Usage-Based Individual Differences in the Probabilistic Processing of Multi-Word Sequences

Kyla McConnell* and Alice Blumenthal-Dramé

English Department, University of Freiburg, Freiburg, Germany

While it is widely acknowledged that both predictive expectations and retrodictive integration influence language processing, the individual differences that affect these two processes and the best metrics for observing them have yet to be fully described. The present study aims to contribute to the debate by investigating the extent to which experienced-based variables modulate the processing of word pairs (bigrams). Specifically, we investigate how age and reading experience correlate with lexical anticipation and integration, and how this effect can be captured by the metrics of forward and backward transition probability (TP). Participants read more and less strongly associated bigrams, paired in sets of four to control for known lexical covariates such as bigram frequency and semantic meaning (i.e., *absolute control*, *total control*, *absolute silence*, *total silence*) in a self-paced reading (SPR) task. They additionally completed assessments of exposure to print text (Author Recognition Test, Shipley vocabulary assessment, Words that Go Together task) and provided their age. Results show that both older age and lesser reading experience individually correlate with stronger TP effects. Moreover, TP effects differ across the spillover region (the two words following the noun in the bigram).

Keywords: individual differences, predictive processing, lexical integration, self-paced reading, age, reading ability, transition probability

INTRODUCTION

Comprehending language is one of the most complicated tasks that humans perform on a regular basis, yet we do it with astounding proficiency and ease. One mechanism that may support this effortless comprehension is probabilistic processing, a framework that has gained traction in recent empirical research in linguistics (Ambridge et al., 2015; Bod et al., 2015; Blumenthal-Dramé, 2016b; Behrens and Pfänder, 2016; Kuperberg and Jaeger, 2016; Armeni et al., 2017; Divjak, 2019). Probabilistic models of language acquisition and processing are situated within the broader paradigm of probabilistic cognition, which assumes that humans learn about and construct a mental representation of the world based on distributional information from the environment and apply this statistical knowledge in interacting with and predicting future states of the world. These abilities for statistical learning and processing are claimed to be key ingredients to cognition in domains as different as motor control, visual perception, causal learning, inferential reasoning and language (Chater and Oaksford, 2008; Perfors et al., 2011; Tenenbaum et al., 2011; Griffiths et al., 2012; Lupyan and Clark, 2015; Ordin et al., 2020).

A myriad of studies confirm that comprehenders are sensitive to word-level distributional patterns such as word frequency (see, for example, Gries and Divjak, 2012). Experimental

evidence shows that these statistical constraints can be utilized in the service of prediction on diverse levels of language comprehension, including the first sound of an upcoming word (DeLong et al., 2005; but see ; Nieuwland et al., 2018; Yan et al., 2017), the grammatical gender of an upcoming word (Wicha et al., 2003), the semantic features of expected lexical items (Federmeier and Kutas, 1999), and specific words that fit sentence-level context (Otten and Van Berkum, 2008). Strong views of prediction in language processing, such as the view marshalled by the "predictive coding" framework, posit that the brain's central function is to predict across all types of sensory input, from completing a friend's sentence to driving down a familiar road. According to this view, the mind predicts constantly and sensory input is largely ignored unless it disconfirms these predictions (Clark, 2013; Hohwy, 2013; Lupyan and Clark, 2015).

However, just because predictive processing is plausible doesn't mean that it is strategy that the processor must always implement (Huettig and Mani, 2015; Pickering and Gambi, 2018). There is growing consensus for the view that "perhaps prediction occurs less or not at all in challenging situations and in less proficient language users" (Huettig, 2015, 131). Prediction also seems to be susceptible to task demands; it is reduced when processing must occur rapidly and results can be significantly altered by experimental design factors (Wlotko and Federmeier, 2015; Huettig and Guerra, 2019). The fact that not all comprehenders predict in all situations implies that "predictive processing may not be the best—or even a viable—strategy for all individuals at all phases of the lifespan and/or in all processing situations" (Federmeier, 2007, 495). Indeed, it may be simply one of many strategies that the processor can select from depending on task demands and capacity (Huettig and Mani, 2015; Huettig and Janse, 2016).

Further, comprehenders' knowledge of the distributional patterns of language are not necessarily utilized in a strictly forward-looking direction. Rather, probabilistic knowledge of language patterns may facilitate integration of incoming input to unfolding meanings and structures; that is, more constrained input may simply be easier to link to existing context once encountered, facilitating comprehension (Ferreira and Chantavarin, 2018). These two routes, prediction and integration, must not be fully independent or exclusive, and may indeed be "two sides of the same coin" (Ferreira and Chantavarin, 2018, 447). In the following paper, we explore the extent to which experience-based individual differences affect lexical anticipation and integration as captured by two metrics that have taken center stage in probabilistic linguistics: forward and backward transition probability.

Experience-Based Individual Differences

Adult native speakers were long thought to achieve a certain ideal attainment, in which they had a complete and uniform understanding of their native language. However, usage-based approaches posit that human skills are highly plastic and shaped by experience (Evans and Levinson, 2009; Dąbrowska, 2015; Dąbrowska, 2018). Recent research has highlighted that

language attainment within adult native speakers is modulated by both endogenous constraints (e.g., executive functions, statistical learning abilities, personality traits) and exogenous, experience-related variables (notably, the quality and quantity of the input) (Andringa and Dąbrowska, 2019; Dąbrowska, 2019; Frost et al., 2019; Medimorec et al., 2019; Divjak and Milin, 2020; Kidd and Donnelly, 2020; Ryskin et al., 2020; Kerz and Wiechmann, 2021).

As far as the exogenous variables are concerned, different lines of experimental research converge to suggest that subjects' language processing strategies are fine-tuned to their linguistic experience. Members of the same speech community process language differently based on their personal experience, which includes, among other things: language exposure (Brysbaert et al., 2018; Dąbrowska, 2019), world knowledge (Ryskin et al., 2019; Venhuizen et al., 2019; Troyer and Kutas, 2020), vocabulary size (Yap et al., 2012; Milin et al., 2017; Kidd et al., 2018), print exposure (Falkauskas and Kuperman, 2015), domain expertise (Verhagen et al., 2018) and social network size (Lev-Ari, 2019). Further, language processing changes across the lifespan as experience accumulates (and the brain changes) (Brysbaert et al., 2016; Hanulíková et al., 2020). Frequency effects also reflect the impact of experience on language processing; even structurally identical multi-morphemic sequences are processed differently depending on their usage frequency, with higher-frequency sequences (like *government* or *I don't know*) eliciting greater processing ease and chunked access relative to rarer ones (like *amazement* or *You don't swim*) (Blumenthal-Dramé, 2016a; Blumenthal-Dramé et al., 2017; Carrol and Conklin, 2019).

Reading Experience

Written texts have different distributional patterns compared to spoken language; they generally have a higher proportion of low frequency words and a wider variety of complex syntactic structures (Biber, 1995; Roland et al., 2007; Dąbrowska, 2018). Not only does initially acquiring reading skills reshape the brain and improve language skills, but literate adult readers also vary considerably in their reading frequency and ability (Huettig and Mishra, 2014; Dehaene et al., 2015). Reading experience has been found to affect language processing, especially in the domain of lexical-level effects, where more proficient readers show reduced sensitivity to word frequencies, potentially indicating that they allocate less effort to word-level decoding and lexical access (Kuperman and Van Dyke, 2011; Yap et al., 2012; Adelman et al., 2014; Falkauskas and Kuperman, 2015; but see ; Kuperman and Van Dyke, 2013).

Although more proficient readers show reduced word frequency effects, more reading experience seems to pattern with heightened prediction effects (Kukona et al., 2016; Favier et al., 2021). Populations with smaller vocabularies, such as non-native speakers and low literates, rely less on predictions (Mishra et al., 2012; Huettig and Mani, 2015; Pickering and Gambi, 2018). Children's predictive ability in comprehension covaries with their language production ability, which may be linked to their

experience with language overall (Mani and Huettig, 2012). Domain-specific text experience can also encourage prediction; job-seekers have lower voice onset times in predictively completing job-related multi-word sequences ("good communication...") (Verhagen et al., 2018). Neurolinguistic studies also support the idea that increased reading experience leads to greater sensitivity to predictability, though the effect seems to be sensitive to task demands, especially time pressure (Ng et al., 2017; Tabullo et al., 2020).

The conclusion that increased reading experience leads to larger predictability effects is not undisputed, however. Slattery and Yates find that better readers show a lessened effect of context-based predictability on gaze durations, while Whitford and Titone report that reading experience modulates frequency effects but not predictability effects (Whitford and Titone, 2014; Slattery and Yates, 2018). One reason for these inconsistencies might be that it is not always clear which level of prediction the study is actually measuring, e.g., the level of letters, morphemes, lexemes or meanings (Huettig, 2015). Another reason might be that research on predictive processing and reading experience has operationalized predictability in different ways. Quantifying predictability via cloze scores or other context-based measures is likely to tap into general world knowledge, while relying on statistics extracted from corpora is more likely to capture language-specific distributional knowledge. However, it may also be that more experienced readers have better bottom-up word recognition skills and thus a reduced need for probabilistic processing, as summarized by Hersch and Andrews (2012):

> "although skilled readers are more effective at using context and previous knowledge to facilitate the higher order integration and inferential processes required for comprehension (...) they can retrieve lexical and semantic information for most words using bottom-up information before contextually based predictions become available (Perfetti, 1992)" (Hersch and Andrews, 2012, 241).

The question thus becomes whether probabilistic processing, whether prediction or integration, is a compensatory mechanism or a luxury (or perhaps it can serve both purposes, depending on the person and the situation). Furthermore, do the conclusions based on cloze-based predictability, which likely draws at least partially on world knowledge, extend to probabilistic processing on the level of lexical co-occurrence? In the current study, we investigate the effect of reading experience on the prediction and integration of lexical co-occurrence, while also considering the experience-based factor of age, to address these questions and determine how language experience interacts with processing.

Age

Age has widely been assumed to have a modulatory effect on predictive processing; however, the direction of this effect is still under debate (Gordon et al., 2016; Dave et al., 2018; Payne and Silcox, 2019). On the one hand, older adults generally have greater *crystallized intelligence* (i.e., knowledge accumulated throughout the lifespan), which is comprised of non-linguistic

world knowledge, vocabulary (Brysbaert et al., 2016; Sánchez-Izquierdo and Fernández-Ballesteros, 2021), schematic or generalized representations of common occurrences (Ghosh and Gilboa, 2014), and more entrenched distributional knowledge (such as which units of language typically co-occur in language use) (Ramscar et al., 2014; Whitford and Titone, 2019). It has been hypothesized that older adults may rely on their superior crystallized knowledge to engage more readily in linguistic prediction, possibly as a strategy to compensate for perceptual (auditory, visual) decline or for age-related slowing (for review, see Gordon et al., 2016; Payne and Silcox, 2019).

In line with this view, several eye-movement studies have found evidence in support of the assumption that older readers rely on top-down knowledge more strongly, though potentially in a qualitatively different manner than younger readers. Word frequency effects are larger in older than younger adults and older readers show differential, and sometimes stronger, sensitivity to predictability in spoken and written word processing (Kliegl et al., 2004; Laubrock et al., 2006; Rayner et al., 2006; Rayner et al., 2011; Rayner et al., 2013; Huettig and Janse, 2016; Moers et al., 2017; Choi et al., 2017; Steen-Baker et al., 2017; Whitford and Titone, 2017; Whitford and Titone, 2019).

On the other hand, electrophysiological event-related potentials research suggests reduced and delayed predictability effects of sentential context (as assessed in terms of cloze probability for the sentence-final word) in older adults (Federmeier et al., 2010; DeLong et al., 2012; Wlotko et al., 2012; Wlotko and Federmeier, 2012), sometimes in combination with preserved or increased lexical-level effects (word frequency and orthographic neighborhood density; Payne and Federmeier, 2018).

The inconsistency between behavioral and electrophysiological results about age effects and probabilistic processing might be attributable to differences in experimental design. For example, eye tracking studies typically allow for relatively naturalistic reading strategies, whereas EEG studies tend to adopt a rapid serial visual presentation format, where subjects do not move their eyes and cannot read at a natural pace, since subsequent words of a sentence overwrite each other as they are presented at the same location of the screen. Moreover, studies have adopted different ways of assessing predictability (e.g., cloze predictability vs distributional statistics derived from corpora) and have not consistently differentiated between different sources of top-down knowledge (e.g., world knowledge plausibility, co-occurrence statistics in corpora, lexical frequency; see Whitford and Titone, 2017). Finally, many studies have focused on prediction effects for the last words of sentences, which might be confounded by wrap-up processes known to change across the lifespan (Stine-Morrow and Payne, 2016).

In summary, we expected healthy older people to show heightened effects of lexical-level association patterns than their younger counterparts, independent of their reading experience. This effect could be driven by two different sources: probabilistic processing may be relied on more

heavily in cases where the processor has sensory processing deficits or additional experience may simply lead to more efficient language processing in general.

Transition Probabilities in Language Learning and Processing

Broadly speaking, forward transition probability (FTP) can be defined as the probability of occurrence of a unit given the preceding context, whereas backward transition probability (BTP) is the probability of occurrence of a unit given the following context October 01, 2021 15:54:00.[1] FTP and BTP have been shown to have a number of correlates in language acquisition, representation and processing at different linguistic levels: from phones and syllables (Aslin, 2017; Hartshorne et al., 2019), to morphemes, lexemes and parts of speech (Solomyak and Marantz, 2010; Lewis et al., 2011; Henderson et al., 2016; Aslin, 2017; Blumenthal-Dramé et al., 2017; Çöltekin, 2017; Hartshorne et al., 2019), and even syntactic structures and semantic representations (Linzen and Jaeger, 2016; Venhuizen et al., 2019).

Most prominently, FTP and BTP have been related to the formation and use of multi-unit "chunks" at varying levels of representation (Pelucchi et al., 2009; McCauley and Christiansen, 2019; Roete et al., 2020). Chunking works in two directions: from wholes to parts (*segmentation*), and from parts to wholes (*grouping*) (Christiansen and Arnon, 2017). On the one hand, language users (and, notably, children acquiring their first language) learn to segment the continuous sensory stream of language into sub-units at different grain sizes (e.g., words, morphemes, syllables, phones). On the other hand, they learn to treat frequently co-occurring low-level items as configurations that can be manipulated holistically at a higher level of representation (thus, a succession of morphemes like *govern-* and *-ment* can be handled as a unit at the lexical level or a succession of syntactic categories like (Det) (N) can be treated as a phrasal unit). Importantly, research suggests that language users capitalize on the fact that TPs are low between segments and high within segments to extract and construct the building blocks of their language (Blumenthal-Dramé et al., 2017; Isbilen et al., 2020).

