitated phrasal metacognition, whereas Yi et al. (2023) reported a similar effect but noted that word frequency inhibited association strength estimation in a mixed L1-L2 group. This discrepancy may reflect measurement differences: earlier studies considered only frequency, which does not fully capture linguistic exposure and mental entrenchment, whereas our metric integrates frequency and contextual distribution for a more comprehensive assessment. Interaction effects showed that lexical statistical regularities most strongly influenced intermediate L2 speakers, followed by advanced L2 and L1 speakers in phrasal statistical metacognition. This supports the shallow structure hypothesis (Clahsen & Felser, 2006), which posits that lower-proficiency learners rely more on analytical, surface-level cues in processing.

Last, word length and neighborhood negatively affected L1 and L2 speakers' metacognition of phrase-level statistical regularities. Complex phonological or orthographic features reduced binomial metacognition, consistent with prior findings on word length effects in collocation processing (Siyanova-Chanturia & Spina, 2015; Yi et al., 2023). This may be because processing words with greater length or neighborhood density imposes greater cognitive demands, reducing attention to phrasal patterns. Alternatively, larger neighborhoods may enhance word recognition (Yates et al., 2005), increasing constituent salience and diverting attention from phrase-level

structure. The finding supports the dual-route model, suggesting word-level cues can influence phrase-level statistical metacognition (Carrol & Conklin, 2014). Interaction analyses showed the strongest word-level influence in intermediate L2 speakers, followed by advanced L2 and then L1 speakers. This aligns with the shallow structure hypothesis (Clahsen & Felser, 2006), which suggests lower-proficiency learners rely on surface-level features, such as word length and neighborhood, in processing MWEs.

Combined, our findings suggest that consciously accessible types of statistical knowledge are interconnected within a shared cognitive framework in both L1 and L2 speakers. The multi-level activation of statistical and word-form features in meta-cognitive judgments supports the dual-route model of MWE processing (Carrol & Conklin, 2014). However, L1-L2 differences emerge. Advanced and intermediate L2 learners—likely due to richer explicit learning (Ellis, 2009)—more readily draw on diverse statistical regularities when making explicit judgments about one type. In contrast, lower-proficiency L2 learners rely more on word-level cues, possibly due to limited usage-based implicit experience (Ellis, 2009), higher cognitive load, shallow processing (Clahsen & Felser, 2006), and a more analytical approach to MWEs (Carrol & Conklin, 2014).

### *Limitations and future directions*

Several limitations should be noted. First, background factors like social status were not controlled, as the study focused on general language use rather than sociolinguistic variation. Specific records of learning trajectories and detailed exposure modes (e.g., written vs. spoken) over the past decades were limited due to the complexity of data collection. Additionally, the L2 group was highly diverse (25+ L1s, few speakers per language), preventing analysis of cross-linguistic influence. Future research could explore how these individual differences shape metacognition of statistical regularities. Second, objective proficiency tests were not used to classify L2 participants. Standardized scores (e.g., IELTS, TOEFL) were available for only 14 participants in our study. Future studies could enhance consistency by administering standardized tests post-recruitment. Third, although COCA is widely used, it cannot fully reflect individual language experience. Future research could adopt individualized input tracking (e.g., diaries) to build longitudinal, speaker-specific corpora for more accurate input characterization. Finally, our findings on binomials offer a starting point rather than a comprehensive account of statistical metacognition across all MWE types. Future research should extend this approach to more structurally variable (e.g., lexical bundles) and semantically opaque MWEs (e.g., idioms) to uncover broader cognitive mechanisms.

### *Conclusion*

This study investigates the speaker-to-corpus consistency, sensitivity, and other contributing factors of statistical metacognition for frequency, association strength, and dispersion in binomials—a key type of MWEs—among L1, advanced L2, and intermediate L2 learners. Our findings have several theoretical implications. Both L1 and L2 speakers can develop statistical metacognition for binomials, suggesting shared statistical learning mechanisms beyond the word level. This extends statistical learning theories (Christiansen, 2019) and usage-based approaches (Goldberg, 1995) by showing that statistical regularities can be explicitly and consciously accessed and reflected upon as metacognition. Moreover, word-level statistics and word-form features influence statistical metacognition, supporting both the dual-route model (Carrol & Conklin, 2014) and

the interplay between statistical learning mechanisms and linguistic knowledge. We also found heterogeneity across statistical learning mechanisms: occurrence- and associative-based knowledge was more accessible in reflective judgments, while context-based knowledge remained less explicitly accessible. Advanced L2 learners outperformed other groups in speaker-to-corpus consistency and sensitivity of statistical metacognition. As the rating task involves explicit metacognitive judgment of implicitly acquired statistical patterns, this may suggest that the combination of relatively extensive explicit learning experiences and usage-based implicit exposure in advanced L2 speakers facilitates the development of statistical metacognition. Finally, lower-proficiency speakers were more influenced by word-level statistical and linguistic information in phrase-level metacognition, likely due to more limited exposure, reduced cognitive resources, and more analytical processing strategies.

