The many timescales of context in language processing

Rachel Ryskin* **and Xinzhu Fang***
University of California, Merced, CA, United States
*Corresponding authors e-mail address: rryskin@ucmerced.edu; xfang4@ucmerced.edu

Contents

Abstract

Language input is often noisy and ambiguous. Yet, humans are able to successfully communicate their thoughts to one another by leveraging contextual information to make inferences about the intended meaning of the input. Though there is broad agreement on the role and importance of context, the term is used so flexibly that it can be difficult to know what it refers to and how to make progress on studying it. Here, we propose that a key dimension which spans domains of context is the timescale over which the contextual information is sampled for the purpose of inferring meaning. We review the literature on context effects in language comprehension, organizing our discussion around a few relevant temporal windows, milliseconds, seconds, and minutes, and then propose that the individual's lifetime of language experience and historical time both constitute relevant contexts for language processing as well. We then discuss how construing of context in this way reveals gaps in the existing research landscape and propose new avenues of inquiry to address them.

Psychology of Learning and Motivation
ISSN 0079-7421
https://doi.org/10.1016/bs.plm.2021.08.001

1. Introduction

Language is the best tool that humans have for sharing thoughts between minds. It enables us to order a cup of coffee, learn how our loved one's day was, and transmit scientific knowledge across generations. Yet, at first glance, it seems like an imperfect system. Opportunities for miscommunication arise at every level of language processing. The same acoustic features can map onto different phonemes depending on who is saying them, words and sentences can have multiple—related or unrelated—meanings, speakers may leave the exact referent of their utterance ambiguous or make errors, and listeners can incorrectly perceive the input.

How humans overcome these "imperfections" to successfully exchange thoughts remains a fundamental open question in cognitive science, yet one piece of the answer has emerged as uncontroversial in the literature: *language comprehension is context-dependent*. As language input is processed, a cascade of (probabilistic) inferences unfolds allowing the language comprehender to resolve ambiguities and compensate for noise at all levels of linguistic representation. These inferences leverage contextual information across many domains. For instance, during speech perception, listeners use the earliest phonemic cues in a word to anticipate the identity of the word before they hear the end. Conversation partners can rely on their memory of the dialog to know what the speaker meant when they said "Now hand me the striped one." And individuals familiar with multiple dialects will interpret the same word differently depending on the speaker's accent (e.g., "a flat" likely refers to a flat car tire in American English but an apartment in British English). But this breadth of evidence for the role of context has yet to yield its full potential in terms of advancing our understanding of language processing in the human mind.

One barrier to developing a unified account of human language processing may be the fact that it is not always clear what is meant by "context." Indeed, the term "context" is used flexibly to refer to a seemingly infinite number of information sources beyond the immediately processed stimulus. In fact, what is stimulus and what is context is often a matter of perspective or granularity of measurement. A word-initial phoneme serves as context to identify subsequent phonemes but that same word-initial phoneme may be more or less predictable given the preceding words in the sentence and who is producing the sentence. These contexts appear to rely on different knowledge representations (e.g., phonology vs word frequency vs world knowledge), likely supported by distinct neural machinery (Braga, DiNicola, Becker, & Buckner, 2020; Fedorenko, Behr, & Kanwisher, 2011; Norman-Haignere, Kanwisher, & McDermott, 2015). What unites the phoneme-context, the sentence-context, the speaker-context, and many other contexts is that they are sources of information which the comprehender uses to derive a probabilistic inference over meanings intended by the producer. As a first step toward articulating a unifying framework, we here propose that a key organizing dimension which spans domains of context is the timescale over which the contextual information is sampled for the purpose of inferring meaning (schematized in Fig. 1).

This is not an entirely new idea. Previous proposals have appealed to lossy compression at multiple timescales to provide a unified explanation for language acquisition, processing, and evolution (Christiansen & Chater, 2016). Similarly, one proposal for the neural architecture of language focuses on the temporal windows over which linguistic information is integrated (Lerner, Honey, Silbert, & Hasson, 2011, c.f. Blank & Fedorenko, 2020).

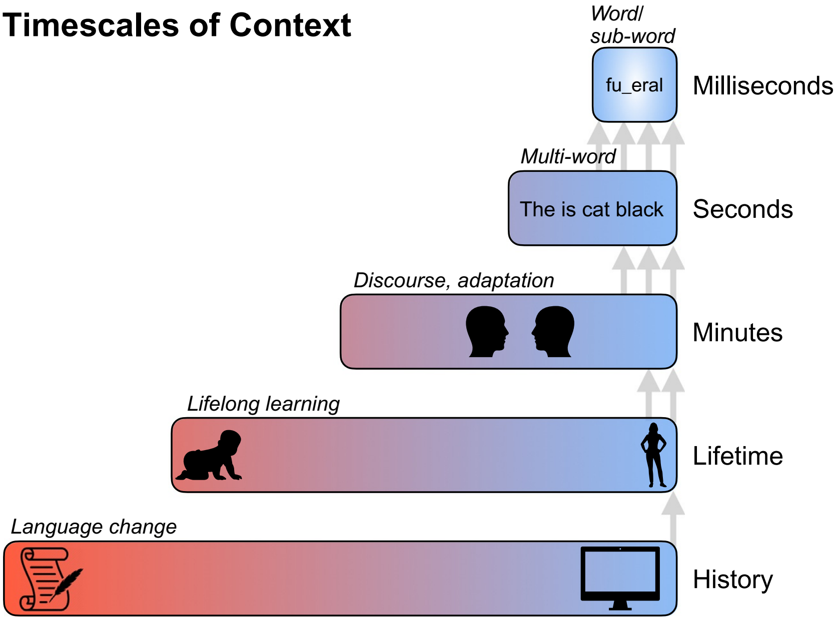

Fig. 1 Schematic illustration of timescales over which contextual information can be sampled to support language processing. As language unfolds, comprehenders use context on multiple timescales, both prospectively and retrospectively, to make inferences about the intended meaning of the input.

Unlike those accounts, the current perspective does not aim to make any mechanistic claims regarding *how* context is processed, but rather to recast disparate findings within an existing computational-level framework—one of rational inference under uncertainty (Gibson, Bergen, & Piantadosi, 2013; Hawkins, Franke, et al., 2021; Kleinschmidt & Jaeger, 2015; Levy, 2008b; Levy, Bicknell, Slattery, & Rayner, 2009, inter alia)—in an attempt to clarify *what* constitutes a context.

In the present piece, we briefly review the literature on context effects in language comprehension, setting aside issues of whether effects of context reflect prediction and/or integration.[a] We structure the discussion around a few relevant temporal windows[b]: milliseconds, seconds, minutes, the lifespan, and historical time. In doing so, we cut across traditional lines of inquiry (e.g., syntax vs semantics vs pragmatics) and define context more broadly to include all information that unfolds on a certain timescale, both that which precedes the stimulus and that which follows it (when applicable). We then discuss how construing of context in this way reveals gaps in the existing research landscape and propose new avenues of inquiry to address them. In particular, we argue that bridging across timescales and delineating the constraints on context-dependence will be crucial for illuminating how humans are able to decode each other's thoughts from language.

2. Context as information guiding inferences on multiple timescales

2.1 Milliseconds of context

The identity of a word or word segment is inferred from a combination of bottom-up sensory input and the surrounding acoustic, phonemic, and visual cues (Brown, Tanenhaus, & Dilley, 2021). When hearing a sound that is ambiguous between /d/ and /t/ followed by "… ask," resulting either in a nonword ("dask") or a common word ("task"), English speakers tend to infer that the sound was /t/. If the word continuation is "… esk" they are biased to infer that the sound was /d/ (as "tesk" is not a word and "desk" is; Ganong, 1980). If visual information about the speaker's production is

---

[a] For reviews see e.g., Federmeier, 2007; Ferreira & Chantavarin, 2018; Huettig & Mani, 2016; Kuperberg & Jaeger, 2016; Pickering & Gambi, 2018.

[b] Breaking the timescales up into these coarse-grained windows is done simply for expository purposes. There is likely no categorical distinction between the kind of information that can be sampled within a time window of 999 ms and one of 1100 ms.

available (i.e., their lip movements), visual cues and bottom-up auditory cues are combined to infer the nature of the syllable (Mcgurk & Macdonald, 1976). When a letter within a word is masked by acoustic noise (e.g., the "s" in "legislatures" is covered by a cough) listeners often fail to notice this absence and perceive the word as being error-free (Cole, 1973; Samuel, 1981; Warren, 1970). Critically, this "phoneme restoration effect" is reduced when the masked phoneme is inserted in a pseudo-word.

Similarly, readers are more likely to correctly recall which letter appeared in a briefly presented (and subsequently masked) string if the string was a word (e.g., "work") rather than a nonword (e.g., "owrk") (word superiority effect; Reicher, 1969; Wheeler, 1970). Readers also often fail to notice transposed letters in the middle of a word (e.g. "avitaion" instead of "aviation") because the surrounding letters provide sufficient context to identify the word (Bruner & O'Dowd, 1958; Chambers, 1979; Fischer-Baum, Charny, & McCloskey, 2011).