Moreover, empirical research has demonstrated that bigrams with lower TPs are more effortful to process, as measured in terms of processing speed, accuracy and brain activation (McDonald and Shillcock, 2003a; McDonald and Shillcock, 2003b; Boston et al., 2008; Demberg and Keller, 2008; Levy, 2008; Frank and Bod, 2011; Frank, 2013; Smith and Levy, 2013; Frank et al., 2015; Hale, 2016; Willems et al., 2016; Lopopolo et al., 2017; Nelson et al., 2017; Lowder et al., 2018). High TPs have also been shown to support language comprehension under challenging speech or reading circumstances, arguably because they allow for the top-down activation of missing or degraded lower-level components (Lorenz and Tizón-Couto, 2019). Likewise, it has been demonstrated that partial input primes high- but not low-TP

completions (e.g., *govern* primes *government* more than it primes *governor*, because of higher TPs between *govern-* and *-ment* than between *govern-* and *-or*) (Blumenthal-Dramé, 2017; Blumenthal-Dramé et al., 2017). These different strands of empirical research, along with the inherent forward-looking directionality of FTP, have been taken to suggest that FTP taps into predictive online processing (for discussion, see Aurnhammer and Frank, 2019). BTP, by contrast, does not tap into prediction processes, which by definition must occur before the sensory signal has been encountered. Rather, BTP is likely to correlate with the ease of retrodiction or integration processes, which connect a physically present sensory signal to the previous input to generate a unified and coherent mental representation (Ferreira and Chantavarin, 2018; van Paridon and Alday, 2020; Onnis et al.).

The dual mechanisms of prediction and integration must not be separable or mutually exclusive and neither must forward and backward transition probability; adult native speakers are likely to be sensitive to both (Perruchet and Desaulty, 2008) and they may be descriptive in parallel (McCauley and Christiansen, 2019). With this in mind, we operationalize these two core linguistic processes through the metrics of BTP and FTP, in order to assess how they are modulated by experience. Ultimately, we aim to answer whether greater experience leads to greater reliance on statistical regularities, making them a "luxury" for the best-equipped processors (such as those who have accumulated the most experience through age or reading exposure), or rather they are a backup or compensatory mechanism for those who have deficits in other linguistic areas (either due to age-related cognitive decline or lower exposure to reading).

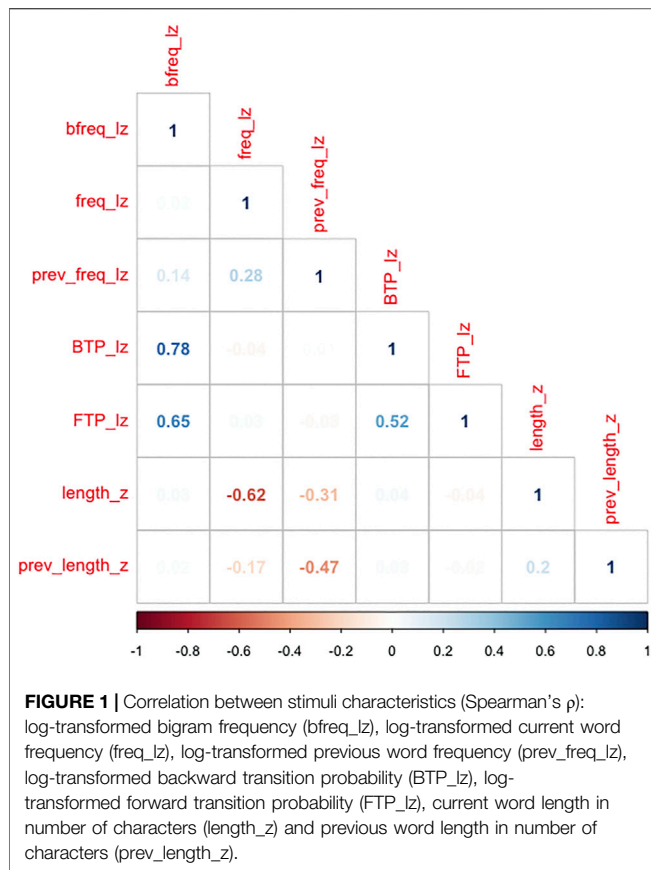## MATERIALS AND METHODS

Stimuli
Bigram Extraction
To maximize objectivity in stimuli selection while simultaneously achieving naturalistic stimuli, modifier-noun bigrams were extracted computationally from the Corpus of Contemporary American English (COCA), which contains a total of about one billion words from 1990 to 2019. Stimuli selection included three main steps: First, an initial list of 1,133 modifier-noun bigrams[2] was extracted from the corpus over a smooth range of log-transformed bigram frequency (MacCallum et al., 2002; Baayen, 2010). Bigram frequency was calculated on the lemma level and ranged from 1 to 6,060 (approx. 6 per million words), with a median value of 38.

Bigrams were rejected if they were emotionally arousing (violent, sexual, or religious) or specialist terms from restricted domains like sports or medicine, so pairs like "primal screams" and "canonical coefficients" were not eligible. Additionally, only bigrams containing one lexical (base) morpheme and up to two derivational and

---

[1]In the following, findings related to the closely related information-theoretic metric of *surprisal* (which can be defined as negative log FTP), will be subsumed under FTP.

[2]Initially, 5,000 items were extracted, but the majority of these were lemma duplicates, so the list had to be automatically reduced to only one item per lemma (i.e., to remove a plural noun if the singular noun was present).

FIGURE 1 | Correlation between stimuli characteristics (Spearman's ρ): log-transformed bigram frequency (bfreq_lz), log-transformed current word frequency (freq_lz), log-transformed previous word frequency (prev_freq_lz), log-transformed backward transition probability (BTP_lz), log-transformed forward transition probability (FTP_lz), current word length in number of characters (length_z) and previous word length in number of characters (prev_length_z).

inflectional morphemes were permitted, because words made up of more than one lexical morpheme could themselves be seen as collocations at the morphological level ("widespread," "lifestyles"). Similarly, words beginning with removable prefixes were not included ("inexperienced," "unusual"). Both words in the bigram also had to have consistent spelling in both British and American English and be less than 12 letters long.

In addition to these criteria, idioms and compounds were also removed. Idioms were identified when at least one word was used "in a figurative, technical, or de-lexical sense only found in the context of a limited number of collocates" (Howarth, 1996; Howarth, 1998; presented in Gyllstad and Wolter, 2016, 299). Thus, bigrams like "eager beaver" and "tidy profits" were ineligible. Bigrams were removed for being compounds if they were not separable, did not allow modification of the first element and/or had non-compositional meaning, e.g., "botanic gardens" and "instant messaging" (Bauer, 2004, 11).

Bigram Pairing

Of course, word associations and other lexical-level factors are not the only force affecting lexical processing; n-gram (or chunk) frequency is also a major factor (Lorenz and Tizón-Couto, 2019; Supasiraprapa, 2019). Usage-based approaches posit that the comprehender does not access, concatenate, or integrate the component words of high-frequency n-grams but rather retrieves chunks of varying sizes holistically (Blumenthal-Dramé, 2017; Ambridge, 2020; Havron

and Arnon, 2021). Further, higher n-gram frequency correlates with greater speed and accuracy in comprehension, production, and acquisition regardless of other lexical-level factors, and this effect is consistently found in self-paced reading paradigms (McConnell and Blumenthal-Dramé, 2019).

However, controlling for bigram frequency in a statistical model that focuses on transition probabilities is difficult; not only are the metrics often highly correlated, but they are also intrinsically linked because bigram frequency is an essential component in calculating transition probabilities[3]. Thus, we matched each of the 501 remaining bigrams with another bigram with the same first word and a maximally similar log-transformed bigram frequency (within a window of ± 0.25). By keeping bigram frequency constant, the effect of the second word in terms of its association to the first word could be isolated. The repetition of the first word (W1) also establishes a natural control item. That is, if W1 provides a processing advantage regardless of both frequency and association strength, this advantage should be seen in both bigrams. By comparing these bigrams, we can isolate the processing advantage of the word-level associations captured by forward and backward transition probability. The matched items were checked along the same criteria presented above and a random sample of 75 pairs were extracted from the eligible items.

Although creating two bigrams with the same W1 allows us to "zoom in" on the level of association between individual words, it would still be impossible to rule out that differences to RTs on the second word (W2) were not based on specific characteristics of that word. That is, hypothetically, if "bad luck" is read faster than "bad idea," this could reflect that "luck" is easier to process than "idea" for any reason, regardless of the word that proceeds it. To avoid this problem and to counteract any possible effects of meaningfulness (Jolsvai et al., 2020), we establish a baseline by adding another pair of bigrams which contain the same W2s as the original bigrams, but paired with synonymous adjectives. This allows for the comparison of bigrams on the basis of association strength between the two words, without influence of chunk frequency or any word-level characteristics of either individual word.

For each bigram, FTP was calculated as the raw frequency of the bigram divided by the (raw) frequency of the first word (range: 0.1894–0.00001; mean: 0.0101). BTP was quantified as bigram frequency divided by the frequency of the second word (range: 0.0955–0.00001; mean: 0.0035). TPs were log-transformed then centered and scaled using the scale() function in R. In our sample, forward and backward transition probability were correlated with each other (Spearman's ρ = 0.52) and with bigram frequency (ρ = 0.65 and ρ = 0.78, respectively; see **Figures 1**, **2**).

Sentences

Critical bigrams were embedded in sentence onsets that were designed to be equally plausible with any of the four matched bigrams, so that onsets could be counterbalanced across participants. Critical words were followed by three-word spillover regions that were held as consistent as possible over

---

[3]FTP = W1 frequency/bigram frequency; BTP = W2 frequency/bigram frequency
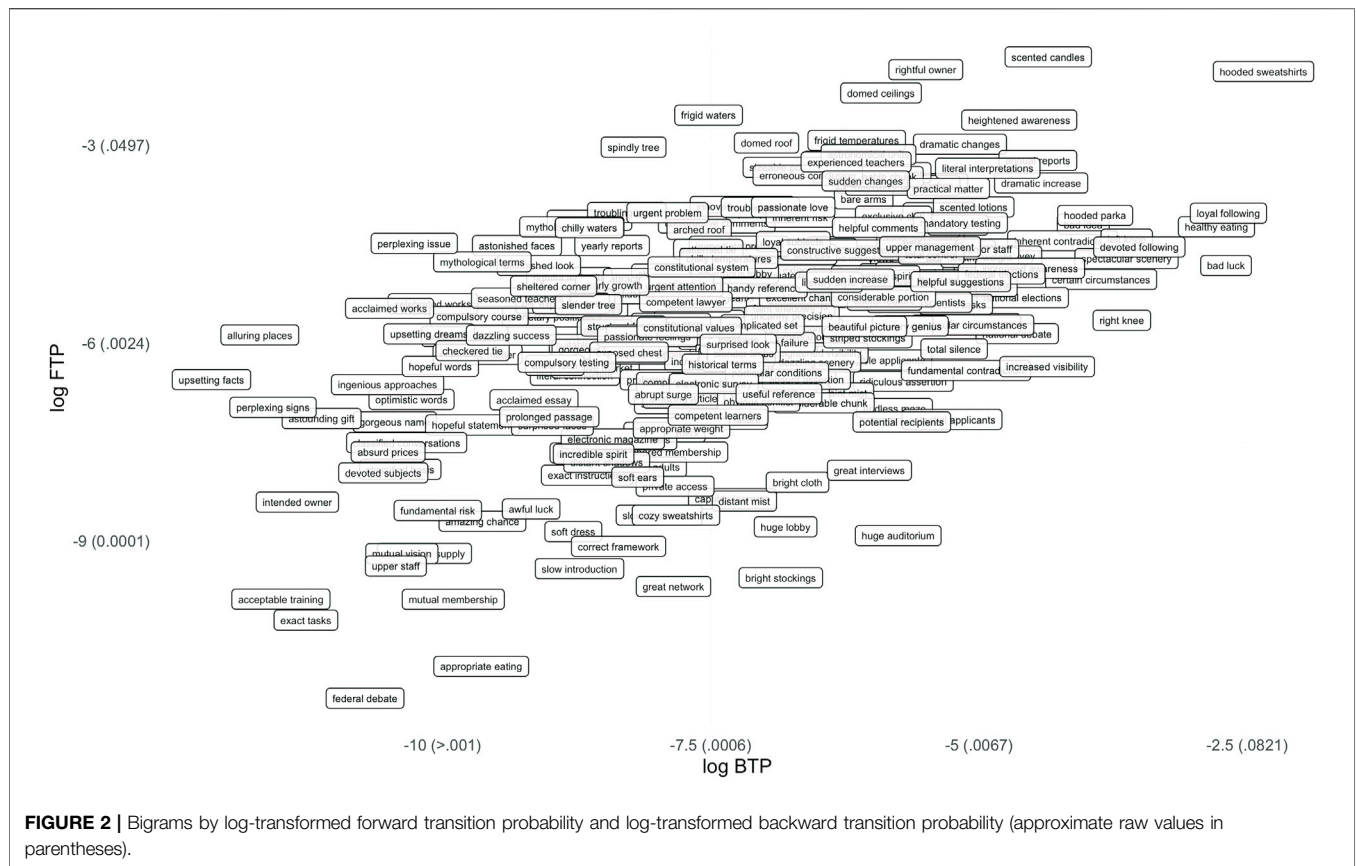
FIGURE 2 | Bigrams by log-transformed forward transition probability and log-transformed backward transition probability (approximate raw values in parentheses).

TABLE 1 | Example stimuli (from lists A and B).