**Supplementary material.** The supplementary material for this article can be found at <http://doi.org/10.1017/S0272263125101411>.

**Acknowledgments.** We are deeply grateful for the invaluable and constructive feedback provided by the editors and anonymous reviewers during the review process.

**Data availability statement.** The data in this study are available at: <https://osf.io/bjrx9/>.

**Competing interests.** None.

## References

- Adolphs, S., & Durow, V. (2004). Social-cultural integration and the development of formulaic sequences. *Formulaic sequences: Acquisition, processing and use (Language Learning & Language Teaching, 9)*, 107–126. <https://doi.org/10.1075/llt.9.07ado>
- Baayen, R. H., Milin, P., & Ramscar, M. (2016). Frequency in lexical processing. *Aphasiology, 30*(11), 1174–1220. <https://doi.org/10.1080/02687038.2016.1147767>
- Backman, J. (1978). *Subjective structures in linguistic recurrence*. Umeå universitet.
- Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., ... & Bolker, M. B. (2015). Package “lme4.” *Convergence, 12*(1), 2.
- Batstone, R. (2002). Contexts of engagement: A discourse perspective on “intake” and “pushed output.” *System, 30*(1), 1–14. [https://doi.org/10.1016/S0346-251X\(01\)00055-0](https://doi.org/10.1016/S0346-251X(01)00055-0)
- Boulenger, V., Shtyrov, Y., & Pulvermüller, F. (2012). When do you grasp the idea? MEG evidence for instantaneous idiom understanding. *Neuroimage, 59*(4), 3502–3513. <https://doi.org/10.1016/j.neuroimage.2011.11.011>
- Camblin, C. C., Gordon, P. C., & Swaab, T. Y. (2007). The interplay of discourse congruence and lexical association during sentence processing: Evidence from ERPs and eye tracking. *Journal of Memory and Language, 56*(1), 103–128. <https://doi.org/10.1016/j.jml.2006.07.005>
- Carrol, G., & Conklin, K. (2014). Getting your wires crossed: Evidence for fast processing of L1 idioms in an L2. *Bilingualism: Language and Cognition, 17*(4), 784–797. <https://doi.org/10.1017/S1366728913000795>
- Christiansen, M. H. (2019). Implicit statistical learning: A tale of two literatures. *Topics in Cognitive Science, 11*(3), 468–481. <https://doi.org/10.1111/tops.12332>
- Clahsen, H., & Felser, C. (2006). Continuity and shallow structures in language processing. *Applied Psycholinguistics, 27*(1), 107–126. <https://doi.org/10.1017/S0142716406060206>
- Conklin, K., & Carrol, G. (2021). Words go together like ‘bread and butter’: The rapid, automatic acquisition of lexical patterns. *Applied Linguistics, 42*(3), 492–513. <https://doi.org/10.1093/applin/amaa034>
- Davies, M. (2009). The 385+ million word Corpus of Contemporary American English (1990–2008+): Design, architecture, and linguistic insights. *International Journal of Corpus Linguistics, 14*(2), 159–190. <https://doi.org/10.1075/ijcl.14.2.02dav>
- DeKeyser, R. (2020). Skill acquisition theory. In *Theories in second language acquisition* (pp. 83–104). Routledge.

- Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition*, 27(2), 141–172. <https://doi.org/10.1017/S0272263105050096>
- Ellis, R. (2009). Implicit and explicit learning, knowledge and instruction. In *Implicit and explicit knowledge in second language learning, testing and teaching*, 42 (pp. 3–25). <https://doi.org/10.21832/9781847691767-003>
- Ellis, N. C., & Wulff, S. (2014). Usage-based approaches to SLA. In *Theories in second language acquisition* (pp. 87–105). Routledge.
- Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. University of Chicago Press.
- Gries, S. T. (2013). 50-something years of work on collocations: What is or should be next .... *International Journal of Corpus Linguistics*, 18(1), 137–166.
- Gries, S. T. (2024). *Frequency, dispersion, association, and keyness: Revising and tupleizing corpus-linguistic measures*. Studies in Corpus Linguistics, vol. 115. John Benjamins. <https://doi.org/10.1075/scl.115>
- Gries, S. T., & Ellis, N. C. (2015). Statistical measures for usage-based linguistics. *Language Learning*, 65(S1), 228–255. <https://doi.org/10.1111/lang.12119>
- Marian, V., Bartolotti, J., Chabal, S., & Shook, A. (2012). CLEARPOND: Cross-linguistic easy-access resource for phonological and orthographic neighborhood densities. *PLoS One*. <https://doi.org/10.1371/journal.pone.0043230>
- Martin, J. R. (2014). A contextual theory of language. In *The powers of literacy (RLE Edu I)* (pp. 116–136). Routledge.
- Morgan, E., & Levy, R. (2016). Abstract knowledge versus direct experience in processing of binomial expressions. *Cognition*, 157, 384–402. <https://doi.org/10.1016/j.cognition.2016.09.011>
- Öksüz, D., Brezina, V., & Rebuschat, P. (2021). Collocational processing in L1 and L2: The effects of word frequency, collocational frequency, and association. *Language Learning*, 71(1), 55–98. <https://doi.org/10.1111/lang.12427>
- R Core Team. (2025). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Schmitt, N., & Dunham, B. (1999). Exploring native and non-native intuition of word frequency. *Second Language Research*, 15(4), 389–411. <https://doi.org/10.1191/026765899669633186>
- Siyanova-Chanturia, A. (2015). On the “holistic” nature of formulaic language. *Corpus Linguistics and Linguistic Theory*, 11(2), 285–301. <https://doi.org/10.1515/cllt-2014-0016>
- Siyanova-Chanturia, A., Conklin, K., & Van Heuven, W. J. (2011). Seeing a phrase “time and again” matters: The role of phrasal frequency in the processing of multiword sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(3), 776.
- Siyanova-Chanturia, A., & Pellicer-Sánchez, A. (2018). Formulaic language: Setting the scene. In *Understanding formulaic language* (pp. 1–15). Routledge.
- Siyanova-Chanturia, A., & Martinez, R. (2015). The idiom principle revisited. *Applied Linguistics*, 36(5), 549–569. <https://doi.org/10.1093/applin/amt054>.
- Siyanova-Chanturia, A., & Schmitt, N. (2008). L2 learner production and processing of collocation: A multi-study perspective. *Canadian Modern Language Review*, 64(3), 429–458. <https://doi.org/10.3138/cmlr.64.3.429>
- Siyanova-Chanturia, A., & Spina, S. (2015). Investigation of native speakers and second language learners’ intuition of collocation frequency. *Language Learning*, 65(3), 533–562. <https://doi.org/10.1111/lang.12125>
- Sonbul, S., El-Dakhs, D. A. S., Conklin, K., & Carrol, G. (2023). “Bread and butter” or “butter and bread”? Nonnatives’ processing of novel lexical patterns in context. *Studies in Second Language Acquisition*, 45(2), 370–392. <https://doi.org/10.1017/S0272263122000237>
- Tomaschek, F., Hendrix, P., & Baayen, R. H. (2018). Strategies for addressing collinearity in multivariate linguistic data. *Journal of Phonetics*, 71, 249–267. <https://doi.org/10.1016/j.wocn.2018.09.004>
- Tremblay, A., Derwing, B., Libben, G., & Westbury, C. (2011). Processing advantages of lexical bundles: Evidence from self-paced reading and sentence recall tasks. *Language Learning*, 61(2), 569–613. <https://doi.org/10.1111/j.1467-9922.2010.00622.x>
- Tryk, H. E. (1968). Subjective scaling of word frequency. *The American Journal of Psychology*, 81(2), 170–177. <https://doi.org/10.2307/1421261>
- Wolter, B., & Yamashita, J. (2018). Word frequency, collocational frequency, L1 congruency, and proficiency in L2 collocational processing: What accounts for L2 performance? *Studies in Second Language Acquisition*, 40(2), 395–416. <https://doi.org/10.1017/S0272263117000237>

- Yates, M. (2005). Phonological neighbors speed visual word processing: Evidence from multiple tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(6), 1385. <https://doi.org/10.1037/0278-7393.31.6.1385>
- Yi, W. (2018). Statistical sensitivity, cognitive aptitudes, and processing of collocations. *Studies in Second Language Acquisition*, 40(4), 831–856. <https://doi.org/10.1017/S0272263118000141>
- Yi, W., Lu, S., & Ma, G. (2017). Frequency, contingency and online processing of multiword sequences: An eye-tracking study. *Second Language Research*, 33(4), 519–549. <https://doi.org/10.1177/0267658317708009>
- Yi, W., Man, K., & Maie, R. (2023). Investigating first and second language speaker intuition of phrasal frequency and association strength of multiword sequences. *Language Learning*, 73(1), 266–300. <https://doi.org/10.1111/lang.12521>
- Yi, W., & Zhong, Y. (2024). The processing advantage of multiword sequences: A meta-analysis. *Studies in Second Language Acquisition*, 46(2), 427–452. <https://doi.org/10.1017/S0272263123000542>

---

**Cite this article:** Zhong, Y., Todd, S., Gries, S. Th., & Brehm, L. (2025). Metacognition of frequency, directional association strength, and dispersion of MWEs in first and second language speakers. *Studies in Second Language Acquisition*, 1–27. <https://doi.org/10.1017/S0272263125101411>