2.1.1 Prospective context

Preceding acoustic cues allow the listener to constrain the set of potential words/phonemes under consideration. As the speech signal unfolds, the listener can typically identify that a speech segment corresponds to a specific word within a few hundred milliseconds, well before the word's offset (Marslen-Wilson, 1975). Visual world paradigm eye-tracking studies (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995)—where listeners hear an auditory stimulus (e.g., "Pick up the beaker") while viewing a visual display which includes multiple candidate referents (e.g., a beaker, a beetle, a speaker, and a carriage)—reveal that listeners use acoustic and phonemic cues immediately and incrementally to circumscribe the set of potential referents (Allopenna, Magnuson, & Tanenhaus, 1998; McMurray, Clayards, Tanenhaus, & Aslin, 2008; Toscano & McMurray, 2012, 2015). For example, after hearing only the onset of the word "beaker," listeners look more to referents that are compatible with this onset phoneme, the beaker and the beetle, and less to referents which have different onset phonemes, the speaker and carriage. Further, because speech production planning is affected by the immediately preceding and following sounds (i.e., the vowel in "the" will sound somewhat different if the subsequent word will be "fish" vs "ladder"), listeners can use these coarticulatory cues to predict what word onset they are about to hear (Dahan, Magnuson, Tanenhaus, & Hogan, 2001; Salverda, Kleinschmidt, & Tanenhaus, 2014).

2.1.2 Retrospective context

The effects of preceding context are not deterministic, however (Apfelbaum, Bullock-Rest, Rhone, Jongman, & McMurray, 2014). Listeners maintain uncertainty about what they have heard and draw inferences based on the subsequent material. For example, in Allopenna et al. (1998), the image of the speaker (or the physical speaker object in some versions of the study) receives some renewed consideration after the listener hears the later portion of the word "beaker," because the later segments of the words "beaker" and "speaker" overlap, even though the onset of "speaker" is inconsistent with the auditory input. (This is known as the "rhyme" effect.) Similarly, uncertainty about the identity of a word onset segment (or gradient acoustic information) can be maintained until later syllables in the word disambiguate it. For example, when voice onset time is near the /b/ vs /p/ category boundary, the words *barri<u>cade</u>* and *para<u>keet</u>* are onset competitors and the identity of the word is ambiguous to listeners until they hear the last (underlined) portion of the word (Gwilliams, Linzen, Poeppel, & Marantz, 2018; McMurray, Tanenhaus, & Aslin, 2009).

2.2 Seconds of context

Readers and listeners understand a phrase or sentence in part through the compositional meaning of the words that constitute it. But when the literal meaning has low probability (e.g., "The mother gave the candle the daughter"), comprehenders often infer that a more plausible meaning (e.g., that the daughter received the candle) was intended (Christianson, Hollingworth, Halliwell, & Ferreira, 2001; Ferreira, 2003; Ferreira & Patson, 2007; Gibson, Bergen, & Piantadosi, 2013; Levy, 2008b; Poppels & Levy, 2016). In particular, noisy-channel accounts of language processing (Gibson, Piantadosi, et al., 2013; Levy, 2008b) propose that $P(s_i|s_p)$, the probability of inferring that a particular sentence/meaning, s_i, was intended given what was perceived, s_p, can be computed, following Bayes' rule (Eq. 1), based on the prior probability that the sentence was intended in the first place, $P(s_i)$, and the probability of the potential noise corruption that might have generated the perceived string from the intended string, $P(s_p|s_i)$.

$$P(s_i|s_p) \propto P(s_p|s_i) \cdot P(s_i) \tag{1}$$

For example, $P(s_i =$ "The mother gave the candle the daughter") is low. The probability of a close alternative sentence, $P(s_i =$ "The mother gave the candle to the daughter"), is higher, and the probability that the more plausible sentence was intended by the producer but corrupted (e.g., by the deletion of "to") into the implausible version that was perceived, $P(s_p =$ "The mother gave the candle the daughter" $|s_i =$ "The mother gave the candle to the daughter"), is relatively high. As a result, $P(s_i =$ "The mother gave the candle to the daughter" $|s_p =$ "The mother gave the candle the daughter") is higher than $P(s_i =$ "The mother gave the candle the daughter" $|s_p =$ "The mother gave the candle the daughter"). Offline comprehension questions (e.g., Did the daughter receive something?) reveal that readers are more likely to respond according to the more plausible meaning, which differs from the literal string in front of them, when the noise corruption that has to be posited is minor (e.g., deletion of "to") than when it is less probable (e.g., two deletions and two insertions) (Gibson, Bergen, & Piantadosi, 2013). In addition, and directly analogously to the transposed letter effect in words (Bruner & O'Dowd, 1958), a sentence with transposed words is judged as ungrammatical more slowly than a string of words of the same length which cannot be transformed into a syntactically correct sentence by transposing a pair of words (Mirault, Snell, & Grainger, 2018). In other words, the meaning of a sentence/phrase does not simply reflect the literal composition of the words that constitute it but rather is the result of an inference process based on the surrounding context consisting of information available within a few seconds.

2.2.1 Prospective context

A typical sentence unfolds over several seconds and the processing of each incoming word is affected by the words that preceded it. Current computational accounts of language comprehension have formalized context effects on comprehension in terms of the probability of a word given the preceding sequence of words (Jurafsky, 1996), in particular, surprisal (Eq. 2; Hale, 2001; Levy, 2008a), the negative log probability of the word, w_i, given the preceding context, where the context encompasses the previous words $w_1, \ldots, w_{i-1}$, and any context outside of the sentence, C.

$$surprisal(w_i) = -logP(w_i|w_1, \ldots, w_{i-1}, C) \tag{2}$$

When a word is highly probable given its context, surprisal is low. When a word is not likely given the context, surprisal is high. Surprisal is typically estimated based on the conditional probabilities computed from large corpora (with n-gram models or neural networks) or cloze task responses (sentence completions collected from large samples of people). It is noteworthy that the "true" surprisal of a word would include all manner of information about the world and the setting of the linguistic exchange, contained in *C*, but this is intractable to estimate, so the necessary simplifying assumption is that *C* is constant and therefore surprisal estimates reflect primarily the statistics of the language.[c] Nonetheless, word-by-word surprisal provides a good approximation to behavioral and neural measures obtained during natural reading (Frank, Otten, Galli, & Vigliocco, 2015; Smith & Levy, 2013).

Controlled experiments also clearly demonstrate that when a word is predictable based on the preceding words (e.g., the word "shark" in "The coast guard warned that someone had seen a <u>shark</u> off the north shore of the island."), it is read faster—and even sometimes skipped entirely—compared to when it is not predictable (e.g., "The zookeeper explained that the life span of a <u>shark</u> is much longer than those of other animals."; Ehrlich & Rayner, 1981). Similarly, visual world paradigm studies reveal that listeners incrementally look to objects in the scene that are likely to be referred to next based on the semantic and syntactic constraints of the preceding words. For instance, participants make more fixations to a cake after hearing "The boy will eat the…" compared to "The boy will move the…" (Altmann & Kamide, 1999; Snedeker & Trueswell, 2004; Trueswell, Tanenhaus, & Kello, 1993). Pragmatically motivated predictions operate on a similar timescale: hearing a scalar adjective (e.g., "Show me the big…") leads listeners to anticipate that the referent that will follow should be one that is a member of a contrast set that differs along the relevant dimension (e.g., a big cup and a small cup; Sedivy, Tanenhaus, Chambers, & Carlson, 1999), the logic being that the speaker would not have included the modifier "big" if the referent was a singleton (Grice, 1975).

Some of the clearest demonstrations of preceding words serving as a context for those that follow come from studies using recordings of electrical

---

[c] How much world knowledge/semantic information can be recovered from models trained on linguistic datasets alone is a matter of active ongoing investigation (e.g., Grand, Blank, Pereira, & Fedorenko, 2018; Huebner & Willits, 2018; Kim, Elli, & Bedny, 2019; Lewis, Zettersten, & Lupyan, 2019; Mikolov, Sutskever, Chen, Corrado, & Dean, 2013; Pennington, Socher, & Manning, 2014). On the one hand, many sentences of a language describe possible events, e.g., "Flowers bloom in spring," particularly within didactic genres. On the other hand, adults are unlikely to talk about things that are self-evident, e.g., "That cat has fur," and are instead more likely to mention things when they deviate from the norm, e.g., "That cat doesn't have any fur!," suggesting that representations learned from linguistic data alone may have some biases: despite providing a reasonable match to human judgments of similarity (Hill, Reichart, & Korhonen, 2016).

activity over the scalp (the electroencephalogram or EEG). In particular, the N400 event-related potential (ERP) component is a negative-going deflection of the EEG occurring approximately 400 ms after a stimulus and thought to index the process of semantic access or updating the activation state of the comprehender's semantic network in response to the stimulus (Federmeier & Laszlo, 2009; Fitz & Chang, 2019; Kutas & Federmeier, 2011; Rabovsky, Hansen, & McClelland, 2018). Its (negative) amplitude is reduced to words that are predictable in context (e.g., "He planted string beans in his garden.") relative to those which are not (e.g., "He planted string beans in his car."; Kutas & Hillyard, 1984; inter alia Dambacher, Kliegl, Hofmann, & Jacobs, 2006; Dimigen et al., 2011). Moreover, the N400 is more negative for words that are earlier in the sentence than those that are later (Payne, Lee, & Federmeier, 2015; Van Petten & Kutas, 1991), suggesting that, as contextual information accumulates, in a typical sentence, the state of the semantic network changes less with each subsequent word.