List	Sentence onset	Critical bigram	Spillover
A1	John saw the	absolute control	of the company as a bad thing
A1	James ensured	absolute silence	throughout the entire exam
A2	Charlie ensured	total control	throughout the entire process
A2	Mike saw the	total silence	of the children as a bad sign
B1	James ensured	absolute control	throughout the entire process
B1	John saw the	absolute silence	of the children as a bad sign
B2	Mike saw the	total control	of the company as a bad thing
B2	Charlie ensured	total silence	throughout the entire exam

the two pairs to maximize comparability. These spillover regions were followed by at least one further word to catch sentence-end wrap-up effects. See **Table 1** for example stimuli. Each participant saw one list (A or B), thus participants did not read highly similar sentences. Similar sentences (i.e., both labeled A1) were spaced maximally within a set, and shuffled in order across participants.
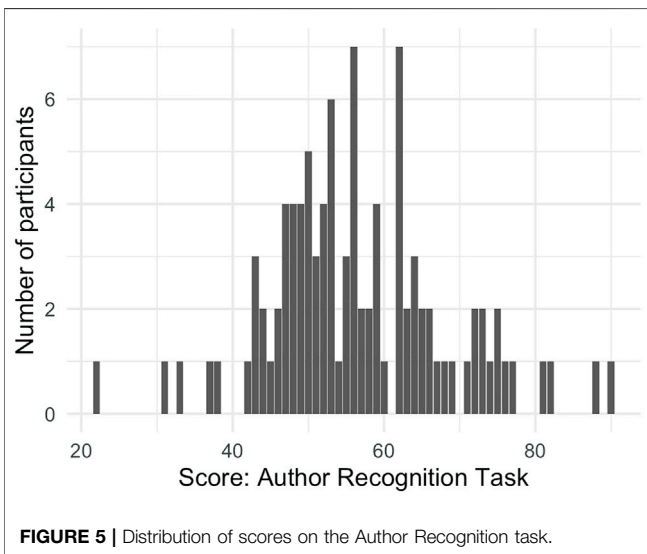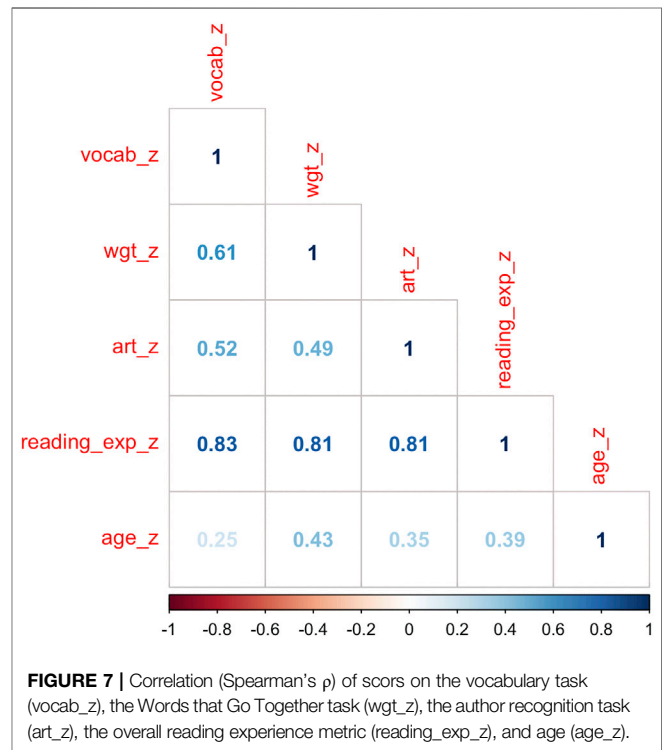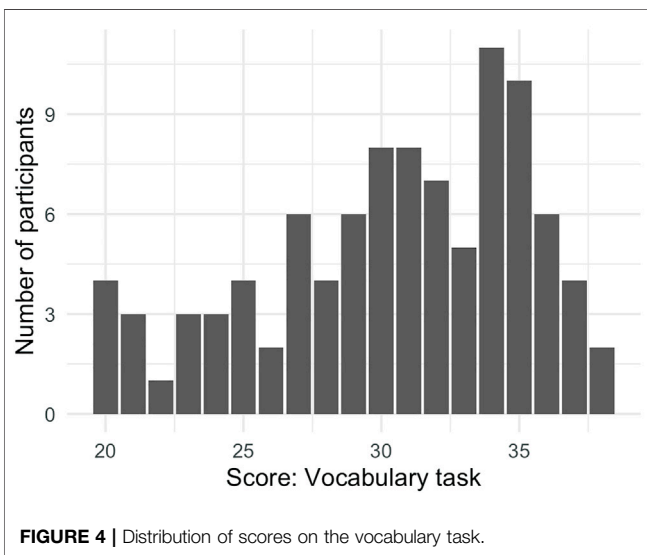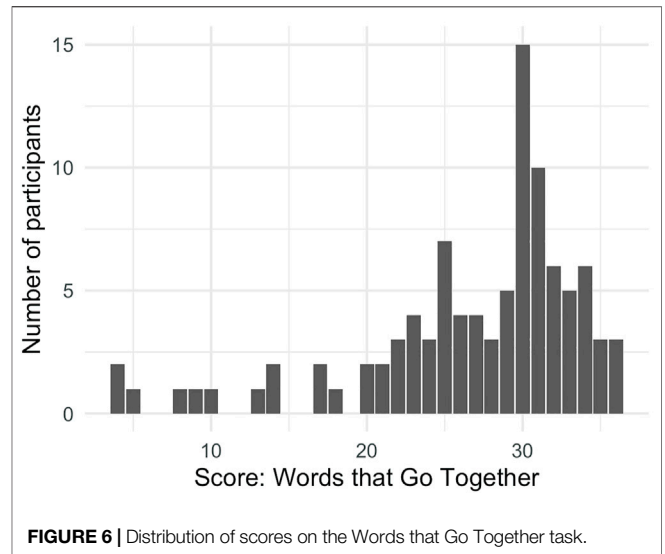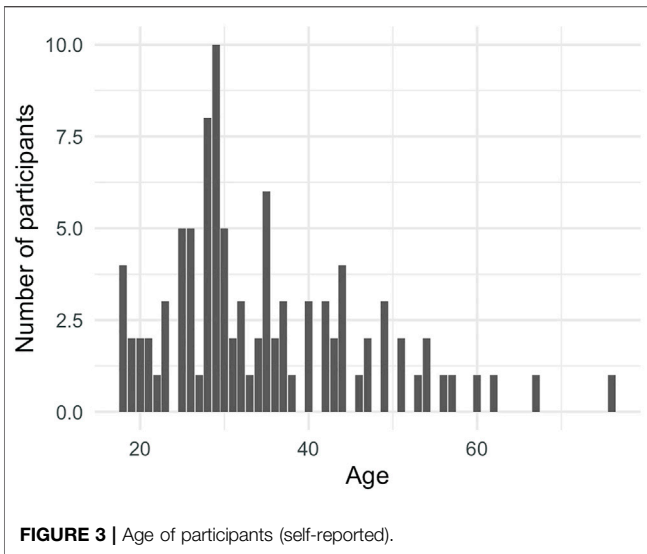
Sentence onsets were created to be maximally semantically neutral to avoid priming or prediction before the reader encountered the first word in the bigram (the modifier). To prevent uncontrolled semantic relations or priming, role nouns were excluded from sentence heads—instead, proper names, group nouns, or occasionally, inanimate subjects were used. Names were not repeated and the

repetition of other lexical items was also restricted except when it came to highly frequent items (e.g., "little"). Sentences were further designed to have the critical bigram in different semantic and syntactic roles, to reduce predictability of stimuli. This means that in some items, the bigram took the role of the subject, in others it was the direct or indirect object, and in others still, it functioned as part of a fronted subordinate clause. Sentences varied in length from 8 to 21 words (median: 13) and did not contain words with regional spellings.

After collecting data but before analyzing results, 23 items from the synonym condition were removed because they were unattested in COCA (for example, "cozy parka," "checkered paper") and six items were removed because they (unintentionally) had the same first word as another item.

Self-Paced Reading

Not including the removed items, participants read 259 sentences. Approximately 33% of sentences were followed by a multiple-choice comprehension question, which usually had three options expect for small proportion in a yes/no format (16%). Questions were designed to probe all levels of processing, from the surface recall of proper names and lexical items or phrases (Who was inspired by the dress? "Katrina," "Catherine," "Louisa") to overall comprehension of the sentence (Why was the building evacuated? "a part collapsed," "someone saw smoke,"

FIGURE 3 | Age of participants (self-reported).

FIGURE 4 | Distribution of scores on the vocabulary task.

FIGURE 5 | Distribution of scores on the Author Recognition task.

FIGURE 6 | Distribution of scores on the Words that Go Together task.

FIGURE 7 | Correlation (Spearman's ρ) of scors on the vocabulary task (vocab_z), the Words that Go Together task (wgt_z), the author recognition task (art_z), the overall reading experience metric (reading_exp_z), and age (age_z).

"renovations were being made") and higher-level inferences (Did the conversations lead to a solution? "yes," "no").

Participants were instructed to read as normally as possible and were informed that the sentences were not related to each other. They were told to answer questions with the best answer, and to keep short breaks (such as drinking water) to the question screens, not the reading sections. The reading section was estimated to take about 45 min.

TABLE 2 | Model output, linear mixed effects model with forward transition probability.

Linear mixed model fit by maximum likelihood

logRT ~ 1 + FTP_lz + reading_exp_z + age_z + position + trial_number_z + word_number_z + length_z + prev_length_z + education + FTP_lz & reading_exp_z + FTP_lz & age_z + reading_exp_z & age_z + FTP_lz & position + reading_exp_z & position + age_z & position + FTP_lz & reading_exp_z & age_z + FTP_lz & reading_exp_z & position + FTP_lz & age_z & position + reading_exp_z & age_z & position + FTP_lz & reading_exp_z & age_z & position + (1 + trial_number_z + length_z + prev_length_z | id) + (1 + trial_number_z | w1) + (1 + trial_number_z | w2)

	logLik	−2 logLik	AIC	AICc	BIC
	−2001.35	4,002.70	4,098.70	4,098.77	4,540.67

Variance components

		Variance	Std.Dev.	Corr.		
id	(Intercept)	0.0723	0.2689			
	trial_number_z	0.0037	0.0608	−0.25		
	length_z	0.0005	0.0232	0.65	−0.02	
	prev_length_z	0.0008	0.0278	0.5	0.1	0.53
w2	(Intercept)	0.0002	0.0138			
	trial_number_z	0.0001	0.0093	−0.42		
w1	(Intercept)	0.0004	0.0200			
	trial_number_z	0.0001	0.0082	0.08		
Residual		0.0603	0.2456			

Number of obs: 73,690; levels of grouping factors: 97, 143, 135

Fixed-effects parameters

	Coef.	Std. Error	z	Pr(>|z|)
(Intercept)	5.7113	0.0293	195.0300	< 0.0001
FTP_lz	−0.0027	0.0018	−1.4600	0.1437
reading_exp_z	−0.0305	0.0235	−1.3000	0.1947
age_z	0.0977	0.0235	4.1500	< 0.0001
position: noun	−0.0080	0.0021	−3.8000	0.0001
position: spillover_1	0.0158	0.0015	10.4900	< 0.0001
trial_number_z	−0.1125	0.0063	−17.7800	< 0.0001
word_number_z	−0.0060	0.0014	−4.2800	< 0.0001
length_z	0.0123	0.0027	4.6200	< 0.0001
prev_length_z	0.0266	0.0031	8.4600	< 0.0001
education: Trade school	−0.0980	0.0312	−3.1400	0.0017
education: Undergraduate	−0.0008	0.0145	−0.0500	0.9581
education: Grad school	0.0111	0.0143	0.7800	0.4366
FTP_lz & reading_exp_z	0.0017	0.0012	1.4400	0.1494
FTP_lz & age_z	−0.0026	0.0012	−2.2600	0.0240
reading_exp_z & age_z	0.0019	0.0198	0.1000	0.9229
FTP_lz & position: noun	0.0059	0.0014	4.3300	< 0.0001
FTP_lz & position: spillover_1	−0.0008	0.0014	−0.6200	0.5351
reading_exp_z & position: noun	−0.0102	0.0019	−5.4500	< 0.0001
reading_exp_z & position: spillover_1	0.0038	0.0016	2.4100	0.0160
age_z & position: noun	0.0048	0.0019	2.5800	0.0100
age_z & position: spillover_1	-0.0011	0.0016	−0.6800	0.4980
FTP_lz & reading_exp_z & age_z	0.0012	0.0010	1.2500	0.2125
FTP_lz & reading_exp_z & position: noun	0.0008	0.0015	0.5200	0.6029
FTP_lz & reading_exp_z & position: spillover_1	0.0002	0.0015	0.1600	0.8716
FTP_lz & age_z & position: noun	0.0007	0.0015	0.4700	0.6389
FTP_lz & age_z & position: spillover_1	−0.0008	0.0015	-0.5500	0.5828
reading_exp_z & age_z & position: noun	−0.0008	0.0016	−0.5000	0.6169
reading_exp_z & age_z & position: spillover_1	−0.0013	0.0013	−0.9800	0.3255
FTP_lz & reading_exp_z & age_z & position: noun	0.0003	0.0012	0.2300	0.8192
FTP_lz & reading_exp_z & age_z & position: spillover_1	0.0004	0.0012	0.3100	0.7580