Other downstream ERP components are similarly sensitive to context. A later anterior positive deflection appears to index situations where a strong prediction was generated and then disconfirmed by an alternative that is not incongruous (e.g., "The groom took the bride's hand and placed the ring on her dresser." ["finger" is predicted]; Federmeier, Kutas, & Schul, 2010; Federmeier, Wlotko, De Ochoa-Dewald, & Kutas, 2007, see also Kuperberg, Brothers, & Wlotko, 2020). The P600—a positivity most pronounced 600–900 ms after word onset—was originally thought to reflect syntactic integration difficulty but has since come to be seen as indexing a more general process related to error-monitoring or correction (Leckey & Federmeier, 2020; Ryskin et al., 2021; van de Meerendonk, Indefrey, Chwilla, & Kolk, 2011). In particular, words that have low probability given the preceding context (e.g., "Before lunch he has to deposit his paycheck at the bark) but are orthographically close to a plausible continuation (e.g., bank)—such that the incongruity of the word could be attributed to noise corruption (e.g., a typographical error)—elicit smaller N400 effects and larger P600 effects than those which are incongruous but have no such orthographic neighbors (Ito, Corley, Pickering, Martin, & Nieuwland, 2016; Ito, Martin, & Nieuwland, 2017; Laszlo & Federmeier, 2009). Furthermore, the N400 effect is additionally reduced and the P600 is more positive whenever the probability of the plausible alternative word (e.g., bank) is higher given the context (Ito et al., 2017) and the probability of a reader recovering the intended, plausible continuation from the "noisy" one is higher (Ryskin et al., 2021). Analogously to the millisecond level, preceding context on the order of seconds is used to make inferences about the intended meaning from noisy or ambiguous bottom-up input.

2.2.2 Retrospective context

A recent extension of the surprisal-based account introduces a lossy representation of the context (Futrell, Gibson, & Levy, 2020), consistent with the observation that memory for preceding sentence material is imperfect (Lewis & Vasishth, 2005; Lewis, Vasishth, & Van Dyke, 2006; McElree, Foraker, & Dyer, 2003). Memory representations of recently heard or read material can be corrupted by noise and therefore revised in light of subsequent input. For example, listeners maintain perceptual uncertainty about the identity of an ambiguous target word (e.g., with a word-initial voice onset time that makes it ambiguous between "tent" and "dent") until they reach a word later in the sentence which provides a semantic cue regarding the word's identity (e.g., in "when the [tent/dent] in the forest was …" forest is tent-biasing while in "when the [tent/dent] in the fender was …" fender is dent-biasing; Bicknell, Jaeger, & Tanenhaus, 2016; Brown-Schmidt & Toscano, 2017; Bushong & Jaeger, 2019; Connine, Blasko, & Hall, 1991; Falandays, Brown-Schmidt, & Toscano, 2020; Szostak & Pitt, 2013; Zellou & Dahan, 2019).

Similarly, in a visual world paradigm eye-tracking study, when listeners hear a sentence with a syntactic agreement error (e.g., "The key to the cabinet *were on the table"), soon after the error, they look to items in the display which would have been more appropriate subjects (e.g., keys) but were inconsistent with the preceding context (Brehm, Jackson, & Miller, 2021). Plausibly, listeners are uncertain about whether they heard the previous words correctly—a deletion of '-s' is a particularly plausible noise corruption. Furthermore, Levy et al. (2009) showed that, when a later portion of a sentence renders the most likely early parse of a sentence improbable (e.g., "The coach smiled at the player tossed a frisbee."), readers are more likely to look back to previous locations in the sentence which are probable loci of noise corruptions. For instance, they make increased fixations to "at" because $P(\text{“}at\text{”} \mid \text{“}as\text{”})$ is high and "The coach smiled as the player tossed a frisbee." is a more probable sentence, relative to when they read a sentence where the initial parse remained plausible (e.g., "The coach smiled at the player thrown a frisbee."). In other words, readers maintain uncertainty about the preceding input as they process a sentence and infer intended sentence meaning and structure in light of both preceding and subsequent context (see also Keshev & Meltzer-Asscher, 2021).

2.3 Minutes/hours of context

How listeners and readers interpret a word or sentence is also affected by the larger environment that they experience over the course of minutes or even hours, whether they are listening to a story, engaging in a conversation, or reading a text. Moreover, in most psycholinguistic experiments, participants' behaviors are measured repeatedly (c.f. von der Malsburg, Poppels, & Levy, 2020) such that the surrounding laboratory environment and experimental stimuli can shape the language distributions from which probabilistic inferences are drawn.

2.3.1 Prospective context

Listeners track the phonetic cues in an environment and adapt their mappings from sounds to phonemes/words to the context (e.g., Bradlow & Bent, 2008; Dahan, Drucker, & Scarborough, 2008; Kleinschmidt & Jaeger, 2015). For example, after listening to a story in which the talker had a novel accent—vowel lowering relative to a standard American English accent such that the pronunciation of "witch" would be *wetch* in this accent—participants judged the novel phonetic forms to be words in a lexical decision task, but did not do so when they heard the same story presented in the standard accent (Maye, Aslin, & Tanenhaus, 2008).

Similarly, listeners can adapt their interpretation of a contrastive adjective (e.g., "big") based on how it is used in the experimental context (Grodner & Sedivy, 2011). In Ryskin, Kurumada, and Brown-Schmidt (2019), listeners were exposed either to a speaker who used size adjectives felicitously, to disambiguate a referent from its pair that differed in size (e.g., when there is a big cup and a small cup in the display), or to an infelicitous speaker who consistently failed to use a size adjective to disambiguate the referent or included a superfluous adjective when there was no ambiguity (i.e., there is only one cup). Eye-tracking indicated that listeners were less likely to predict that the referent would be a member of a pair after exposure to the infelicitous speaker than the felicitous speaker (see also Gardner et al., 2021; Kurumada, Brown, & Tanenhaus, 2018; Schuster & Degen, 2020 for similar findings in the domains of contrastive prosody and uncertainty expressions).

In a dialog, listeners learn about their interlocutor's lexical choice patterns (e.g., what label a speaker is most likely to use to refer to a particular object; Brennan & Clark, 1996; Clark & Wilkes-Gibbs, 1986; Wilkes-Gibbs & Clark, 1992). For example, in a referential communication task where a Director and Matcher discuss multiple abstract "tangram" shapes in order to coordinate on the same arrangement, the ad hoc labels that speakers devise start out fairly verbose (e.g.,“ Um, it looks like a person. They have their arms out to the left and the head is kind of back like they're dancing”). Over the course of multiple conversational turns, interlocutors entrain on jointly developed succinct labels (e.g., “the dancer”). Listeners also adapt their communicative expectations to the informativity of their partner (Hawkins, Gweon, & Goodman, 2021; Ryskin, Stevenson, & Heller, 2020). And they leverage the past history and goals of the conversation to resolve ambiguity (Brown-Schmidt & Konopka, 2011; Brown-Schmidt & Tanenhaus, 2008; Yoon, Koh, & Brown-Schmidt, 2012). The importance of this contextual information can be most clearly observed in ecological settings. In contrast to tightly controlled lab paradigms, in naturalistic conversation, speakers are frequently under-informative (e.g., “Hand me the lego” when there are multiple different-colored legos), yet the interaction continues unimpeded. Listeners are able to resolve this ambiguity and avoid competition from other candidate referents (e.g., the other colored lego), by considering elements of the recent context (on the order of minutes), including which objects have been recently mentioned and which are most relevant to the task.

In reading a text, comprehenders build a mental model of the narrative (Bower & Morrow, 1990). This continuously updated model reflects how probable different events might be in the context of the narrative. For example, after reading a story with anthropomorphic peanuts, comprehenders do not find the sentence “The peanut was in love” unexpected (as indexed by an absence of N400 effects), relative to when such a sentence is not preceded by a supportive discourse context (Nieuwland & Van Berkum, 2006). Moreover, comprehenders appear to rationally adapt their semantic predictions (as indexed by the amplitude of the N400) to the rates of semantic associates in the surrounding stimuli (Brothers, Hoversten, Dave, Traxler, & Swaab, 2019; Delaney-Busch, Morgan, Lau, & Kuperberg, 2019, but see Nieuwland, 2020).

Similarly, readers can learn new dialectal structures (e.g., *needs* + past participle, as in “The car needs washed,” which is idiosyncratic to western Pennsylvania) from exposure to the structure (Fraundorf & Jaeger, 2016; Kaschak & Glenberg, 2004). Comprehenders also learn the subtle distributional statistics of already-known syntactic structures in a given environment. For example, in a self-paced reading experiment in which infrequent syntactic structures are overrepresented among the stimuli, readers experience less processing difficulty for those infrequent structures than they do when the surrounding stimuli do not contain as many of those infrequent structures (Fine, Jaeger, Farmer, & Qian, 2013; Ryskin, Qi, Duff, & Brown-Schmidt, 2017; Tooley, Swaab, Boudewyn, Zirnstein, & Traxler, 2014, c.f. Harrington Stack, James, & Watson, 2018).

Readers also track the distribution of noise in their input. P600 effects are sensitive to the rate of errors in the environment (Coulson, King, & Kutas, 1998). Readers make more noisy-channel inferences—interpreting semantically implausible sentences nonliterally—when the rate of word deletion or insertion errors is higher in the surrounding stimuli, indicating that they adapt their representations of the probabilities of sentences as well as their noise likelihoods (Gibson, Bergen, & Piantadosi, 2013). Moreover, readers can learn the regularities in the *kinds* of errors that the writer tends to introduce (e.g., many of the sentences contain deletion errors vs many of the sentences contain insertion errors) and infer different intended sentences depending on which noise corruption is most probable in the given environment (Ryskin, Futrell, Kiran, & Gibson, 2018). Listeners exposed to a speaker with a foreign accent make more noisy-channel inferences relative to those listening to a speaker with an American accent, even when the content of speech is identical across speakers and error-free (Gibson et al., 2017). In other words, listeners increase their expectation of noise based on the assumption that a speaker with a foreign accent *may* produce more errors. Similarly, in the presence of sporadic acoustic noise (phonemes of nontarget words replaced by radio/white noise of equal duration), listeners maintain more uncertainty about the nature of earlier segments of a word: while listening to a target word (e.g., "beaker") rhyme competitors (e.g., "speaker"), which do not share an onset but overlap in terms of the later phonemes, are considered more strongly (as evidenced by gaze fixations in the visual world paradigm) relative to when no noise is introduced (McQueen & Huettig, 2012).

2.3.2 Retrospective contexts

To our knowledge, very few attempts have been made to investigate retrospective context effects on the order of minutes. Following up on work by Kraljic, Samuel, and Brennan (2008), Kraljic and Samuel (2011), and Liu and Jaeger (2018) showed that listeners exposed to atypical pronunciations from a novel talker (e.g., "dinosaur" pronounced as "dinoshaur") adapt their categorization of subsequent input from that talker but, crucially, do not do so when the speaker is shown to have a pen in their mouth. The pen provides a plausible alternative cause for the atypical pronunciation such that listeners do not expect the atypical pronunciation to persist once this causal element is removed. Further, they find that, once the alternative cause is removed and the atypical pronunciation is revealed to be a feature of the talker's speech, adaptation to the atypical pronunciation is faster than would be expected if listeners had simply ignored the prior evidence (while the pen was in the mouth). These results suggest that, as with the millisecond and second-level contexts, over the course of adaptation, listeners maintain uncertainty about the input and can update their inference about sound/-word identity when later information renders the original interpretation implausible.

2.3.3 Hierarchical contexts

In contrast to shorter timescales, context on the order of minutes can include additional structure. For example, a student may alternate reading paragraphs from a few different textbooks as they skim for relevant information or a listener may be involved in a conversation with several other people who are taking turns speaking. The same kinds of contextual cues as discussed to this point (e.g., preceding and/or subsequent sounds or words) could in principle have different distributions and lead to different inferences depending on the text or speaker. Indeed, this is the premise behind author identification in natural language processing—authors leave behind idiosyncratic traces in their output which can be detected using statistical or neural net models and used to classify documents (Ding, Fung, Iqbal, & Cheung, 2019; Mosteller & Wallace, 1963; Shrestha et al., 2017).

There is also growing experimental evidence that, over the course of a dialog, conversation partners learn about each other's idiosyncrasies and this information from the previous minutes of the linguistic exchange guides interpretation of future utterances by the same speaker (Brown-Schmidt, Yoon, & Ryskin, 2015). Listeners track the fine-grained acoustics of their conversation partner's speech (Creel & Tumlin, 2011; Kleinschmidt & Jaeger, 2015) allowing them to more readily anticipate an upcoming referent. For example, when exposed to a speaker with a familiar local accent (Midwestern American English) and a speaker with an unfamiliar regional accent (characterized by the pronunciation of the "a" in "tag," but not the "a" in "tack," similarly to the "a" in "take"), listeners learned this new pronunciation in a talker-specific way. They were quicker to identify the referent (e.g., "tack") when the auditory stimulus was produced by the "accented" talker, for whom "tag" and "tack" have different onset sounds, than the familiar accent talker, for whom "tag" and "tack" are onset competitors (Trude & Brown-Schmidt, 2012). Conversation partners also develop shared conceptualizations and terms (Brennan & Clark, 1996; Clark & Wilkes-Gibbs, 1986; Yoon & Brown-Schmidt, 2014). Listeners predict upcoming referents in a partner-specific way based on the speaker's past references and knowledge state (Brown-Schmidt, 2009; Metzing & Brennan, 2003), idiosyncratic preferences (Ryskin, Ng, Mimnaugh, Brown-Schmidt, & Federmeier, 2020), spatial viewpoint (Ryskin, Wang, & Brown-Schmidt, 2016), syntactic biases (Kamide, 2012; but see Liu, Burchill, Tanenhaus, & Jaeger, 2017), and even their previous tendency to say unexpected things (Brothers et al., 2019).

Despite these existence of speaker-specific idiosyncracies, the language found in any text or produced by any speaker presumably inherits most properties from the larger environment in which it is situated (e.g., the topic of the student's class, the type of conversation—friendly chat vs business negotiation) and the language as a whole, creating a hierarchically structured context (Hawkins, Franke, et al., 2021; Kleinschmidt & Jaeger, 2015), an idea we will return to in Section 4.2.2.

2.4 Interim summary

In the preceding review of the literature on context effects in language processing, we hoped to highlight the vast variety of phenomena that have been studied under this umbrella. We also proposed that a fruitful lens through which these phenomena can be viewed is that of the temporal window within which information is sampled in the service of inferring the producer's intended meaning. Recasting several results in this light reveals many commonalities across timescales. In particular, comprehenders use prior audio and visual context to constrain the set of possible inferences about the meaning of the linguistic input. They maintain uncertainty about the preceding input such that the representation of what came before can be updated to accommodate downstream context. However, comparatively little is known about downstream/retrospective context effects, particularly on the order of minutes or hours. We return to this in Section 4. In the next section, we extend context effects to encompass information on longer timescales.

3. Broadening context to longer timescales

3.1 A lifetime of context

A lifetime of conversations, reading books, exchanging messages, and generally accumulating linguistic and world experiences provides a unique backdrop to how an individual interprets any given linguistic input. Though the sum of a person's linguistic and nonlinguistic experience prior to entering the lab is rarely referred to as a type of context, we would argue that there is no discontinuity between the context effects operating over various timescales described above and the longer timescales we describe next. Arguably, some of the experimental studies reviewed here have sacrificed ecological validity for the sake of experimental control and it is possible that comprehenders make a categorical distinction between language they experience "in the wild" and language they experience in the laboratory (see Clark, 2021). Nonetheless, even in tightly controlled settings, it is undeniable that the language input that an individual experienced prior to entering the lab plays a critical role in determining how the more temporally local contexts (e.g., the surrounding words in a sentence, the distributional statistics of the experimental stimuli) affect the reader's/ listener's inferences.

At the timescale of several milliseconds, one of the most basic findings in psycholinguistics, that words that appear more frequently in the language are recognized and accessed more quickly (Balota & Chumbley, 1984; Dahan, Magnuson, & Tanenhaus, 2001; Howes & Solomon, 1951; Luce & Pisoni, 1998; Marslen-Wilson & Welsh, 1978), rests on the assumption that some aspect of word representation depends on the frequency with which the individual has experienced it. Spelling errors/letter transpositions are more likely to go unnoticed in higher frequency words than low frequency words (O'Connor & Forster, 1981), ambiguous acoustics are biased to be interpreted as more frequent words than nonwords (Connine, Titone, & Wang, 1993, but see Politzer-Ahles, Lee, & Shen, 2020), and the phoneme restoration effect is stronger for more frequent words (Samuel, 1981). Most current theoretical perspectives on language processing agree that the probabilities of phonemes, words, or constructions conditional on the context are the product of lifelong learning and accumulation of language statistics (Beckner et al., 2009; Chang, Janciauskas, & Fitz, 2012; Dell & Chang, 2013; Goldberg, 2019; Levy, 2008a; MacDonald, 2013; Wells, Christiansen, Race, Acheson, & Macdonald, 2009).

Perhaps even more plainly, a person's lifetime of experience is what determines what events they find plausible or implausible. World knowledge and semantic memory are known to guide language interpretation from the earliest moments of processing (Federmeier & Kutas, 1999; Hagoort, Hald, Bastiaansen, & Petersson, 2004; Kamide, Altmann, & Haywood, 2003; Kutas & Hillyard, 1984, inter alia). For example, the probability of a word in context (at the level of seconds) can be low if that word evokes an event that violates animacy constraints (Kuperberg, Kreher, Sitnikova, Caplan, & Holcomb, 2007) or is inconsistent with cultural norms (Van Berkum, van den Brink, Tesink, Kos, & Hagoort, 2008). On the other hand, a priori knowledge of the fact that the speaker is using their second language results in smaller P600 responses to words containing morphosyntactic errors relative to when those same errors are produced by a native speaker (Hanulíková, van Alphen, van Goch, & Weber, 2012), presumably because these forms are higher in probability in the speech of a non-native speaker (see also Cai et al., 2017, for effects of speaker dialect on word-sense retrieval).