Participants

Participants were recruited online *via* Prolific (Damer and Bradley, 2018). Submissions for which less than 80% of SPR comprehension questions were answered correctly were not considered in the analysis. Recruitment stopped when 100 participants with useable submissions were reached; however, only 9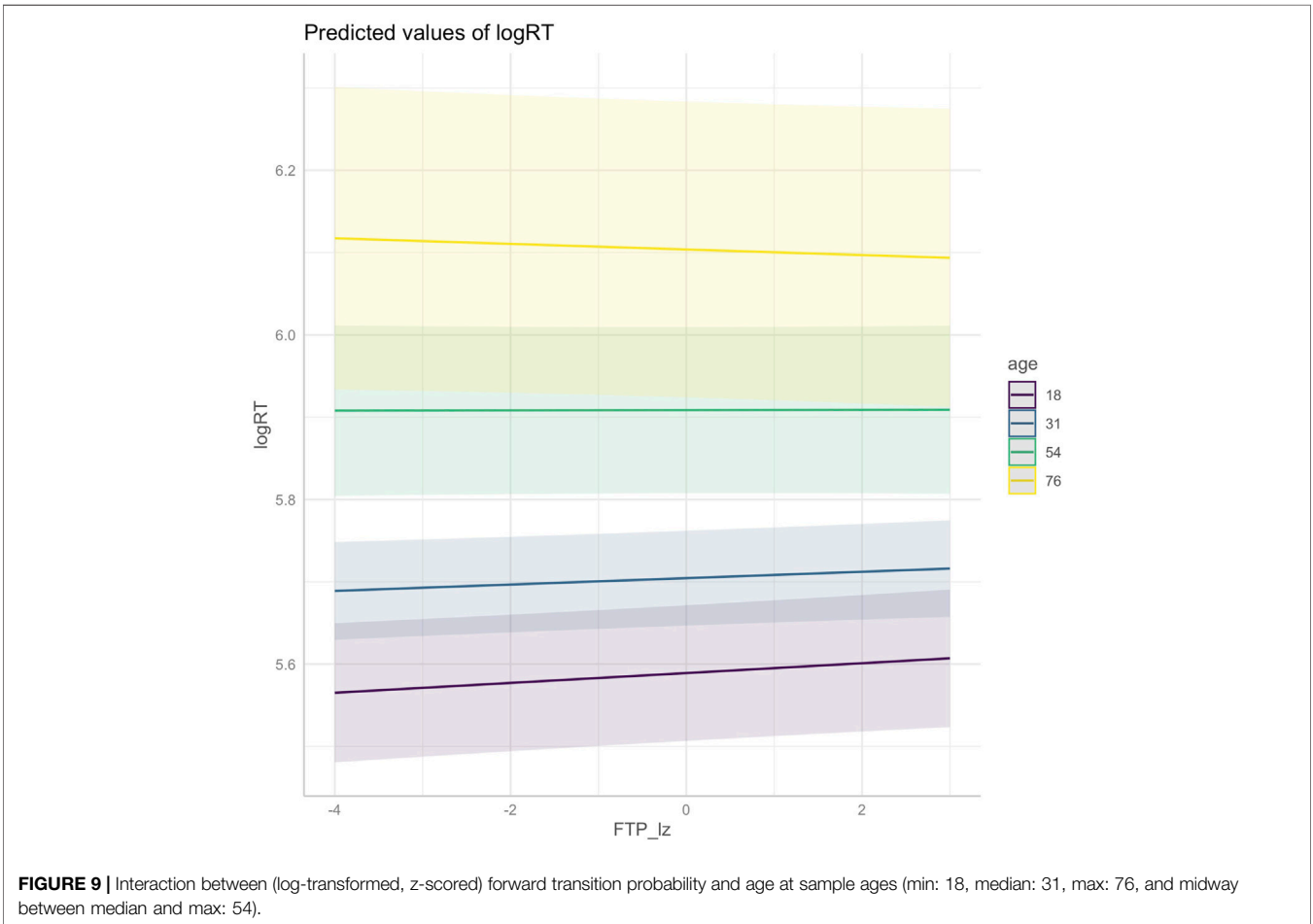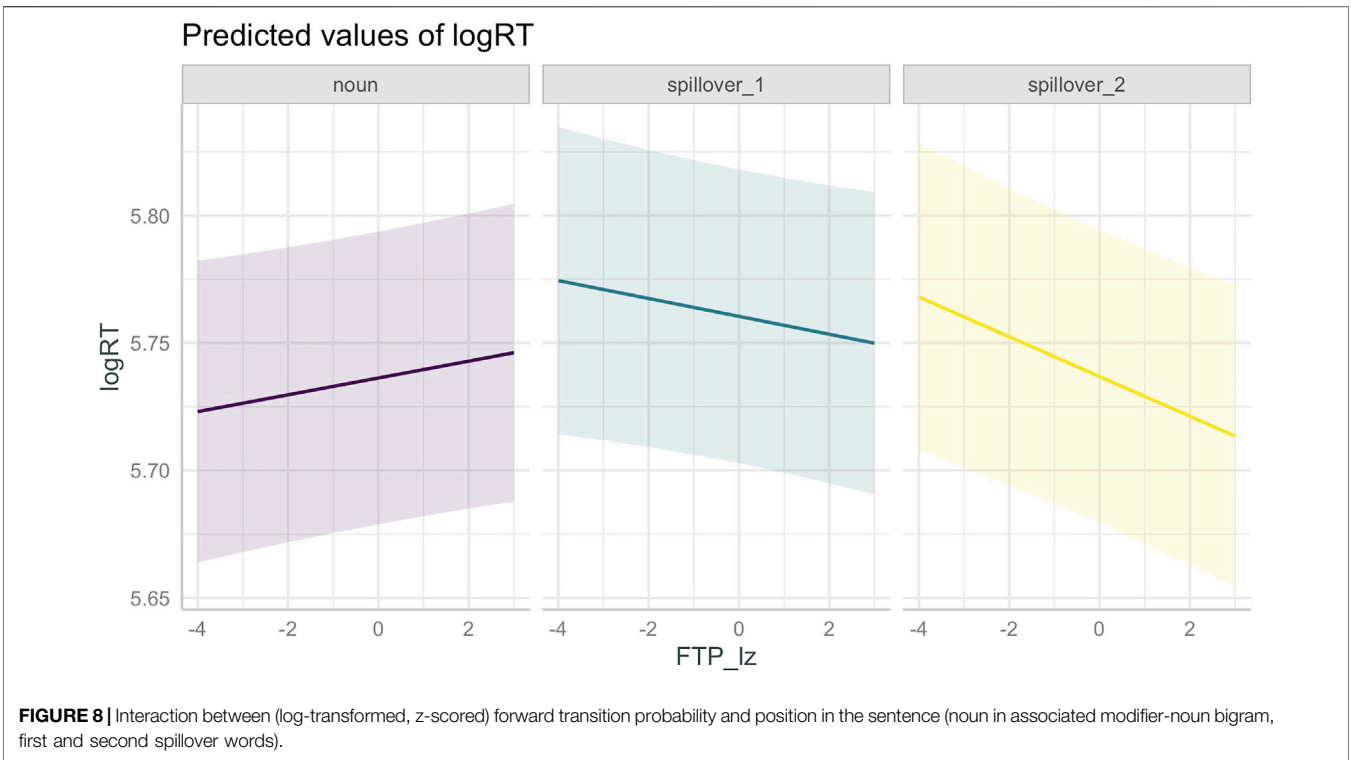7 participants were used in the current study because two did not complete the vocabulary task and one received a score of only 28%. 56 participants were female. Education was measured on a 4-point scale: 19 participants had at least some high school education, 21 had attended technical or trade school, 43 had a bachelor's degree and 14 had an advanced degree (Master's or Ph.D.). All participants were required to be monolingual native speakers of English; 60
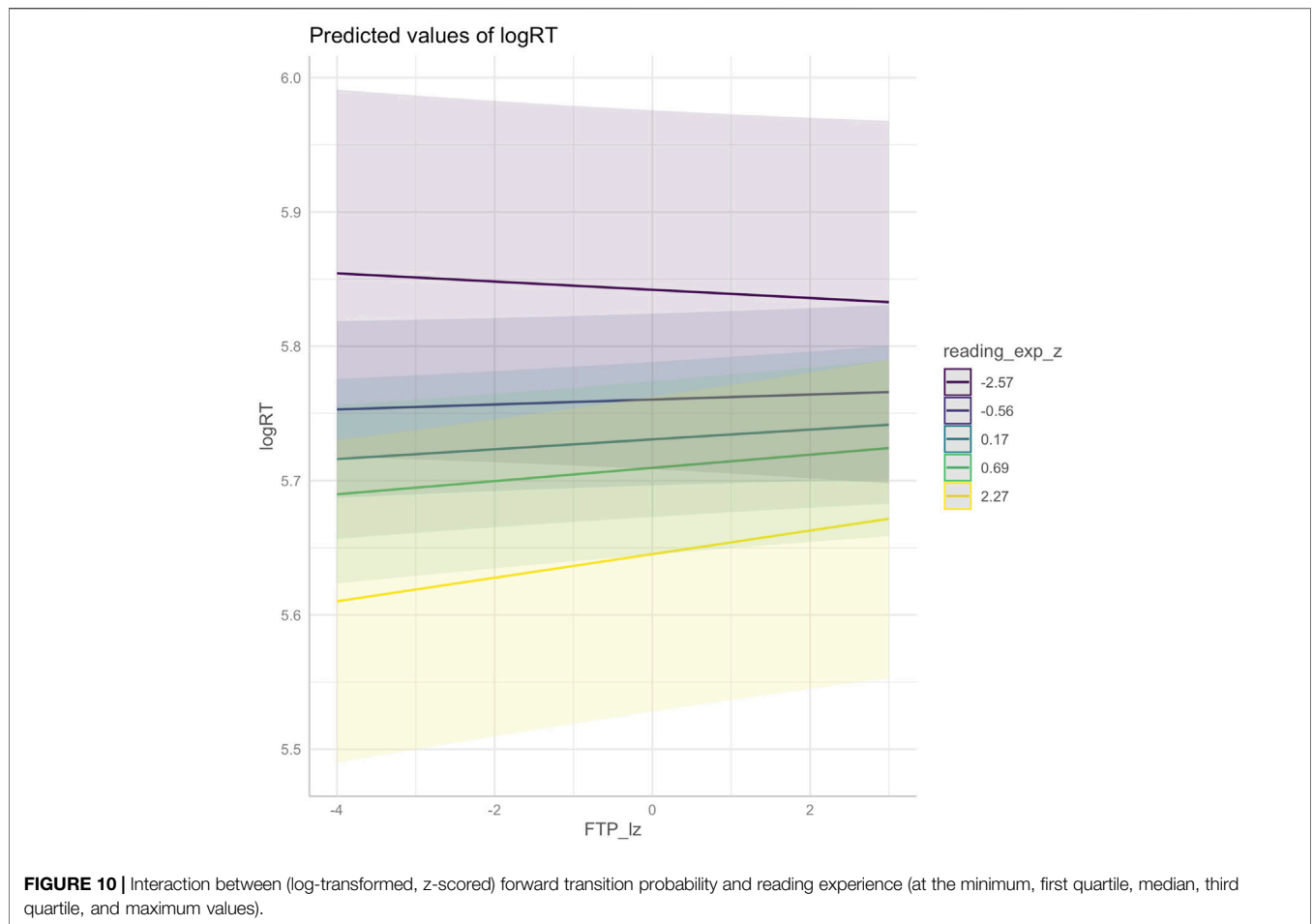
TABLE 3 | Model output linear mixed effects model with backward transition probability.

Linear mixed model fit by maximum likelihood

logRT ~ 1 + BTP_lz + reading_exp_z + age_z + position + trial_number_z + word_number_z + length_z + prev_length_z + education + BTP_lz & reading_exp_z + BTP_lz & age_z + reading_exp_z & age_z + BTP_lz & position + reading_exp_z & position + age_z & position + BTP_lz & reading_exp_z & age_z + BTP_lz & reading_exp_z & position + BTP_lz & age_z & position + reading_exp_z & age_z & position + BTP_lz & reading_exp_z & age_z & position + (1 + trial_number_z + length_z + prev_length_z | id) + (1 + trial_number_z | w1) + (1 + trial_number_z | w2)

	logLik	−2 logLik	AIC	AICc	BIC		
	−2006.17	4,012.34	4,108.34	4,108.41	4,550.31		

Variance components

		Variance	Std.Dev.	Corr.			
id	(Intercept)	0.0723	0.2689				
	trial_number_z	0.0037	0.0609	−0.24			
	length_z	0.0005	0.0232	0.65	−0.02		
	prev_length_z	0.0008	0.0278	0.5	0.1		
						0.52	
w2	(Intercept)	0.0002	0.0145				
	trial_number_z	0.0001	0.0093	−0.36			
w1	(Intercept)	0.0003	0.0186				
	trial_number_z	0.0001	0.0082	0.06			
Residual		0.0603	0.2457				

Number of obs: 73,690; levels of grouping factors: 97, 143, 135

Fixed-effects parameters

	Coef.	Std. Error	Z	Pr(>|z|)
(Intercept)	5.7106	0.0293	194.9700	< 0.0001
BTP_lz	−0.0052	0.0018	−2.8700	0.0040
reading_exp_z	−0.0289	0.0235	−1.2300	0.2200
age_z	0.0954	0.0236	4.0500	< 0.0001
position: noun	−0.0091	0.0021	−4.3300	< 0.0001
position: spillover_1	0.0158	0.0015	10.4300	< 0.0001
trial_number_z	−0.1125	0.0063	−17.7500	< 0.0001
word_number_z	−0.0059	0.0014	−4.2400	< 0.0001
length_z	0.0128	0.0027	4.7900	< 0.0001
prev_length_z	0.0278	0.0032	8.8100	< 0.0001
education: Trade school	−0.0979	0.0312	−3.1300	0.0017
education: Undergraduate	−0.0008	0.0146	−0.0500	0.9562
education: Grad school	0.0113	0.0143	0.7900	0.4318
BTP_lz & reading_exp_z	0.0024	0.0011	2.1300	0.0336
BTP_lz & age_z	−0.0022	0.0011	−2.0000	0.0452
reading_exp_z & age_z	0.0034	0.0198	0.1700	0.8630
BTP_lz & position: noun	0.0005	0.0014	0.3500	0.7251
BTP_lz & position: spillover_1	−0.0031	0.0014	−2.2400	0.0251
reading_exp_z & position: noun	−0.0100	0.0019	−5.3500	< 0.0001
reading_exp_z & position: spillover_1	0.0036	0.0016	2.2600	0.0241
age_z & position: noun	0.0045	0.0019	2.4100	0.0160
age_z & position: spillover_1	−0.0010	0.0016	−0.6300	0.5305
BTP_lz & reading_exp_z & age_z	0.0007	0.0009	0.7700	0.4395
BTP_lz & reading_exp_z & position: noun	0.0023	0.0015	1.5500	0.1202
BTP_lz & reading_exp_z & position: spillover_1	−0.0010	0.0015	−0.6800	0.4939
BTP_lz & age_z & position: noun	0.0002	0.0015	0.1500	0.8840
BTP_lz & age_z & position: spillover_1	−0.0007	0.0015	-0.4800	0.6311
reading_exp_z & age_z & position: noun	−0.0006	0.0016	−0.3900	0.6932
reading_exp_z & age_z & position: spillover_1	−0.0014	0.0013	−1.0500	0.2939
BTP_lz & reading_exp_z & age_z & position: noun	0.0013	0.0013	1.0500	0.2957
BTP_lz & reading_exp_z & age_z & position: spillover_1	0.0009	0.0012	0.7000	0.4833

FIGURE 8 | Interaction between (log-transformed, z-scored) forward transition probability and position in the sentence (noun in associated modifier-noun bigram, first and second spillover words).

FIGURE 9 | Interaction between (log-transformed, z-scored) forward transition probability and age at sample ages (min: 18, median: 31, max: 76, and midway between median and max: 54).

FIGURE 10 | Interaction between (log-transformed, z-scored) forward transition probability and reading experience (at the minimum, first quartile, median, third quartile, and maximum values).

participants were British nationals and 37 were United States nationals[4].

Participant age ranged from 18 to 76 (mean: 34.6, median: 31, SD: 11.7), see **Figure 3**. By recruiting participants within this range, we look specifically at age effects during general adulthood without considering old age or youth **Figure 4**. This allows for maximally comparable participants that are not in the prime years of acquisition but also not so advanced in age as to be incomparable in terms of sensory and cognitive ability as well as technological savvy. We expected age effects to appear in these mid-range adult years based on prior research showing that cognitive abilities vary across the entire adult lifespan, not only at the extremes (Hartshorne and Germine, 2015).
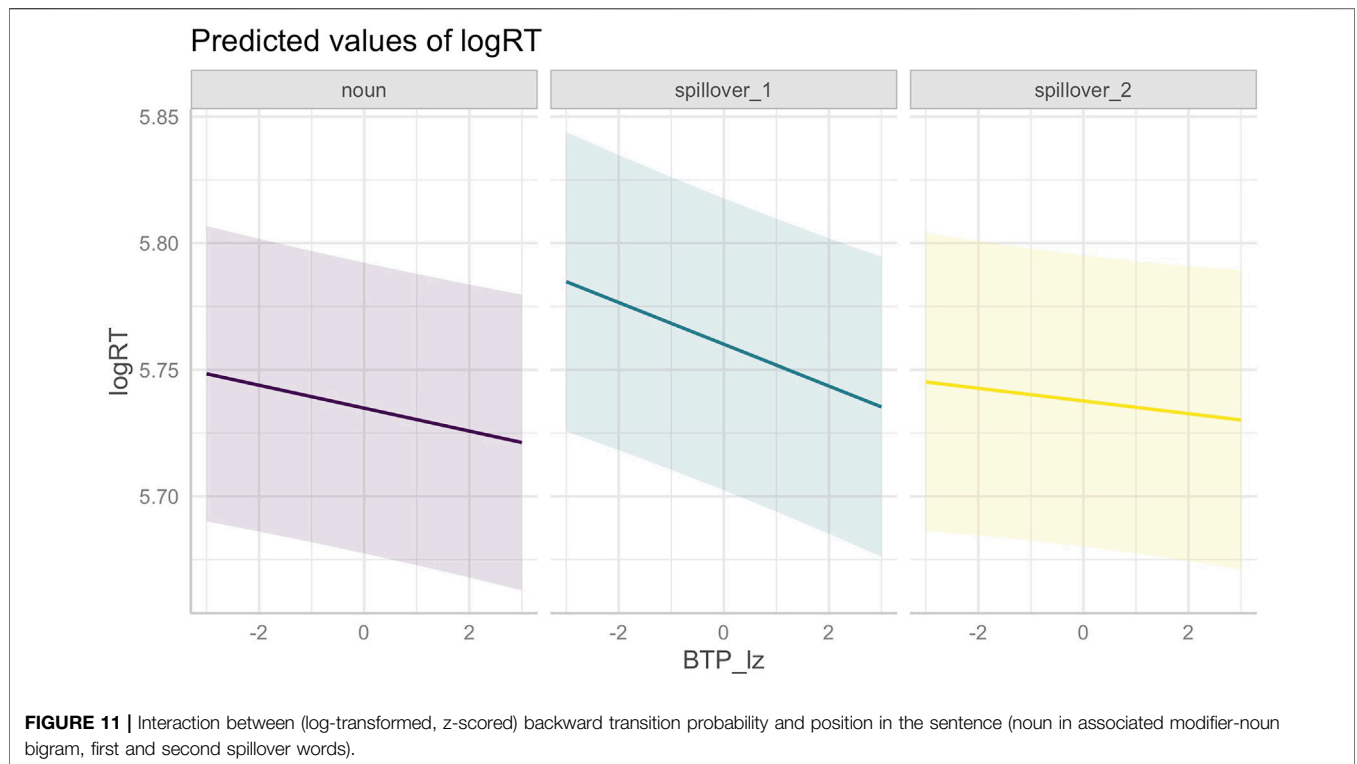
Individual Difference Assessments

Before starting with any tasks, participants provided their age, nationality and level of education **Figure 5**. They completed the self-paced reading section as one block and a total of 10 individual difference assessments in another block; both blocks were completed in the same sitting with the option for a break in between and the order of the blocks as well as the tasks within the individual difference block were counter-balanced over subjects **Figure 6**. Three reading-based individual differences are presented below; the other tasks were used separately for a different experiment.