Finally, the context of language outside the laboratory affects how participants in language experiments adapt to the distributions within the lab/experiment environment—context on the order of minutes/hours. Listeners appear less likely to adapt to distributions which are in direct conflict with their lifetime of experience, consistent with a Bayesian belief updating model of adaptation to the linguistic environment (Kleinschmidt, Fine, & Jaeger, 2012). For example, as reviewed in Section 2.2.1, listeners anticipate that a noun phrase starting with a size adjective (e.g., "Show me the big…") will ultimately refer to an object in a contrast set (big cup vs small cup). They continue to apply this pragmatic inference even in the face of a very high rate of infelicitous uses of size adjectives in the experimental environment (e.g., on >90% of trials the speaker uses a size adjective when none is needed for disambiguation or fails to use one when there are multiple candidate referents of the same kind differing only in size), though this inference is reduced relative to cases where the use of adjectives is always felicitous (Grodner & Sedivy, 2011; Ryskin et al., 2019). Plausibly, the new input distribution experienced in the lab is partially outweighed by the prior experience in which the presence of the adjective is highly diagnostic of size contrast presence (Sedivy, 2005).

3.1.1 Variability in lifetime context

Psycholinguistic experiments and computational models which investigate probabilistic inferences based on contextual information typically make a key assumption: The probability distributions experienced by the participants (or the distributions in the dataset on which a model is trained) are the same as those which guided stimulus-making (or test dataset creation). For example, for a frequency effect to work in the lab, participants must have experienced words with relative frequencies that are similar to those assumed by the experimenter. If the researcher analyzes the data assuming that "apex" is an infrequent word and "top" is a frequent word but all of their participants have encountered both words an equal number of times in their life, the researcher will presumably fail to observe a frequency effect. For this reason, care is taken to measure the probability of a word, or a word in a sentence, across a large cohort of individuals before creation of linguistic stimuli. The assumption is that these "norms" will reflect the bulk of experience of any individual that comes to participate in the experiment and any variability between people's experiences will be reflected in small, random deviations which can be modeled as noise. Yet, this past experience may vary substantially across language users, in *systematic* ways, and is both of interest in and of itself and an important variable which may lead to spurious conclusions if ignored (Ryskin, Levy, & Fedorenko, 2020).

Starting with the clearest example, a toddler in the midst of acquiring their first language has accumulated very little knowledge of the relative distributions of this language. They are less able to use context to constrain word recognition (Castles, Davis, Cavalot, & Forster, 2007). Children's grammaticality judgments diverge from those of adult, native speakers (Ambridge, Pine, Rowland, & Young, 2008) and they appear less able to predict upcoming input based on contextual information from previous words in the sentence (Gambi, Gorrie, Pickering, & Rabagliati, 2018; Rabagliati, Gambi, & Pickering, 2016). On the other hand, children with more developed vocabularies are more likely to show such context effects (e.g., looking to the treasure after hearing "The pirates hid the…") than those with smaller vocabularies (Borovsky, Elman, & Fernald, 2012; Ylinen, Bosseler, Junttila, & Huotilainen, 2017), suggesting that additional language experience, rather than brain maturation alone, makes the child's language system more adult-like.

Beyond the quantity of experience, the nature of language experience plays a role as well. Children's books contain a diverse inventory of words and an overrepresentation of certain syntactic structures relative to spoken, child-directed language. As a result, children who have more exposure to these books expect different distributions of lexical items and syntactic structures than their peers who primarily receive spoken input (Montag, Jones, & Smith, 2015; Montag & MacDonald, 2015). Similarly, second-language learners have had less experience with the patterns of their non-native language(s) (Ellis, 2002, 2013), which may explain their apparent lack of use of sentential context (Kaan, 2014) and differences in subcategorization biases (Dussias & Cramer Scaltz, 2008; Dussias, Marful, Gerfen, & Molina, 2010). They may additionally experience transfer effects from their first language (MacWhinney, 1995, 1997).

Often, this lifetime of context is assumed to be very similar across adult native speakers of the same language. Yet, exposure to particular genres or books affects the expectations that adult readers bring to an experiment (Troyer & Kutas, 2020) and habitual reading experience is known to explain variance in online and offline reading behaviors (James, Fraundorf, Lee, & Watson, 2018; MacDonald & Christiansen, 2002; Moore & Gordon, 2015). Everyday activities and interests influence language processing as well. For example, job seekers are more likely to predict "skills" after reading "good communication..." than those who have not spent as much time reading job ads (Verhagen, Mos, Backus, & Schilperoord, 2018), and habitual rowers interpret polysemous words in terms of their (typically nondominant) rowing-related meanings (Rodd et al., 2016).

Similarly, older adults have experienced much more linguistic input than the average undergraduate student, which may partly explain the puzzle of aging and language—despite well-documented declines in other cognitive domains (Craik, 1994; Craik & Byrd, 1982; Verhaeghen & Salthouse, 1997), older adults perform better on language comprehension than young adults (Hartshorne & Germine, 2015). Due to the additional years of experience, they tend to have richer, more diverse vocabularies (Meylan & Gahl, 2014; Verhaeghen, 2003) and are more likely to have experienced rare words (Heaps, 1978; Ramscar, Hendrix, Shaoul, Milin, & Baayen, 2014). Network analysis approaches, using words as nodes and the strengths of associations between words (in a free association task) as edges, indicate that the lexicon in older adults is larger but less efficiently structured (Dubossarsky, De Deyne, & Hills, 2017). In particular, the average node in the older adults' network has fewer nodes connected to it, the shortest number of steps required to connect a pair of nodes in the network tends to be longer, and the local clustering coefficient (number of edges between neighboring nodes over the total possible number of edges) is lower, as compared to young adults' networks (Wulff, De Deyne, Jones, & Mata, 2019).

The effects of this additional and more diverse language experience on language processing in context have yet to be fully explored (for reviews see Payne & Silcox, 2019; Peelle, 2019). At the millisecond level, visual word recognition appears to be relatively similar across adulthood (Cohen-Shikora & Balota, 2016). When acoustic information is incomplete, older adults are less successful at inferring the identity of the word than their younger counterparts, but they are able to almost entirely compensate for this deficit when the word is embedded in a sentence (Wingfield, Aberdeen, & Stine, 1991). More fine-grained investigations using ERPs suggest that older adults may be less likely to obtain benefits from the preceding context (on the order of seconds) than young adults. For instance, the effects of a word's predictability on the post-N400 positivity mentioned in Section 2.2.1 are reduced in older adults (Federmeier et al., 2010; Payne & Federmeier, 2018; Wlotko & Federmeier, 2012, see also Dave et al., 2018).

At the level of minutes of context, adaptation to the experimental environment is also likely affected by the distributions that comprehenders have experienced prior to entering the lab. For instance, children appear to be more susceptible to adapting their linguistic distributions than adults (Peter, Chang, Pine, Blything, & Rowland, 2015), possibly because they have less prior knowledge to fall back on so every new learning event represents a larger proportion of their lifetime learning experience. Similarly, second-language learners' patterns of adaptation reflect the relative probabilities of syntactic structures in their prior experience across languages. In a written cumulative priming study, both native English speakers and second-language learners of English whose first language was Korean tuned their probability of producing a syntactic alternative to their input and the change in their probabilities was larger when they were exposed to a structure that was initially less frequent to them: prepositional phrase datives for the native English speakers and double object datives for the second-language learners (Kaan & Chun, 2018). Though no study has, to our knowledge, directly investigated this question across the lifespan, Ryskin, Qi, Covington, Duff, and Brown-Schmidt (2018) found that, in a small sample, older adults failed to adapt their lexico-syntactic distributions to new co-occurrence statistics of verbs and syntactic structures, in contrast to young adults (Ryskin et al., 2017), providing some preliminary suggestion that older adults' language statistics may be less malleable because the lab input is outweighed by their lifetime of experience. In Section 4.2.1, we discuss future directions for addressing this question.

3.2 Historical context

The content of an individual's lifetime of context depends in part on the socio-cultural and historical context within which they exist. Languages evolve and change (e.g., Biber & Finegan, 1989; Bréal, 1904; Bybee, 2015; Lieberman, Michel, Jackson, Tang, & Nowak, 2007; Wolk, Bresnan, Rosenbach, & Szmrecsanyi, 2013). Some pronunciations shift (e.g., the disappearance of the Mid-Atlantic accent), words and structures fall out of fashion (e.g., "davenport"), while others are born (e.g., "to google," "because [Noun]"). Dialects cleave off (Bresnan & Ford, 2010; Labov, 1969; Szmrecsanyi et al., 2017; Tagliamonte, Smith, & Lawrence, 2005), and language contact transforms previously isolated languages (Millar & Trask, 2015). The driver of this language variation and change is beyond the scope of the current piece but several proposals argue that language is shaped by cognitive constraints and/or a pressure for efficient communication (Bybee, 2015; Christiansen & Chater, 2016; Gibson et al., 2019; Kemp, Xu, & Regier, 2018; Kirby, 2017; Mahowald, Fedorenko, Piantadosi, & Gibson, 2013; Piantadosi, Tily, & Gibson, 2011; Zaslavsky, Kemp, Regier, & Tishby, 2018).

To our knowledge, little prior work has explicitly considered language change in relation to context effects during comprehension on shorter timescales. Perhaps this is due to the assumption that language change does not typically occur on a timeframe that should be relevant to the standard psycholinguistic research question (e.g., do co-articulatory cues influence word identification?). The extent to which this is true may differ by domain. For example, lexico-semantic change may be faster than pragmatic change; new innovations or concepts pop up in the world and new labels are created or meanings of old labels are expanded (e.g., tweet, tablets). Indeed, many such changes occur within the lifespan of a typical psycholinguistic participant (Hamilton, Leskovec, & Jurafsky, 2016). Structural changes seem to occur on a more protracted timescale (Lieberman et al., 2007; Wolk et al., 2013) but recent ERP studies have observed markers of structural language change phenomena, as they are happening, in languages (Icelandic, Agnonese dialect of Southern Italian) that are undergoing change presently (Bambini et al., 2021; Bornkessel-Schlesewsky, Roehm, Mailhammer, & Schlesewsky, 2020).

Furthermore, the possibility that linguistic distributions change sufficiently rapidly to impact experiments in the lab is implicit in the practice of updating norms which are used for experimental stimulus design with some regularity. Indeed, Lahar, Tun, and Wingfield (2004) examined sentence completion norms across multiple age cohorts (young: 17–29, middle: 30–59, young-old: 60–74, old-old: 75–91). They observed substantial consistency across the age groups in terms of the most likely completions they provided, particularly for highly constraining sentence contexts, but also some divergences. Intriguingly, they observed a gradation in the correlation for the dominant sentence completion probability such that responses from groups that were closer together in age were more strongly correlated.

They also reported weak correlations with a set of norms collected around 25 years prior. This suggests that an experiment aimed at testing context effects using experimental stimuli that were normed several decades prior may not successfully observe context effects. How rapidly context effects "decline" is an empirical question.

It is particularly interesting to consider how the historical timescale-interacts with the minute and lifetime timescale. Starting first with the lifetime of context, on the perspective adopted here, comprehenders continue updating their model of the language (and the world) as they go through life such that any measurement instance captures the result of some (lossy) integration of language statistics over the individual's entire lifetime until that point. Assuming a stable language, the central tendencies of linguistic distributions would end up the same across comprehenders regardless of the amount of exposure, and more exposure would only lead to more peaked probability distributions. However, given that the language is itself changing, the additional years of exposure may result in more substantial differences in language statistics between young and old, which may partially explain divergences in predictions/inferences across the lifespan. Indeed, this is consistent with the observation that older adults have a more variable lexical inventory (Dubossarsky et al., 2017; Meylan & Gahl, 2014). Similarly, these different probabilistic models of the language—between individuals from different time periods or individuals from the same time period but of different ages—may adapt differently over the course of a conversation or an experimental session due to their different starting conditions.

4. The future of context

In the present piece, we have reframed context effects on language processing in terms of information sampled, with uncertainty, in service of inference about the intended meaning of a given linguistic input, along a continuum of timescales from milliseconds to centuries. Construing of context in this way highlights the importance of considering longer timescales and opens up new questions regarding how information on different timescales can be simultaneously learned and represented. In what follows, we propose novel avenues for addressing these questions.

4.1 Bridging timescales

Interactions between contexts at different timescales are well-trodden territory in psycholinguistics for short timescales. For example, an error in a word (milliseconds of context) that is more predictable given the preceding words in the sentence (seconds of context) is more quickly corrected (Ito et al., 2016). And ambiguous sounds can be maintained in memory and identified once later discourse (minutes of context) resolves the uncertainty regarding the intended word (milliseconds of context) (Brown-Schmidt & Toscano, 2017). Recently, there has been a wave of interest in the effects of minute/hour-level context on the faster timescales. For instance, over the course of an experimental session, participants learn novel sound-meaning mappings such that millisecond-level context (e.g., the surrounding sounds in a word) operates differently after exposure (e.g., Maye et al., 2008; Trude & Brown-Schmidt, 2012). Similarly, adaptation studies show that exposure to manipulated distributions of linguistic input (e.g., co-occurrences of verbs and syntactic structures) over the duration of an experiment (on the order of minutes or hours) changes the probabilities of the words in a sentence (e.g., Fine et al., 2013; Ryskin et al., 2017). Yet, with a few exceptions (e.g., Bambini et al., 2021; Hawkins, Franke, et al., 2021; Kleinschmidt, 2020; Rodd et al., 2016), little attention has been paid to the longer timescales and how they interact with the shorter ones.

In Section 3, we argued that the probability of a word given the preceding and/or subsequent words, for example, will be a reflection of the representation of the language that each individual has constructed based on their experience up to that point. This representation is in turn a reflection of the diachronic change that the language has undergone before and during the learner's lifetime. This view generates a set of testable predictions. The overarching prediction is that the size of context effects (e.g., response difference between input that is predictable in context vs unpredictable), should be maximized whenever the population on which the conditions were normed is similar, in terms of quantity and nature of linguistic experience, to that which is being tested. A more specific version of this prediction is that a context manipulation developed through norming at time t using the standard approach (i.e., collecting data from a large sample of young adults) will elicit a smaller effect at time t with a group of older adults—due to differences in the lifetime of context—and will elicit a smaller effect at time $t + 20$ years[e] with any group of participants including those that are the same age as the norming population was at time t due to both lifetime and historical context. Further, and perhaps more controversially,

---

[e] 20 is purely arbitrary and is just meant to reflect the minimal time over which one could see language change happening; this could be determined empirically from corpus data.

a context manipulation developed through norming at time *t* using data from older adults, will elicit a smaller effect at time *t* with a group of younger adults.[d] Though these predictions are framed around time and age, these are meant as proxies for quantity and nature of input. A thorough assessment of language experience (e.g., time spent reading, time spent in conversation, genres read, occupation, preferred movie genres) or an experience sampling approach may provide a more accurate picture of the quantity and nature of language experience and how they relate to contextual inferences during language processing.

One major obstacle to testing these predictions, and the role of contexts with longer timescales in general, is the methodological challenge of collecting the requisite data. In order to investigate the effects of a person's lifetime of language exposure or of the historical context, both breadth and depth of measurement are needed. For example, very large sample sizes are needed in order to collect cloze norms from a sufficient sample across the entire lifespan, in particular in order to avoid the pitfalls of extreme groups analyses (Salthouse, 2000). More complicated still is the collection of longitudinal contextual norms which could capture language change as it happens. However, arguably the tools to do this on a large scale are now within reach. The proliferation of online crowdsourcing platforms which can reach a more diverse pool of participants (though some selection bias remains) and the success of collaborative projects across many labs which facilitate pooling resources to develop open resources (Frank, 2016) point to avenues which would help language researchers overcome the practical challenges and collect dense, multifaceted, diachronic measurements of probabilities in context.

Similarly, characterizations of the state of the participants' language experience prior to beginning an experimental session can address the question of how the comprehender's lifetime of experience affects how they adapt to the local environment. For instance, the relative frequencies for a pair of syntactic alternations that participants bring to the experiment may affect how much adaptation is observed over the course of the session. A participant whose frequencies closely match those of the experiment will presumably adapt less than one whose representation was initially

---

[d] This assumes that contextual probability distributions are not entirely stable across the lifespan, which seems reasonable for certain linguistic cues, e.g., lexical or syntactic cues, but less so for others. For example, pragmatic cues such as the presence of a size adjective to indicate a contrast set may be very stable over historic time and across an individual's lifespan. No differences in context effects between young and old would be expected for cues of that sort.

quite different. Carefully connecting prior knowledge to adaptation trajectories has the potential to illuminate the learning mechanisms involved (e.g., Chang et al., 2012; Kleinschmidt et al., 2012). One caveat is that any initial measurement of prior knowledge constitutes a learning instance as well and may subtly shift the learner's language statistics (Ryskin & Brown-Schmidt, 2017).

Finally, the existence of linguistic communities undergoing rapid change (e.g., Icelandic) is a unique opportunity to study the complex dynamics of language over time. Using longitudinal measures of processing could help address questions regarding how individual speakers adapt to the language and how the language simultaneously adapts to the speakers (Beckner et al., 2009; Blythe & Croft, 2021; Hawkins, Franke, et al., 2021).

4.2 Constraints on context

4.2.1 Memory

Not all available information within a few hundred millisecond window can be used to inform inference every single time a listener or reader perceives a word segment. The eyes may skip over a relevant piece of information (Staub, Dodge, & Cohen, 2019) and/or attention may lapse. Similarly, memory for the exact words in a sentence is quickly lost (Potter & Lombardi, 1990) and listeners have poorer recall of the referents and labels that were produced during a conversation than speakers do (Yoon, Benjamin, & Brown-Schmidt, 2016, 2021). A complete account of context effects will incorporate the limitations imposed by human cognitive machinery (Griffiths, 2020). Critically, these limitations may differ by timescale.

On the level of milliseconds and seconds, memory may determine which sounds/words/structures are included in the computation of surprisal, as proposed by Futrell et al. (2020). Can the same principles of lossy memory be applied to context on the order of minutes, the lifespan, or historic time? There may be good reason to think no. For instance, one word or letter being lost to deletion noise is plausible, but a whole utterance in a conversational exchange less so. There also may be nonlinearities in how experience is compressed over the lifespan. The simplest possible model would be one in which the current representation of the language is purely an integration over all past experiences, but other functions are possible as well. For example, experiences from early childhood may hold particular weight and/or be more faithfully represented. Similarly, recent language experience may be subject to less decay. Research in autobiographical memory (Brewer, 1986; Rubin & Schulkind, 1997) suggests that experiences between the ages of 10 and 30 may hold a privileged status, though it is unclear whether the same memory mechanisms are involved. Analyses of historical corpora (Davies, 2017) may shed light on these questions by relating contextual probabilities from different time periods to comprehenders' present-day expectations.