Vocabulary Task

To assess vocabulary, participants completed a timed version of the Shipley Institute of Living scale, in which they had to select the best synonym for a given word (abbrv. "vocab" below) (Shipley, 1940)[5]. For example, participants were asked to select a word that meant the same as "fortify" from the set: "strengthen," "submerge," "vent", "deaden". Participants were told not to use any other browser tabs and were presented the questions in a randomized order

---

[4]The reference corpus is composed of American English; however, the stimuli were designed to not be regionally specific. To ensure that the choice of corpus did not affect results in a way that preferred participants who were United States nationals, we ran a post-hoc model which included an additional fixed effects term for nationality (dummy-coded with United States as the reference level), as well as an interaction term for FTP/BTP by nationality. Neither term was statistically significant. See supplementary materials for the full model output: https://github.com/kyla-mcconnell/inddiff_experience

[5]http://www.musicianbrain.com/documentation/Shipley_Vocab.pdf

FIGURE 11 | Interaction between (log-transformed, z-scored) backward transition probability and position in the sentence (noun in associated modifier-noun bigram, first and second spillover words).

with 8 s to select the best option. Participant scores were calculated as the number of sets for which the correct synonym was selected; they ranged from 20 to 38 correct answers out of 40 questions (median: 31).
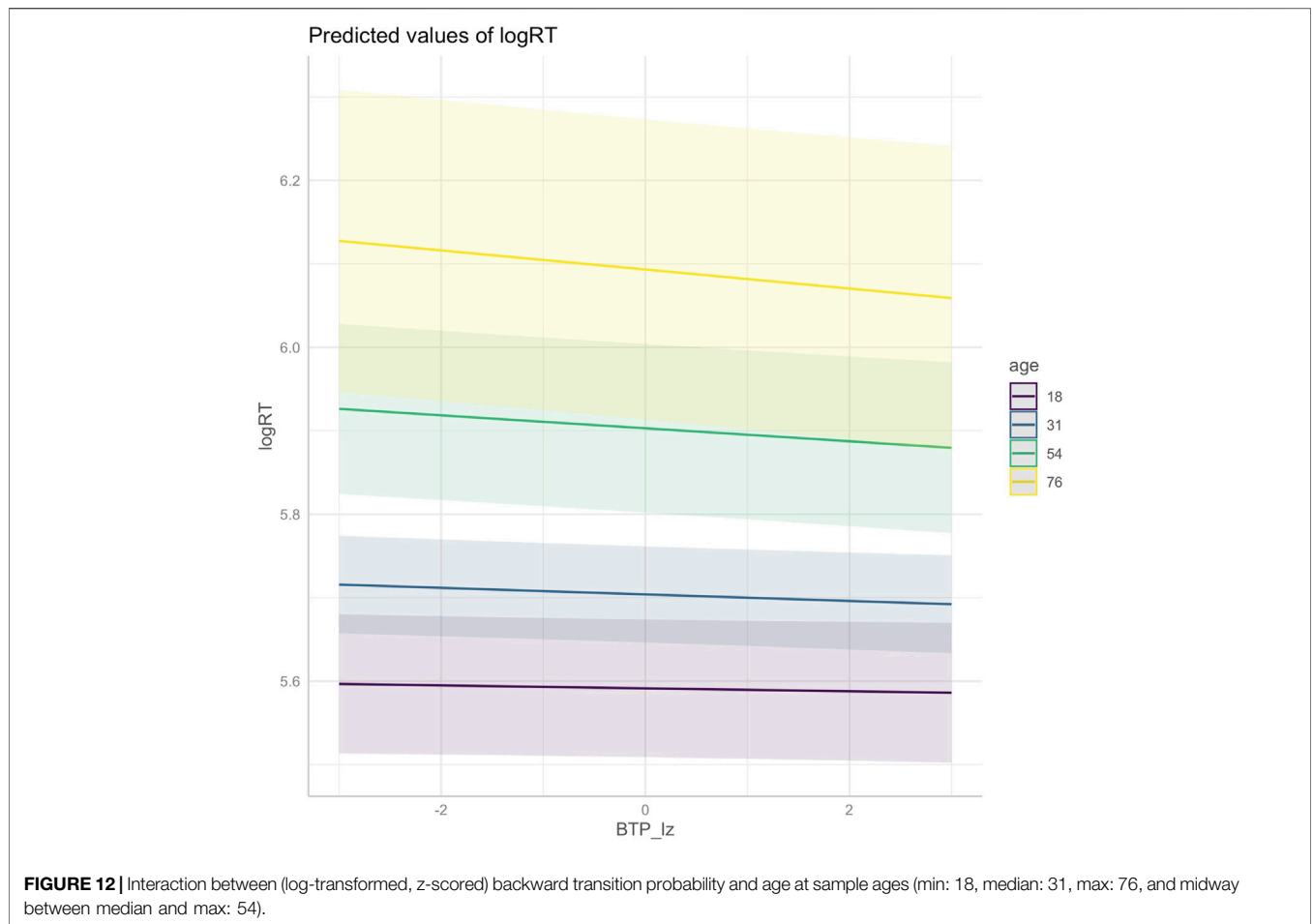
Author Recognition Task

Similarly, participants completed an author recognition task (abbrv. ART), a commonly used assessment of reading amount that circumvents inflated scores that tends to arise from a direct question. By asking participants to identify authors, but not directly asking how much they read, those who see reading as a favorable pastime, but who do not read often themselves, are less able to inflate their own scores. However, it should be kept in mind that participants from social groups that are well-read or educated may also know more authors names without having read any of the texts from those authors. In the current experiment, the ART task contained 65 authors names from Acheson et al., 2008, an update on Stanovich and West, 1989. Though Acheson and colleagues tested the list on a United States undergraduate audience, the median score across all participants in the United Kingdom was 7.5 points higher in our sample (57.5 vs. 50.0). Because many of the real author names were recognized by less than 10% of the undergraduates tested, just 40 additional foils were taken from Mar et al. (2006), which had a similarly structured task and openly published these names. Participants scores were calculated as the total number of correct answers minus a point for each incorrect guess and ranged from 22 to 90 (median: 56).

Words That Go Together Task

Finally, participants were also presented the Words that Go Together task introduced in Dąbrowska, 2015 (abbrv. WGT). The task had 40 multiple-choice questions, each of which contained five collocated bigrams, where the participant was asked to select the pair that "sounds the most natural or familiar." Participants were further instructed to guess if they didn't know which pair sounded best. Sets varied in frequency; one example is: "hazard a guess," "chance a guess," "dare a guess," "gamble a guess," "risk a guess". Participants scores were the number of pairs for which they chose the most highly collocated pair and ranged from 4 to 36 correct answers (median: 29). Across median scores by country, participants from the United Kingdom scored five points higher than United States participants in our sample (30 vs. 25). However, the participants from the United Kingdom scored higher on the US-based ART, this does not seem concerning in terms of a bias of the current task towards United Kingdom Englishes.

All reading-based tasks were positively correlated with age, with the Author Recognition task the most strongly correlated with increasing age. More importantly, the scores on all reading assessments were all strongly correlated with each other (see **Figure 7**). This is an expected effect, as it may reflect a common or at least similar mechanism behind all measures of reading experience. Because both theoretical assumptions and the quantitative correlation suggest that the measures capture the same variance, we used all constructs together as one metric of reading exposure, by adding all centered and scaled individual scores (vocab + ART + WGT) and then again centering and scaling the result across all participants. This doesn't allow us to draw conclusions about specific score types that may be

FIGURE 12 | Interaction between (log-transformed, z-scored) backward transition probability and age at sample ages (min: 18, median: 31, max: 76, and midway between median and max: 54).

driving the effect but further levels of specificity are not necessary to our hypothesis.

Modeling

Data was first prepared in R: in this stage, reading times were joined to frequency data extracted from COCA and word and bigram frequencies were log-transformed. RTs below 100 m or (subsequently) outside of 2.5 standard deviations of each participant's individual mean were removed[6] (total: 2.2%). RTs were log-transformed and then centered and scaled. All other numeric variables were also centered and scaled. The prepared data was then loaded into Julia and analyzed with linear mixed effects models using MixedModels.jl (Bates et al., 2021) and plotted in R with the ggeffects package via JellyMe4.jl.
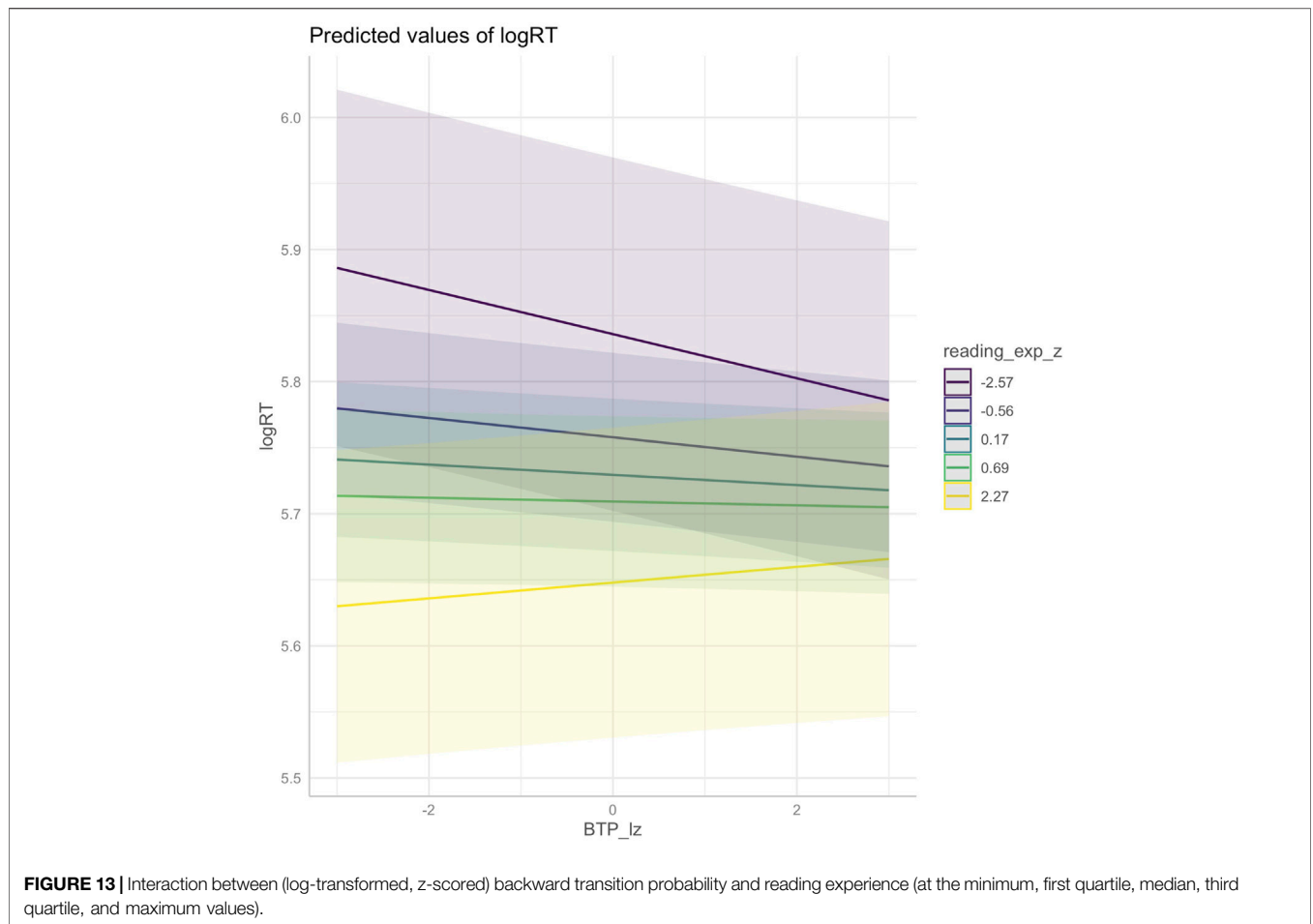
The critical word was defined as the noun in the bigram, because at this point, participants were able to recognize the completed bigram. Our previous research showed that an association effect develops over the spillover region (the critical word and the two following words), so two words following the noun were also

included and the variable "position" was sum-coded with three levels (McConnell and Blumenthal-Dramé, 2019).

Random slopes were calculated on three grouping variables: participant, the first word in the bigram (W1) and the second word in the bigram (W2). Grouping on W1 pairs bigrams paired items based on the automatic frequency-matched bigram extraction. Grouping on W2 pairs bigrams allowed for bigrams to vary based on semantic similarity (see **Table 1**). To control for commonly observed self-paced reading effects, trial number, word length[7] and (effects-coded) position in sentence

---

[6]Using participant means and SDs instead of an absolute upper bound was designed to not disadvantage slower readers, many of whom are older or have less reading experience.

[7]Word length and previous length were used (instead of the commonly used covariate of word frequency), because current and previous word frequencies are essential components to calculating forward and backward transition probability. Since the model covers a 3-word span from word 2 (the critical word, where the bigram can be identified) to two words after the critical word, using word frequency would create the situation that for some words in some models, information about current and/or previous word frequencies is already included in the BTP/FTP predictor, but for other words it isn't. To ensure that this was the correct design decision, we ran a post-hoc LMM comparing a model using word lengths to one using word frequencies and found that the model using length was better fit and that coefficients and statistical significance was essentially unchanged. See supplementary materials for full model: https://github.com/kyla-mcconnell/inddiff_experience

FIGURE 13 | Interaction between (log-transformed, z-scored) backward transition probability and reading experience (at the minimum, first quartile, median, third quartile, and maximum values).

were included as covariates. We additionally control for level of education, which was Helmert coded as a factor with four levels.