In addition, a rich literature on predictive processing in language has revealed a wide array of sources of information that can be used to anticipate upcoming meanings, but less is known about what kind of information comprehenders can maintain uncertainty about, what cues can elicit an update of the previous inference, and how long uncertainty is maintained. Intuitively, one can imagine that previously experienced linguistic information is available for revision for a substantial period of time. For example, following Kraljic et al. (2008) and Liu and Jaeger (2018), if a listener hears a speaker produce a word with a novel pronunciation ("dinoshaur" for "dinosaur"), they will shift their categorization along the acoustic spectrum. If it is later revealed that the speaker had a pen in their mouth, the listener may be able to update their representation of previously heard input and "reverse" their adaptation. This would require maintaining an uncertain representation of the acoustic input or of the phonemic categories over many minutes (c.f. Caplan, Hafri, & Trueswell, 2021).

Similarly, after exchanging written messages with someone who makes many typos, the reader should adapt to the kinds of errors that the writer tends to make (e.g., insertions of 's' due to a sticky keyboard key) and make inferences about the intended meaning of seemingly corrupted words accordingly (e.g., "I sent you the files" will get interpreted as "I sent you the file" if only one attachment is included; Ryskin, Futrell, et al., 2018). However, if the email writer later informs the reader that they got their computer keyboard fixed, the reader may then re-evaluate the prior inference and feel compelled to ask whether they should have received multiple files, since now the probabilities have shifted and the 's' after file is less likely to have been a noise corruption. Whether or not noisy-channel interpretations can be re-evaluated in this way, potentially many minutes later, is an open question.

4.2.2 Complexity

As described in Section 2.3.3, comprehenders learn the idiosyncracies of different speakers such that their model of the linguistic context is hierarchically structured with multiple speakers as clusters of linguistic experiences nested within the larger context (Hawkins, Franke, et al., 2021;

Kleinschmidt & Jaeger, 2015). But other aspects of the environment can fluctuate as well. For example, the light in the office could get switched off midway through reading a paragraph, or a smell could waft into the room during a conversation. As noted in Brown-Schmidt et al. (2015), these are unlikely to elicit any particular adaptation because no direct causal link can be posited between a smell and the particular pronunciation that the listener hears at that moment. In contrast, how a speaker has produced the same phoneme before is clearly informative regarding their upcoming productions (Chodroff & Wilson, 2017). The additional complexity in the model that follows from creating speaker-specific representations of phoneme pronunciation is likely outweighed by the comprehender's enhanced ability to understand their interlocutor. Tracking arbitrary co-occurrences and representing them as sources of variance clustering (e.g., learning smell-specific sound distributions) would likely increase the complexity of the comprehender's language model without improving their ability to accurately infer linguistic meanings. Assuming a pressure for efficient compression both at the level of the language and the language user (e.g., Zaslavsky et al., 2018) predicts that speaker-, environment-, or situation-specific distributions are learned only insofar as they improve inferences and meaning transmission. Corpus analyses and production experiments could reveal what linguistic features are used with high intra-speaker consistency and high inter-speaker variability and should, in principle, be learned in a speaker-specific way. Further, an analogous approach could be used to uncover which additional variables, beyond the speaker identity, constitute important sources of structure in linguistic experience; for instance, genres, registers, modalities (auditory vs visual), age cohorts may all be important sources of clustering in linguistic data. The most efficient hierarchical structure will likely depend on the timescale of context being considered. For example, genres may be an important source of structure when drawing inferences based on longer timescales (minutes and above) but not shorter timescales of context.

5. Conclusion

Language is a noisy and often ambiguous signal, but humans are able to use it to communicate their thoughts by leveraging contextual information on different timescales (see Fig. 1). Context effects on the order of milliseconds and seconds have been extensively studied for decades. For instance, the surrounding phonemes in a word can allow the listener to infer the identity of a missing phoneme. Similarly, readers anticipate upcoming words based on the preceding sequence. They also maintain uncertainty about previously read inputs and can update their inferences in light of downstream context. Recently, substantial work has shown that the larger environment in which language comprehension is taking place affects interpretation as well. Comprehenders track patterns over multiple minutes in order to tune their expectations to the setting (e.g., the conversation partner, the genre, the types of fillers).

In addition, we propose that both an individual's lifetime of language experience and the preceding language change over historical time can also be fruitfully viewed as contexts for language comprehension. In particular, we note that variability in comprehenders' prior language experiences and fluctuations in language probabilities over time plausibly affect the inferences that comprehenders make during lab experiments and, presumably, in everyday conversation. Fruitful future directions for exploring these longer timescale context effects and how they interact with the shorter ones will likely involve innovative approaches to data collection which will allow for larger and more diverse samples as well as more intensive longitudinal psycholinguistic measurements. On the theoretical front, addressing how memory mechanisms and a pressure for efficient compression place constraints on what contextual information is learned and how it is represented holds substantial promise.

References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, *38*(4), 419–439. https://doi.org/10.1006/jmla.1997.2558.

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264. https://doi.org/10.1016/S0010-0277(99)00059-1.

Ambridge, B., Pine, J. M., Rowland, C. F., & Young, C. R. (2008). The effect of verb semantic class and verb frequency (entrenchment) on children's and adults' graded judgements of argument-structure overgeneralization errors. *Cognition*, *106*(1), 87–129. https://doi.org/10.1016/j.cognition.2006.12.015.

Apfelbaum, K. S., Bullock-Rest, N., Rhone, A. E., Jongman, A., & McMurray, B. (2014). Contingent categorisation in speech perception. *Language, Cognition and Neuroscience*, *29*(9), 1070–1082. https://doi.org/10.1080/01690965.2013.824995.

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(3), 340–357.

Bambini, V., Canal, P., Breimaier, F., Meo, D., Pescarini, D., & Loporcaro, M. (2021). Capturing language change through EEG: Weaker P600 for a fading gender value in a southern Italo-Romance dialect. *Journal of Neurolinguistics*, *59*, 101004. https://doi.org/10.1016/j.jneuroling.2021.101004.

Beckner, C., Blythe, R., Bybee, J., Christiansen, M. H., Croft, W., Ellis, N. C., … Schoenemann, T. (2009). Language is a complex adaptive system: Position paper. *Language Learning*, *59*(s1), 1–26. https://doi.org/10.1111/j.1467-9922.2009.00533.x.

Biber, D., & Finegan, E. (1989). Drift and the evolution of English style: A history of three genres. *Language*, *65*(3), 487–517. https://doi.org/10.2307/415220. https://www.jstor.org/stable/415220.

Bicknell, K., Jaeger, T. F., & Tanenhaus, M. K. (2016). Now or … later: Perceptual data are not immediately forgotten during language processing. *Behavioral and Brain Sciences*, *39*. https://doi.org/10.1017/S0140525X15000734. https://www.cambridge.org/core/journals/behavioral-and-brain-sciences/article/now-or-later-perceptual-data-are-not-immediately-forgotten-during-language-processing/BA9A9EB84E2A944630F5F2F4077DB95.

Blank, I. A., & Fedorenko, E. (2020). No evidence for differences among language regions in their temporal receptive windows. *NeuroImage*, *219*, 116925. https://doi.org/10.1016/j.neuroimage.2020.116925.

Blythe, R. A., & Croft, W. (2021). How individuals change language. *PLOS ONE*, *16*(6), e0252582. https://doi.org/10.1371/journal.pone.0252582. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0252582.

Bornkessel-Schlesewsky, I., Roehm, D., Mailhammer, R., & Schlesewsky, M. (2020). Language processing as a precursor to language change: Evidence from Icelandic. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.03013. https://www.frontiersin.org/articles/10.3389/fpsyg.2019.03013/full.

Borovsky, A., Elman, J. L., & Fernald, A. (2012). Knowing a lot for one's age: Vocabulary skill and not age is associated with anticipatory incremental sentence interpretation in children and adults. *Journal of Experimental Child Psychology*, *112*(4), 417–436. https://doi.org/10.1016/j.jecp.2012.01.005.

Bower, G., & Morrow, D. (1990). Mental models in narrative comprehension. *Science*, *247*(4938), 44–48. https://doi.org/10.1126/science.2403694.

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, *106*(2), 707–729. https://doi.org/10.1016/j.cognition.2007.04.005.

Braga, R. M., DiNicola, L. M., Becker, H. C., & Buckner, R. L. (2020). Situating the left-lateralized language network in the broader organization of multiple specialized large-scale distributed networks. *Journal of Neurophysiology*, *124*(5), 1415–1448. https://doi.org/10.1152/jn.00753.2019.

Bréal, M. (1904). *Essai de sémantique (science des significations)*. Hachette.

Brehm, L., Jackson, C. N., & Miller, K. L. (2021). Probabilistic online processing of sentence anomalies. *Language, Cognition and Neuroscience*, 1–25. https://doi.org/10.1080/23273798.2021.1900579.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(6), 1482–1493.