Because forward and backward transition probability were strongly correlated (ρ = 0.53) and our predictions for them were disparate, they were not included in the same model. In a first step, two maximal models were fit with the same fixed effects structure, one with forward transition probability and one with backward transition probability. The random effects structure of the maximal models were assessed via PCA and were initially overfit, as expected. Without looking at fixed effects coefficients or *p*-values, the random effects structure was first reduced; terms that described a small portion of the variance in both models were removed to reduce overfitting until rePCA results were satisfactory (i.e., all random effects terms added to the variance described) (Matuschek et al., 2017; Bates et al., 2018). After fitting, model diagnostic plots were visually inspected and no major issues were detected. Analysis files are available in a Github repository: https://github.com/kyla-mcconnell/inddiff_experience

RESULTS

For each model, we fit a four-way interaction between age, reading experience, position, and FTP/BTP, as we hypothesized that the

effect of experience may change over the critical region as a factor of these two types of experience. However, no three- or four-way interactions were statistically significant.

In both the model with forward transition probability (**Table 2**) and the model with backward transition probability (**Table 3**), slower responses followed longer current or previous words (length_z, prev_length_z) and proceeding towards the end of the experiment (trial_number_z) or the end of the sentence (word_number_z) led to faster response times, as expected. In terms of experience, older participants (age_z) had slower response times in general, but there was no main effect of reading experience (reading_exp_z) in either model. In both models, higher transition probabilities, whether forward (FTP) or backward (BTP), led to quicker reading times in general; however, only the main effect of BTP was statistically significant (β = -0.0052 (SE: 0.0018), *p* = 0.0040[8]).

---

[8]A β-value of −0.0052 means that for an increase of one standard deviation in the predictor (BTP), the outcome variable (log-transformed RT) decreases by 0.0052 standard deviations. For response times in this sample, one standard deviation corresponds to approximately 149.5 m. Thus, the effect of a change in BTP by one standard deviation (i.e., a change of approximately 0.01) results in a response time decrease of about 0.8 m, all other covariates held constant.

Response times to the noun in the critical bigram (position: noun) also tended to be faster than to the word immediately following it (position: spillover_1). The effect of transition probability also varied by position: Higher FTP correlated with faster response times in the spillover region but slower response time on the noun (β = 0.0059 (SE: 0.0014), p = < 0.001, **Figure 8**). Higher BTP also correlated with faster response times particularly on the first spillover word (β = −0.0031 (SE: 0.0014), p = 0.0251, **Figure 11**).

In terms of experience-based predictors, both FTP and BTP had significant interactions with age (β = −0.0026 (SE: 0.0012), p = 0.0240 and β = −0.0022 (SE: 0.0011), p = 0.0452, respectively). Both also interact with reading experience, though this is only statistically significant for BTP (β = 0.0024 (SE: 0.0011), p = 0.0336; FTP: β = 0.0017 (SE: 0.0012), p = 0.1494). Interestingly, age and reading experience show opposite effects: age enhances the speeding up effects associated with transition probability, while greater reading experience reduces the effects. See the marginal effects plots in **Figures 9**–**13** for a visualization of these interactions.

DISCUSSION

In this self-paced reading experiment, we investigated probabilistic prediction and integration as parallel mechanisms in language processing, operationalized in terms of forward (FTP) and backward (BTP) transition probability. In stimuli controlled for bigram frequency, there was a significant main effect of BTP but no significant main effect of FTP, confirming our previous research, in which BTP was also a more suitable predictor of reading times to bigrams (McConnell and Blumenthal-Dramé, 2019). BTP and FTP also unfolded differently over the critical and spillover region in this dataset, highlighting that they capture at least partially unique reading mechanisms.

In the case of age, higher age correlated with an increased processing (speed) benefit for bigrams with higher transition probabilities. This is in line with previous research, which has consistently found that older adults rely more heavily on predictive processing and have more entrenched probabilistic knowledge. Older participants also consistently read more slowly than younger participants in this study, which may have allowed them more time for lexical access to individual words and the distributional statistics associated with them (an effect that could even be heightened in a self-paced reading set-up).

We initially considered two explanations for why older readers may process more probabilistically: because they have greater experience with language and world knowledge in general, or because they are compensating for perceptual decline and cognitive slowing. In isolation, the results for age do not answer this question; however, the effect of reading experience provides insight to whether an increased reliance on distributional knowledge is a luxury, in which case the most experienced readers should be better able to exploit

distributional information to facilitate processing, or if it is rather a compensatory mechanism, in which case participants with *less* experience should show greater effects of transition probability.

Although in our sample, greater age and increased reading experience were positively correlated (ρ = 0.39, see **Figure 7**), they show effects in different directions. Thus, the two individual difference factors can be disentangled and even have competing effects on language comprehension processes. Interestingly, participants with less reading experience show increased speeding up effects of BTP and FTP, which seems to contradict the previous research that has found that more reading experience makes a comprehender more predictive (see Section 1.1.1). However, the majority of these studies operationalized prediction via cloze scores, which may rely more heavily on general world knowledge, ability to effectively use context, or other mechanisms (McDonald and Shillcock, 2003a; McDonald and Shillcock, 2003b; Boston et al., 2008, but see Frisson et al., 2005).

Taken together, we find that those participants who may have the most need for compensation, either because of greater age or due to reduced experience with reading, exhibit a stronger speeding-up effect in the processing of modifier-noun bigrams with higher BTP, and, to a lesser extent, FTP. This could be tied to the finding that older age and less reading experience correlated with slower reading in general: On the one hand, slower reading could be a side effect of accessing distributional patterns such as FTP and BTP from long-term memory. On the other hand, perhaps this slowing is indicative of a need to devote more effort to word-level decoding and lexical access in general, and this effort intensifies the difference in processing bigrams with low vs. high transition probabilities. It could also simply be the case that our younger and more experienced groups were already at ceiling for transition probability effects in our stimuli.

Although we did not find a three-way interaction with transition probability, we did find that average readers are qualitatively different to highly skilled readers particularly in terms of the time course of effects, as previous research has also confirmed (Ashby et al., 2005). Bigrams with higher BTP had a general speeding up effect across the entire spillover region, but especially on the first spillover word. However, higher FTP elicited slower reading times on the noun and faster reading times on the first spillover word and slowing on the noun was particularly pronounced for both older and less experienced readers. Perhaps, then, slower reading allows more time for access to distributional information at the noun, as mentioned above. On the other hand, accessing distributional information at the noun may cost time for those who are older or weaker readers to begin with, but assist later in the spillover region. If this is the case, then probabilistic processing as a compensatory mechanism seems to be efficient in supporting processing.

There are of course limitations to the current design. The self-paced reading paradigm, for example, may have affected

some groups differentially; more experienced readers, younger readers, and potentially faster readers in general may have been disrupted by the somewhat unnatural constraints of self-paced reading, which don't allow for phenomena like word skipping, which is especially common in highly predictable phrases. By presenting lexical units separately, and disallowing parafoveal preview, the experimental paradigm may have disrupted normal reading, especially in highly associated bigrams that may naturally be read as multi-word units.

Additionally, the current research cannot speak to how TPs at different grain sizes interact (see Section 1.2). For example, if *discussion* is more predictable after *helpful* than after *careful* but the morpheme *-ful* is less predictable after *help-* than it is after *care-*[9], does the comparatively low morphological predictability within *helpful* undermine the comparatively high lexical predictability within *helpful discussion*? Or does the presumed chunk status of *helpful discussion* give cognitive precedence to the whole over the constituents parts (Blumenthal-Dramé, 2017)? Or do different levels of TP analysis have equal weight and balance each other off? And does the extent to which subjects rely on probabilistic cues at different levels depend on their world and reading experience? For example, word length effects suggest that beginning readers read words in a more letter-by-letter fashion than advanced readers, who adopt a more holistic reading style (Barton et al., 2014). Other research shows that in reading development, phonological awareness emerges before morphological awareness. This indicates that younger readers may track TPs on the letter level (and at the level of letter-sound correspondences), whereas more advanced readers might adopt a wider TP sampling window (e.g., inter-morphemic or inter-lexical TPs), which may or may not supersede lower-level knowledge (Mann and Singson, 2003).

Follow-up research is encouraged to consider TPs at more fine-grained levels of analysis (letters, phonemes, syllables, etc.), TPs between longer stretches of language [e.g. the predictability of the unit (*helpful discussion*), given the preceding context], and TPs at abstract levels of analysis [e.g., between the syntactic categories (Adj) and (N)] (Celata, 2020). Ultimately, it would also be desirable take into account TPs that cut across levels [e.g. between the morpheme *-ful* and the lexeme *discussion*, or between the lexeme *helpful* and the syntactic category (N)] and even TPs between context and language strings (e.g., how predictable is *helpful discussion* in an academic context?) Of course, this should be conducted from the perspective of individual differences, among the groups posited here as well as those not focused on in this study.

CONCLUSION

In conclusion, we found that older and less experienced readers show heightened effects of transition probability. Against this background, we suggested that increased reliance on probabilistic processing should be seen a compensatory mechanism rather than a luxury. We also found initial evidence for a different time course of effects based on both individual differences in experience and the reliance on either prediction (FTP) or integration (BTP), but more large-scale research is necessary to support this initial claim. Taken together, our research reveal that experience-based individual differences can critically affect the reliance on distributional statistics in language processing

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The study was approved by the ethics committee of Freiburg University. The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

Following the CRediT system (https://casrai.org/credit/). KMC: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data Curation, Writing – Original Draft, Writing – Review and Editing, Visualization, Funding acquisition. AB-D: Conceptualization, Methodology, Writing – Review and Editing, Supervision. Both authors contributed to the article and approved the submitted version.

FUNDING

ACKNOWLEDGMENTS

---

[9]This seems to be the case in COCA: *helpful discussion* (FTP 0.00299) vs. *careful discussion* (FTP 0.000292) but *help-ful* (FTP 0.04073) vs. *care-ful* (0.1230 FTP).

REFERENCES

Acheson, D. J., Wells, J. B., and MacDonald, M. C. (2008). New and Updated Tests of Print Exposure and reading Abilities in College Students. *Behav. Res.* 40, 278–289. doi:10.3758/BRM.40.1.278

Adelman, J. S., Sabatos-DeVito, M. G., Marquis, S. J., and Estes, Z. (2014). Individual Differences in reading Aloud: A Mega-Study, Item Effects, and Some Models. *Cogn. Psychol.* 68, 113–160. doi:10.1016/j.cogpsych.2013.11.001

Ambridge, B. (2020). Against Stored Abstractions: A Radical Exemplar Model of Language Acquisition. *First Lang.* 40, 509–559. doi:10.1177/0142723719869731

Ambridge, B., Kidd, E., Rowland, C. F., and Theakston, A. L. (2015). The Ubiquity of Frequency Effects in First Language Acquisition. *J. Child. Lang.* 42, 239–273. doi:10.1017/S030500091400049X

Andringa, S., and Dąbrowska, E. (2019). Individual Differences in First and Second Language Ultimate Attainment and Their Causes. *Lang. Learn.* 69, 5–12. doi:10.1111/lang.12328

Armeni, K., Willems, R. M., and Frank, S. (2017). Probabilistic Language Models in Cognitive Neuroscience: Promises and Pitfalls. *Neurosci. Biobehav Rev.* 83, 579–588. doi:10.1016/j.neubiorev.2017.09.001

Aslin, R. N. (2017). Statistical Learning: a Powerful Mechanism that Operates by Mere Exposure. *Wires Cogn. Sci.* 8, e1373. doi:10.1002/wcs.1373

Ashby, J., Rayner, K., and Clifton, C. (2005). Eye Movements of Highly Skilled and Average Readers: Differential Effects of Frequency and Predictability. *The Quarterly Journal of Experimental Psychology Section A* 58 (6), 1065–1086. doi:10.1080/02724980443000476

Aurnhammer, C., and Frank, S. L. (2019). Evaluating Information-Theoretic Measures of Word Prediction in Naturalistic Sentence reading. *Neuropsychologia* 134, 107198. doi:10.1016/j.neuropsychologia.2019.107198

Baayen, R. H. (2010). A Real experiment Is a Factorial experiment. *ML* 5, 149–157. doi:10.1075/ml.5.1.06baa

Barton, J. J. S., Hanif, H. M., Eklinder Björnström, L., and Hills, C. (2014). The Word-Length Effect in reading: A Review. *Cogn. Neuropsychol.* 31, 378–412. doi:10.1080/02643294.2014.895314

Bates, D., Alday, P., Kleinschmidt, D., Calderón, J. B. S., Zhan, L., Noack, A., et al. (2021). JuliaStats/MixedModels.jl: v3.4.1. Zenodo. doi:10.5281/ZENODO.4613347

Bates, D., Alday, P., Kleinschmidt, D., Calderón, J. B. S., Zhan, L., Noack, A., et al. (2021). *JuliaStats/MixedModels.jl: v4.5.0 (v4.5.0)*. Zenodo. Available at: https://doi.org/10.5281/zenodo.5672870.

Bauer, L. (2004). Adjectives, Compounds, and Words. *Nordic J. English Stud.* 3, 7–22. doi:10.35360/njes.18

Biber, D. (1995). *Dimensions of Register Variation: A Cross-Linguistic Comparison*. Cambridge University Press.

Blumenthal-Dramé, A. (2016a). "6. Entrenchment from a Psycholinguistic and Neurolinguistic Perspective," in *Entrenchment and the Psychology of Language Learning How We Reorganize and Adapt Linguistic Knowledge* (Berlin, Boston: De Gruyter). doi:10.1515/9783110341423-007

Blumenthal-Dramé, A. (2017). "Entrenchment from a Psycholinguistic and Neurolinguistic Perspective," in *Entrenchment and the Psychology of Language Learning: How We Reorganize and Adapt Linguistic Knowledge*. Editor H.-J, Schmid (Washington DC: De Gruyter Mouton).

Blumenthal-Dramé, A., Glauche, V., Bormann, T., Weiller, C., Musso, M., and Kortmann, B. (2017). Frequency and Chunking in Derived Words: A Parametric fMRI Study. *J. Cogn. Neurosci.* 29, 1162–1177. doi:10.1162/jocn_a_01120

Blumenthal-Dramé, A. (2016b). What Corpus-Based Cognitive Linguistics Can and Cannot Expect from Neurolinguistics. *Cogn. Linguistics* 27, 493–505. doi:10.1515/cog-2016-0062

Boston, M. F., Hale, J., Kliegl, R., Patil, U., and Vasishth, S. (2008). Parsing Costs as Predictors of reading Difficulty: An Evaluation Using the Potsdam Sentence Corpus. *JEMR* 2. doi:10.16910/jemr.2.1.1

Brysbaert, M., Mandera, P., and Keuleers, E. (2018). The Word Frequency Effect in Word Processing: An Updated Review. *Curr. Dir. Psychol. Sci.* 27, 45–50. doi:10.1177/0963721417727521

Brysbaert, M., Stevens, M., Mandera, P., and Keuleers, E. (2016). How Many Words Do We Know? Practical Estimates of Vocabulary Size Dependent on Word Definition, the Degree of Language Input and the Participant's Age. *Front. Psychol.* 7. doi:10.3389/fpsyg.2016.01116

Carrol, G., and Conklin, K. (2019). Is All Formulaic Language Created Equal? Unpacking the Processing Advantage for Different Types of Formulaic Sequences Unpacking the Processing Advantage for Different Types of Formulaic Sequences. *Lang. Speech* 63, 95–122. doi:10.1177/0023830918823230

Celata, C. (2020). Bottom-up Probabilistic Information in Visual Word Recognition: Interactions with Phonological and Morphological Functions. *Lang. Sci.* 78, 101267. doi:10.1016/j.langsci.2019.101267

Choi, W., Lowder, M. W., Ferreira, F., Swaab, T. Y., and Henderson, J. M. (2017). Effects of Word Predictability and Preview Lexicality on Eye Movements during reading: A Comparison between aoung and Older Adults. *Psychol. Aging* 32, 232–242. doi:10.1037/pag0000160

Christiansen, M. H., and Arnon, I. (2017). More Than Words: The Role of Multiword Sequences in Language Learning and Use. *Top. Cogn. Sci.* 9, 542–551. doi:10.1111/tops.12274

Clark, A. (2013). Whatever Next? Predictive Brains, Situated Agents, and the Future of Cognitive Science. *Behav. Brain Sci.* 36, 181–204. doi:10.1017/S0140525X12000477

Çöltekin, Ç. (2017). Using Predictability for Lexical Segmentation. *Cogn. Sci.* 41, 1988–2021. doi:10.1111/cogs.12454

Dąbrowska, E. (2018). Experience, Aptitude and Individual Differences in Native Language Ultimate Attainment. *Cognition* 178, 222–235. doi:10.1016/j.cognition.2018.05.018

Dąbrowska, E. (2019). Experience, Aptitude, and Individual Differences in Linguistic Attainment: A Comparison of Native and Nonnative Speakers. *Lang. Learn.* 69, 72–100. doi:10.1111/lang.12323

Dąbrowska, E. (2015). Words that Go Together. *Ml* 9, 401–418. doi:10.1075/ml.9.3.02dab

Damer, E., and Bradley, P. (2018). Prolific. Available at: https://prolific.ac.

Dave, S., Brothers, T. A., Traxler, M. J., Ferreira, F., Henderson, J. M., and Swaab, T. Y. (2018). Electrophysiological Evidence for Preserved Primacy of Lexical Prediction in Aging. *Neuropsychologia* 117, 135–147. doi:10.1016/j.neuropsychologia.2018.05.023

Dehaene, S., Cohen, L., Morais, J., and Kolinsky, R. (2015). Illiterate to Literate: Behavioural and Cerebral Changes Induced by reading Acquisition. *Nat. Rev. Neurosci.* 16, 234–244. doi:10.1038/nrn3924

DeLong, K. A., Groppe, D. M., Urbach, T. P., and Kutas, M. (2012). Thinking Ahead or Not? Natural Aging and Anticipation during reading. *Brain Lang.* 121, 226–239. doi:10.1016/j.bandl.2012.02.006

DeLong, K. A., Urbach, T. P., and Kutas, M. (2005). Probabilistic Word Pre-activation during Language Comprehension Inferred from Electrical Brain Activity. *Nat. Neurosci.* 8, 1117–1121. doi:10.1038/nn1504

Demberg, V., and Keller, F. (2008). Data from Eye-Tracking Corpora as Evidence for Theories of Syntactic Processing Complexity. *Cognition* 109, 193–210. doi:10.1016/j.cognition.2008.07.008

Divjak, D. (2019). *Frequency in Language: Memory, Attention and Learning*. 1st ed. Cambridge University Press. doi:10.1017/9781316084410

Divjak, D., and Milin, P. (2020). Exploring and Exploiting Uncertainty: Statistical Learning Ability Affects How We Learn to Process Language along Multiple Dimensions of Experience. *Cogn. Sci.* 44. doi:10.1111/cogs.12835

Evans, N., and Levinson, S. C. (2009). The Myth of Language Universals: Language Diversity and its Importance for Cognitive Science. *Behav. Brain Sci.* 32, 429–448. doi:10.1017/S0140525X0999094X

Falkauskas, K., and Kuperman, V. (2015). When Experience Meets Language Statistics: Individual Variability in Processing English Compound Words. *J. Exp. Psychol. Learn. Mem. Cogn.* 41, 1607–1627. doi:10.1037/xlm0000132

Favier, S., Meyer, A. S., and Huettig, F. (2021). Literacy Can Enhance Syntactic Prediction in Spoken Language Processing. *J. Exp. Psychol. Gen.* doi:10.1037/xge0001042

Federmeier, K. D., and Kutas, M. (1999). A Rose by Any Other Name: Long-Term Memory Structure and Sentence Processing. *J. Mem. Lang.* 41, 469–495. doi:10.1006/jmla.1999.2660

Federmeier, K. D., Kutas, M., and Schul, R. (2010). Age-related and Individual Differences in the Use of Prediction during Language Comprehension. *Brain Lang.* 115, 149–161. doi:10.1016/j.bandl.2010.07.006

Federmeier, K. D. (2007). Thinking Ahead: The Role and Roots of Prediction in Language Comprehension. *Psychophysiology* 44, 491–505. doi:10.1111/j.1469-8986.2007.00531.x

Ferreira, F., and Chantavarin, S. (2018). Integration and Prediction in Language Processing: A Synthesis of Old and New. *Curr. Dir. Psychol. Sci.* 27, 443–448. doi:10.1177/0963721418794491

Frank, S. L., and Bod, R. (2011). Insensitivity of the Human Sentence-Processing System to Hierarchical Structure. *Psychol. Sci.* 22, 829–834. doi:10.1177/0956797611409589

Frank, S. L., Otten, L. J., Galli, G., and Vigliocco, G. (2015). The ERP Response to the Amount of Information Conveyed by Words in Sentences. *Brain Lang.* 140, 1–11. doi:10.1016/j.bandl.2014.10.006

Frank, S. L. (2013). Uncertainty Reduction as a Measure of Cognitive Load in Sentence Comprehension. *Top. Cogn. Sci.* 5, 475–494. doi:10.1111/tops.12025

Frisson, S., Rayner, K., and Pickering, M. J. (2005). Effects of Contextual Predictability and Transitional Probability on Eye Movements During Reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 31 (5), 862–877. doi:10.1037/0278-7393.31.5.862

Frost, R. L. A., Monaghan, P., and Christiansen, M. H. (2019). Mark My Words: High Frequency Marker Words Impact Early Stages of Language Learning. *J. Exp. Psychol. Learn. Mem. Cogn.* 45, 1883–1898. doi:10.1037/xlm0000683

Ghosh, V. E., and Gilboa, A. (2014). What Is a Memory Schema? A Historical Perspective on Current Neuroscience Literature. *Neuropsychologia* 53, 104–114. doi:10.1016/j.neuropsychologia.2013.11.010

Gordon, P. C., Lowder, M. W., and Hoedemaker, R. S. (2016). "Reading in Normally Aging Adults," in *Cognition, Language and Aging*. Editor H. H. Wright (Amsterdam: John Benjamins Publishing Company), 165–191. doi:10.1075/z.200.07gor

Gries, S. T., and Divjak, D. (2012). *Frequency Effects in Language Learning and Processing*. Berlin/Boston: Walter de Gruyter.

Griffiths, T. L., Vul, E., and Sanborn, A. N. (2012). Bridging Levels of Analysis for Probabilistic Models of Cognition. *Curr. Dir. Psychol. Sci.* 21, 263–268. doi:10.1177/0963721412447619

Gyllstad, H., and Wolter, B. (2016). Collocational Processing in Light of the Phraseological Continuum Model: Does Semantic Transparency Matter. *Lang. Learn.* 66, 296–323. doi:10.1111/lang.12143

Hale, J. (2016). Information-theoretical Complexity Metrics. *Lang. Linguistics Compass* 10, 397–412. doi:10.1111/lnc3.12196

Hanulíková, A., Ferstl, E. C., and Blumenthal-Dramé, A. (2020). Language Comprehension across the Life Span: Introduction to the Special Section. *Int. J. Behav. Develop.* 45, 379–381. doi:10.1177/0165025420954531

Hartshorne, J. K., and Germine, L. T. (2015). When Does Cognitive Functioning Peak? the Asynchronous Rise and Fall of Different Cognitive Abilities across the Life Span. *Psychol. Sci.* 26, 433–443. doi:10.1177/0956797614567339

Hartshorne, J. K., Skorb, L., Dietz, S. L., Garcia, C. R., Iozzo, G. L., Lamirato, K. E., et al. (2019). The Meta-Science of Adult Statistical Word Segmentation: Part 1. *Collabra: Psychol.* 5, 1. doi:10.1525/collabra.181

Havron, N., and Arnon, I. (2021). Starting Big: The Effect of Unit Size on Language Learning in Children and Adults. *J. Child. Lang.* 48, 244–260. doi:10.1017/S0305000920000264

H. Behrens and S. Pfänder (Editors) (2016). *Experience Counts: Frequency Effects in Language* (Berlin/Boston: Walter de Gruyter). doi:10.1515/9783110346916

Heine, B., Narrog, H., and Bod, R. (2015). "Probabilistic Linguistics," in *The Oxford Handbook of Linguistic Analysis*. Editors H. Narrog and B. Heine (Oxford University Press). doi:10.1093/oxfordhb/9780199677078.013.0025

Henderson, J. M., Choi, W., Lowder, M. W., and Ferreira, F. (2016). Language Structure in the Brain: A Fixation-Related fMRI Study of Syntactic Surprisal in reading. *NeuroImage* 132, 293–300. doi:10.1016/j.neuroimage.2016.02.050

Hersch, J., and Andrews, S. (2012). Lexical Quality and Reading Skill: Bottom-Up and Top-Down Contributions to Sentence Processing. *Scientific Stud. Reading* 16, 240–262. doi:10.1080/10888438.2011.564244

Hohwy, J. (2013). *The Predictive Mind*. 1st ed. Oxford, New York: Oxford University Press.

Howarth, P. A. (1996). *Phraseology in English Academic Writing: Some Implications for Language Learning and Dictionary Making*. Tübingen: M. Niemeyer.

Howarth, P. (1998). Phraseology and Second Language Proficiency. *Appl. Linguistics* 19, 24–44. doi:10.1093/applin/19.1.24

Huettig, F. (2015). Four central Questions about Prediction in Language Processing. *Brain Res.* 1626, 118–135. doi:10.1016/j.brainres.2015.02.014

Huettig, F., and Guerra, E. (2019). Effects of Speech Rate, Preview Time of Visual Context, and Participant Instructions Reveal strong Limits on Prediction in Language Processing. *Brain Res.* 1706, 196–208. doi:10.1016/j.brainres.2018.11.013

Huettig, F., and Janse, E. (2016). Individual Differences in Working Memory and Processing Speed Predict Anticipatory Spoken Language Processing in the Visual World. *Lang. Cogn. Neurosci.* 31, 80–93. doi:10.1080/23273798.2015.1047459

Huettig, F., and Mani, N. (2015). Is Prediction Necessary to Understand Language? Probably Not. *Lang. Cogn. Neurosci.* 31, 19–31. Available at: http://www.tandfonline.com/doi/abs/10.1080/23273798.2015.1072223 (Accessed May 30, 2017). doi:10.1080/23273798.2015.1072223

Huettig, F., and Mishra, R. K. (2014). How Literacy Acquisition Affects the Illiterate Mind - A Critical Examination of Theories and Evidence. *Lang. Linguistics Compass* 8, 401–427. doi:10.1111/lnc3.12092

Isbilen, E. S., McCauley, S. M., Kidd, E., and Christiansen, M. H. (2020). Statistically Induced Chunking Recall: A Memory-Based Approach to Statistical Learning. *Cogn. Sci.* 44. doi:10.1111/cogs.12848

Jolsvai, H., McCauley, S. M., and Christiansen, M. H. (2020). Meaningfulness Beats Frequency in Multiword Chunk Processing. *Cogn. Sci.* 44, e12885. doi:10.1111/cogs.12885

Kerz, E., and Wiechmann, D. (2021). "Individual Differences," in *The Routledge Handbook of Second Language Acquisition and Corpora*. Editors N. Tracy-Ventura and M. Paquot (Taylor & Francis), 394–406.

Kidd, E., Donnelly, S., and Christiansen, M. H. (2018). Individual Differences in Language Acquisition and Processing. *Trends Cogn. Sci.* 22, 154–169. doi:10.1016/j.tics.2017.11.006

Kidd, E., and Donnelly, S. (2020). Individual Differences in First Language Acquisition. *Annu. Rev. Linguist.* 6, 319–340. doi:10.1146/annurev-linguistics-011619-030326

Kliegl, R., Grabner, E., Rolfs, M., and Engbert, R. (2004). Length, Frequency, and Predictability Effects of Words on Eye Movements in reading. *Eur. J. Cogn. Psychol.* 16, 262–284. doi:10.1080/09541440340000213

Kukona, A., Braze, D., Johns, C. L., Mencl, W. E., Van Dyke, J. A., Magnuson, J. S., et al. (2016). The Real-Time Prediction and Inhibition of Linguistic Outcomes: Effects of Language and Literacy Skill. *Acta Psychologica* 171, 72–84. doi:10.1016/j.actpsy.2016.09.009

Kuperberg, G. R., and Jaeger, T. F. (2016). What Do We Mean by Prediction in Language Comprehension. *Lang. Cogn. Neurosci.* 31, 32–59. doi:10.1080/23273798.2015.1102299

Kuperman, V., and Van Dyke, J. A. (2011). Effects of Individual Differences in Verbal Skills on Eye-Movement Patterns during Sentence reading. *J. Mem. Lang.* 65, 42–73. doi:10.1016/j.jml.2011.03.002

Kuperman, V., and Van Dyke, J. A. (2013). Reassessing Word Frequency as a Determinant of Word Recognition for Skilled and Unskilled Readers. *J. Exp. Psychol. Hum. Perception Perform.* 39, 802–823. doi:10.1037/a0030859

Laubrock, J., Kliegl, R., and Engbert, R. (2006). SWIFT Explorations of Age Differences in Eye Movements during reading. *Neurosci. Biobehavioral Rev.* 30, 872–884. doi:10.1016/j.neubiorev.2006.06.013

Lev-Ari, S. (2019). People with Larger Social Networks Are Better at Predicting what Someone Will Say but Not How They Will Say it. *Lang. Cogn. Neurosci.* 34, 101–114. doi:10.1080/23273798.2018.1508733

Levy, R. (2008). Expectation-based Syntactic Comprehension. *Cognition* 106, 1126–1177. doi:10.1016/j.cognition.2007.05.006

Lewis, G., Solomyak, O., and Marantz, A. (2011). The Neural Basis of Obligatory Decomposition of Suffixed Words. *Brain Lang.* 118, 118–127. doi:10.1016/j.bandl.2011.04.004

Linzen, T., and Jaeger, T. F. (2016). Uncertainty and Expectation in Sentence Processing: Evidence from Subcategorization Distributions. *Cogn. Sci.* 40, 1382–1411. doi:10.1111/cogs.12274

Lopopolo, A., Frank, S. L., van den Bosch, A., and Willems, R. M. (2017). Using Stochastic Language Models (SLM) to Map Lexical, Syntactic, and Phonological Information Processing in the Brain. *PLoS ONE* 12, e0177794. doi:10.1371/journal.pone.0177794

Lorenz, D., and Tizón-Couto, D. (2019). Chunking or Predicting - Frequency Information and Reduction in the Perception of Multi-word Sequences. *Cogn. Linguistics* 30, 751–784. doi:10.1515/cog-2017-0138

Lowder, M. W., Choi, W., Ferreira, F., and Henderson, J. M. (2018). Lexical Predictability during Natural Reading: Effects of Surprisal and Entropy Reduction. *Cogn. Sci.* 42, 1166–1183. doi:10.1111/cogs.12597

Lupyan, G., and Clark, A. (2015). Words and the World. *Curr. Dir. Psychol. Sci.* 24, 279–284. doi:10.1177/0963721415570732

MacCallum, R. C., Zhang, S., Preacher, K. J., and Rucker, D. D. (2002). On the Practice of Dichotomization of Quantitative Variables. *Psychol. Methods* 7, 19–40. doi:10.1037/1082-989X.7.1.19

Mani, N., and Huettig, F. (2012). Prediction during Language Processing Is a Piece of Cake-But Only for Skilled Producers. *J. Exp. Psychol. Hum. Perception Perform.* 38, 843–847. doi:10.1037/a0029284

Mann, V. A., and Singson, M. (2003). "Reading Complex Words," in *Reading Complex Words Cross-Language Studies.* Editors E. M. H. Assink and D. Sandra (Amsterdam: Kluwer Academic). Available at: https://doi.org/10.1007/978-1-4757-3720-2 (Accessed October 1, 2021).

Mar, R. A., Oatley, K., Hirsh, J., dela Paz, J., and Peterson, J. B. (2006). Bookworms versus Nerds: Exposure to Fiction versus Non-fiction, Divergent Associations with Social Ability, and the Simulation of Fictional Social Worlds. *J. Res. Personal.* 40, 694–712. doi:10.1016/j.jrp.2005.08.002

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., and Bates, D. (2017). Balancing Type I Error and Power in Linear Mixed Models. *J. Mem. Lang.* 94, 305–315. doi:10.1016/j.jml.2017.01.001

McCauley, S. M., and Christiansen, M. H. (2019). Language Learning as Language Use: A Cross-Linguistic Model of Child Language Development. *Psychol. Rev.* 126, 1–51. doi:10.1037/rev0000126

McConnell, K., and Blumenthal-Dramé, A. (2019). Effects of Task and Corpus-Derived Association Scores on the Online Processing of Collocations. *Corpus Linguistics Linguistic Theor.* doi:10.1515/cllt-2018-0030

McDonald, S. A., and Shillcock, R. C. (2003a). Eye Movements Reveal the On-Line Computation of Lexical Probabilities during Reading. *Psychol. Sci.* 14, 648–652. doi:10.1046/j.0956-7976.2003.psci_1480.x

McDonald, S. A., and Shillcock, R. C. (2003b). Low-level Predictive Inference in reading: the Influence of Transitional Probabilities on Eye Movements. *Vis. Res.* 43, 1735–1751. doi:10.1016/S0042-6989(03)00237-2

Medimorec, S., Milin, P., and Divjak, D. (2019). Working Memory Affects Anticipatory Behavior during Implicit Pattern Learning. *Psychol. Res.* 85, 291–301. doi:10.1007/s00426-019-01251-w

Milin, P., Feldman, L. B., Ramscar, M., Hendrix, P., and Baayen, R. H. (2017). Discrimination in Lexical Decision. *PLOS ONE* 12, e0171935. doi:10.1371/journal.pone.0171935

Mishra, R. K., Singh, N., Pandey, A., and Huettig, F. (2012). Spoken Language-Mediated Anticipatory Eye-Movements Are Modulated by reading Ability - Evidence from Indian Low and High Literates. *Journal of Eye Movement Research* 5 (1), 1–10. doi:10.16910/jemr.5.1.3

Moers, C., Meyer, A., and Janse, E. (2017). Effects of Word Frequency and Transitional Probability on Word Reading Durations of Younger and Older Speakers. *Lang. Speech* 60, 289–317. doi:10.1177/0023830916649215

N. Chater and M. Oaksford (Editors) (2008). *The Probabilistic Mind: Prospects for Bayesian Cognitive Science* (Oxford University Press). doi:10.1093/acprof:oso/9780199216093.001.0001

Nelson, M. J., El Karoui, I., Giber, K., Yang, X., Cohen, L., Koopman, H., et al. (2017). Neurophysiological Dynamics of Phrase-Structure Building during Sentence Processing. *Proc. Natl. Acad. Sci. USA* 114, E3669–E3678. doi:10.1073/pnas.1701590114

Ng, S., Payne, B. R., Steen, A. A., Stine-Morrow, E. A. L., and Federmeier, K. D. (2017). Use of Contextual Information and Prediction by Struggling Adult Readers: Evidence from Reading Times and Event-Related Potentials. *Scientific Stud. Reading* 21, 359–375. doi:10.1080/10888438.2017.1310213

Nieuwland, M. S., Politzer-Ahles, S., Heyselaar, E., Segaert, K., Darley, E., Kazanina, N., et al. (2018). Large-scale Replication Study Reveals a Limit on Probabilistic Prediction in Language Comprehension. *eLife* 7. doi:10.7554/eLife.33468

Onnis, L., Lim, A., Cheung, S., and Huettig, F. Is the mind inherently forward looking? Experimental evidence for 'retrodiction' in language processing. 22.

Ordin, M., Polyanskaya, L., and Soto, D. (2020). Neural Bases of Learning and Recognition of Statistical Regularities. *Ann. N.Y. Acad. Sci.* 1467, 60–76. doi:10.1111/nyas.14299

Otten, M., and Van Berkum, J. J. A. (2008). Discourse-Based Word Anticipation during Language Processing: Prediction or Priming. *Discourse Process.* 45, 464–496. doi:10.1080/01638530802356463

Payne, B. R., and Federmeier, K. D. (2018). Contextual Constraints on Lexico-Semantic Processing in Aging: Evidence from Single-word Event-Related Brain Potentials. *Brain Res.* 1687, 117–128. doi:10.1016/j.brainres.2018.02.021

Payne, B. R., and Silcox, J. W. (2019). "Aging, Context Processing, and Comprehension," in *Psychology of Learning and Motivation* (Elsevier), 215–264. doi:10.1016/bs.plm.2019.07.001

Pelucchi, B., Hay, J. F., and Saffran, J. R. (2009). Learning in Reverse: Eight-Month-Old Infants Track Backward Transitional Probabilities. *Cognition* 113, 244–247. doi:10.1016/j.cognition.2009.07.011

Perfetti, C. A. (1992). "The Representation Problem in reading Acquisition," in *Reading Acquisition* (Hillsdale, NJ, US: Lawrence Erlbaum Associates), 145–174.

Perfors, A., Tenenbaum, J. B., Griffiths, T. L., and Xu, F. (2011). A Tutorial Introduction to Bayesian Models of Cognitive Development. *Cognition* 120, 302–321. doi:10.1016/j.cognition.2010.11.015

Perruchet, P., and Desaulty, S. (2008). A Role for Backward Transitional Probabilities in Word Segmentation. *Mem. Cogn.* 36, 1299–1305. doi:10.3758/MC.36.7.1299

Pickering, M. J., and Gambi, C. (2018). Predicting while Comprehending Language: A Theory and Review. *Psychol. Bull.* 144, 1002–1044. doi:10.1037/bul0000158

Ramscar, M., Hendrix, P., Love, B., and Baayen, R. H. (2014). Learning Is Not Decline. *Ml* 8, 450–481. doi:10.1075/ml.8.3.08ram

Rayner, K., Reichle, E. D., Stroud, M. J., Williams, C. C., and Pollatsek, A. (2006). The Effect of Word Frequency, Word Predictability, and Font Difficulty on the Eye Movements of Young and Older Readers. *Psychol. Aging* 21, 448–465. doi:10.1037/0882-7974.21.3.448

Rayner, K., Yang, J., Castelhano, M. S., and Liversedge, S. P. (2011). Eye Movements of Older and Younger Readers when reading Disappearing Text. *Psychol. Aging* 26, 214–223. doi:10.1037/a0021279

Rayner, K., Yang, J., Schuett, S., and Slattery, T. J. (2013). Eye Movements of Older and Younger Readers when Reading Unspaced Text. *Exp. Psychol.* 60, 354–361. doi:10.1027/1618-3169/a000207

Roete, I., Frank, S. L., Fikkert, P., and Casillas, M. (2020). Modeling the Influence of Language Input Statistics on Children's Speech Production. *Cogn. Sci.* 44. doi:10.1111/cogs.12924

Roland, D., Dick, F., and Elman, J. L. (2007). Frequency of Basic English Grammatical Structures: A Corpus Analysis. *J. Mem. Lang.* 57, 348–379. doi:10.1016/j.jml.2007.03.002

Ryskin, R., Levy, R. P., and Fedorenko, E. (2020). Do domain-general Executive Resources Play a Role in Linguistic Prediction? Re-evaluation of the Evidence and a Path Forward. *Neuropsychologia* 136, 107258. doi:10.1016/j.neuropsychologia.2019.107258

Ryskin, R., Ng, S., Mimnaugh, K., Brown-Schmidt, S., and Federmeier, K. D. (2019). Talker-specific Predictions during Language Processing. *Lang. Cogn. Neurosci.* 35, 797–812. doi:10.1080/23273798.2019.1630654

Sánchez-Izquierdo, M., and Fernández-Ballesteros, R. (2021). Cognition in Healthy Aging. *IJERPH* 18, 962. doi:10.3390/ijerph18030962

Shipley, W. C. (1940). A Self-Administering Scale for Measuring Intellectual Impairment and Deterioration. *J. Psychol.* 9, 371–377. doi:10.1080/00223980.1940.9917704

Slattery, T. J., and Yates, M. (2018). Word Skipping: Effects of Word Length, Predictability, Spelling and reading Skill. *Q. J. Exp. Psychol.* 71, 250–259. doi:10.1080/17470218.2017.1310264

Smith, N. J., and Levy, R. (2013). The Effect of Word Predictability on reading Time Is Logarithmic. *Cognition* 128, 302–319. doi:10.1016/j.cognition.2013.02.013

Solomyak, O., and Marantz, A. (2010). Evidence for Early Morphological Decomposition in Visual Word Recognition. *J. Cogn. Neurosci.* 22, 2042–2057. doi:10.1162/jocn.2009.21296

Stanovich, K. E., and West, R. F. (1989). Exposure to Print and Orthographic Processing. *Reading Res. Q.* 24, 402. doi:10.2307/747605

Steen-Baker, A. A., Ng, S., Payne, B. R., Anderson, C. J., Federmeier, K. D., and Stine-Morrow, E. A. L. (2017). The Effects of Context on Processing Words during Sentence reading Among Adults Varying in Age and Literacy Skill. *Psychol. Aging* 32, 460–472. doi:10.1037/pag0000184

Stine-Morrow, E. A. L., and Payne, B. R. (2016). Age Differences in Language Segmentation. *Exp. Aging Res.* 42, 83–96. doi:10.1080/0361073X.2016.1108751

Supasiraprapa, S. (2019). Frequency Effects on First and Second Language Compositional Phrase Comprehension and Production. *Appl. Psycholinguistics* 40, 987–1017. doi:10.1017/S0142716419000109

Tabullo, Á. J., Shalom, D., Sevilla, Y., Gattei, C. A., París, L., and Wainselboim, A. (2020). Reading Comprehension and Predictability Effects on Sentence Processing: An Event-Related Potential Study. *Mind, Brain Educ.* 14, 32–50. doi:10.1111/mbe.12205

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., and Goodman, N. D. (2011). How to Grow a Mind: Statistics, Structure, and Abstraction. *Science* 331, 1279–1285. doi:10.1126/science.1192788

Troyer, M., and Kutas, M. (2020). Harry Potter and the Chamber of what?: the Impact of what Individuals Know on Word Processing during reading. *Lang. Cogn. Neurosci.* 35, 641–657. doi:10.1080/23273798.2018.1503309

van Paridon, J., and Alday, P. M. (2020). A Note on Co-occurrence, Transitional Probability, and Causal Inference. *PsyArXiv*. doi:10.31234/osf.io/92zbx

Venhuizen, N. J., Crocker, M. W., and Brouwer, H. (2019). Expectation-based Comprehension: Modeling the Interaction of World Knowledge and Linguistic Experience. *Discourse Process.* 56, 229–255. doi:10.1080/0163853X.2018.1448677

Verhagen, V., Mos, M., Backus, A., and Schilperoord, J. (2018). Predictive Language Processing Revealing Usage-Based Variation. *Lang. Cogn.* 10, 329–373. doi:10.1017/langcog.2018.4

Whitford, V., and Titone, D. (2019). Lexical Entrenchment and Cross-Language Activation: Two Sides of the Same coin for Bilingual reading across the Adult Lifespan. *Bilingualism* 22, 58–77. doi:10.1017/S1366728917000554

Whitford, V., and Titone, D. (2014). The Effects of Reading Comprehension and Launch Site on Frequency-Predictability Interactions during Paragraph Reading. *Q. J. Exp. Psychol.* 67, 1151–1165. doi:10.1080/17470218.2013.848216

Whitford, V., and Titone, D. (2017). The Effects of Word Frequency and Word Predictability during First- and Second-Language Paragraph reading in Bilingual Older and Younger Adults. *Psychol. Aging* 32, 158–177. doi:10.1037/pag0000151

Wicha, N. Y. Y., Bates, E. A., Moreno, E. M., and Kutas, M. (2003). Potato Not Pope: Human Brain Potentials to Gender Expectation and Agreement in Spanish Spoken Sentences. *Neurosci. Lett.* 346, 165–168. doi:10.1016/S0304-3940(03)00599-8

Willems, R. M., Frank, S. L., Nijhof, A. D., Hagoort, P., and van den Bosch, A. (2016). Prediction during Natural Language Comprehension. *Cereb. Cortex* 26, 2506–2516. doi:10.1093/cercor/bhv075

Wlotko, E. W., and Federmeier, K. D. (2012). Age-related Changes in the Impact of Contextual Strength on Multiple Aspects of Sentence Comprehension. *Psychophysiol* 49, 770–785. doi:10.1111/j.1469-8986.2012.01366.x

Wlotko, E. W., Federmeier, K. D., and Kutas, M. (2012). To Predict or Not to Predict: Age-Related Differences in the Use of Sentential Context. *Psychol. Aging* 27, 975–988. doi:10.1037/a0029206

Wlotko, E. W., and Federmeier, K. D. (2015). Time for Prediction? the Effect of Presentation Rate on Predictive Sentence Comprehension during Word-by-word reading. *Cortex* 68, 20–32. doi:10.1016/j.cortex.2015.03.014

Yan, S., Kuperberg, G. R., and Jaeger, T. F. (2017). Prediction (or Not) during language processing. A commentary on Nieuwland et al. (2017) and DeLong et al. (2005). A Commentary On Nieuwland et al. (2017) And Delong et al. (2005). bioRxiv. doi:10.1101/143750

Yap, M. J., Balota, D. A., Sibley, D. E., and Ratcliff, R. (2012). Individual Differences in Visual Word Recognition: Insights from the English Lexicon Project. *J. Exp. Psychol. Hum. Perception Perform.* 38, 53–79. doi:10.1037/a0024177

McConnell, K., and Blumenthal-Dramé, A. (2018). *Effects of task and corpus-derived association scores on the online processing of collocations.* submitted.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article or claim that may be made by its manufacturer is not guaranteed or endorsed by the publisher.