Bresnan, J., & Ford, M. (2010). Predicting syntax: Processing dative constructions in American and Australian varieties of English. *Language*, *86*(1), 168–213. https://doi.org/10.1353/lan.0.0189. http://muse.jhu.edu/content/crossref/journals/language/v086/86.1.bresnan.html.

Brewer, W. F. (1986). What is autobiographical memory? In D. C. Rubin (Ed.), *Autobiographical memory* (pp. 25–49). Cambridge: Cambridge University Press. https://www.cambridge.org/core/books/autobiographical-memory/what-is-autobiographical-memory/AA4D63876B9F5121B6DC0453FE090643.

Brothers, T., Hoversten, L. J., Dave, S., Traxler, M. J., & Swaab, T. Y. (2019). Flexible predictions during listening comprehension: Speaker reliability affects anticipatory processes. *Neuropsychologia*, *135*, 107225. https://doi.org/10.1016/j.neuropsychologia.2019.107225.

Brown, M., Tanenhaus, M. K., & Dilley, L. (2021). Syllable inference as a mechanism for spoken language understanding. *Topics in Cognitive Science*, *13*(2), 351–398. https://doi.org/10.1111/tops.12529.

Brown-Schmidt, S. (2009). Partner-specific interpretation of maintained referential precedents during interactive dialog. *Journal of Memory and Language*, *61*(2), 171–190. https://doi.org/10.1016/j.jml.2009.04.003.

Brown-Schmidt, S., & Konopka, A. E. (2011). Experimental approaches to referential domains and the on-line processing of referring expressions in unscripted conversation. *Information*, *2*(2), 302–326. https://doi.org/10.3390/info2020302. http://www.mdpi.com/2078-2489/2/2/302.

Brown-Schmidt, S., & Tanenhaus, M. (2008). Real-time investigation of referential domains in unscripted conversation: A targeted language game approach. *Cognitive Science: A Multidisciplinary Journal*, *32*(4), 643–684. https://doi.org/10.1080/03640210802066816.

Brown-Schmidt, S., & Toscano, J. C. (2017). Gradient acoustic information induces long-lasting referential uncertainty in short discourses. *Language, Cognition and Neuroscience*, *32*, 1211–1228. https://www.tandfonline.com/doi/abs/10.1080/23273798.2017.1325508.

Brown-Schmidt, S., Yoon, S. O., & Ryskin, R. (2015). People as contexts in conversation. In B. H. Ross (Ed.), *Vol. 62. Psychology of learning and motivation* (pp. 59–99). Elsevier. https://doi.org/10.1016/bs.plm.2014.09.003.

Bruner, J. S., & O'Dowd, D. (1958). A note on the informativeness of parts of words. *Language and Speech*, *1*(2), 98–101. https://doi.org/10.1177/002383095800100203.

Bushong, W., & Jaeger, T. F. (2019). Dynamic re-weighting of acoustic and contextual cues in spoken word recognition. *The Journal of the Acoustical Society of America*, *146*(2), EL135–EL140. https://doi.org/10.1121/1.5119271.

Bybee, J. (2015). *Language change*. Cambridge University Press (May).

Cai, Z. G., Gilbert, R. A., Davis, M. H., Gaskell, M. G., Farrar, L., Adler, S., & Rodd, J. M. (2017). Accent modulates access to word meaning: Evidence for a speaker-model account of spoken word recognition. *Cognitive Psychology*, *98*, 73–101. https://doi.org/10.1016/j.cogpsych.2017.08.003.

Caplan, S., Hafri, A., & Trueswell, J. C. (2021). Now you hear me, later you don't: The immediacy of linguistic Computation and the Representation of speech. *Psychological Science*, *32*(3), 410–423. https://doi.org/10.1177/0956797620968787.

Castles, A., Davis, C., Cavalot, P., & Forster, K. (2007). Tracking the acquisition of orthographic skills in developing readers: Masked priming effects. *Journal of Experimental Child Psychology*, *97*(3), 165–182. https://doi.org/10.1016/j.jecp.2007.01.006.

Chambers, S. M. (1979). Letter and order information in lexical access. *Journal of Verbal Learning and Verbal Behavior*, *18*(2), 225–241. https://doi.org/10.1016/S0022-5371(79)90136-1.

Chang, F., Janciauskas, M., & Fitz, H. (2012). Language adaptation and learning: Getting explicit about implicit learning: Language Adaptation. *Language and Linguistics Compass*, *6*(5), 259–278. https://doi.org/10.1002/lnc3.337.

Chodroff, E., & Wilson, C. (2017). Structure in talker-specific phonetic realization: Covariation of stop consonant VOT in American English. *Journal of Phonetics*, *61*, 30–47.

Christiansen, M. H., & Chater, N. (2016). The now-or-never bottleneck:A fundamental constraint on language. *Behavioral and Brain Sciences*, *39*, E62. https://doi.org/10.1017/S0140525X1500031X. https://www.cambridge.org/core/product/identifier/S0140525X1500031X/type/journal_article.

Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology*, *42*(4), 368–407. https://doi.org/10.1006/cogp.2001.0752.

Clark, H. H. (2021). Anchoring utterances. *Topics in Cognitive Science*, *13*(2), 329–350. https://doi.org/10.1111/tops.12496.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, *22*(1), 1–39. https://doi.org/10.1016/0010-0277(86)90010-7.

Cohen-Shikora, E. R., & Balota, D. A. (2016). Visual word recognition across the adult lifespan. *Psychology and aging*, *31*(5), 488–502. https://doi.org/10.1037/pag0000100. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4980241/.

Cole, R. A. (1973). Listening for mispronunciations: A measure of what we hear during speech. *Perception & Psychophysics*, *13*(1), 153–156. https://doi.org/10.3758/BF03207252. http://link.springer.com/10.3758/BF03207252.

Connine, C. M., Blasko, D. G., & Hall, M. (1991). Effects of subsequent sentence context in auditory word recognition: Temporal and linguistic constraints. *Journal of Memory and Language*, *30*(2), 234–250. https://doi.org/10.1016/0749-596X(91)90005-5.

Connine, C. M., Titone, D., & Wang, J. (1993). Auditory word recognition: Extrinsic and intrinsic effects of word frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(1), 81–94. https://doi.org/10.1037/0278-7393.19.1.81.

Coulson, S., King, J. W., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. In *Language and cognitive processes 13* (pp. 21–58).

Craik, F. I. M. (1994). Memory changes in normal aging. *Current Directions in Psychological Science*, *3*(5), 155–158. https://doi.org/10.1111/1467-8721.ep10770653.

Craik, F. I. M., & Byrd, M. (1982). Aging and cognitive deficits. In F. I. M. Craik, & S. Trehub (Eds.), *Aging and cognitive processes* (pp. 191–211). Boston, MA: Springer US. https://doi.org/10.1007/978-1-4684-4178-9_11.

Creel, S. C., & Tumlin, M. A. (2011). On-line acoustic and semantic interpretation of talker information. *Journal of Memory and Language*, *65*(3), 264–285. https://doi.org/10.1016/j.jml.2011.06.005.

Dahan, D., Drucker, S. J., & Scarborough, R. A. (2008). Talker adaptation in speech perception: Adjusting the signal or the representations? *Cognition*, *108*(3), 710–718. https://doi.org/10.1016/j.cognition.2008.06.003.

Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001). Time course of frequency effects in Spoken-Word recognition: Evidence from Eye movements. *Cognitive Psychology*, *42*(4), 317–367. https://doi.org/10.1006/cogp.2001.0750.

Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes*, *16*(5-6), 507–534. https://doi.org/10.1080/01690960143000074.

Dambacher, M., Kliegl, R., Hofmann, M., & Jacobs, A. M. (2006). Frequency and predictability effects on event-related potentials during reading. *Brain Research*, *1084*(1), 89–103. https://doi.org/10.1016/j.brainres.2006.02.010.

Dave, S., Brothers, T. A., Traxler, M. J., Ferreira, F., Henderson, J. M., & Swaab, T. Y. (2018). Electrophysiological evidence for preserved primacy of lexical prediction in aging. *Neuropsychologia*, *117*, 135–147. https://doi.org/10.1016/j.neuropsychologia.2018.05.023.

Davies, M. (2017). *Corpus of historical American English (COHA)*. https://www.english-corpora.org/coha/.

Delaney-Busch, N., Morgan, E., Lau, E., & Kuperberg, G. R. (2019). Neural evidence for Bayesian trial-by-trial adaptation on the N400 during semantic priming. *Cognition*, *187*, 10–20. https://doi.org/10.1016/j.cognition.2019.01.001.

Dell, G. S., & Chang, F. (2013). The P-chain: Relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1634), 20120394. https://doi.org/10.1098/rstb.2012.0394.

Dimigen, O., Sommer, W., Hohlfeld, A., Jacobs, A. M., Kliegl, R., & Link to external site this link will open in a new window. (2011). Coregistration of eye movements and EEG in natural reading: Analyses and review. *Journal of Experimental Psychology: General*, *140*(4), 552–572. https://doi.org/10.1037/a0023885. https://search.proquest.com/docview/876246171/abstract/3754057B68B54CF5PQ/1.

Ding, S. H. H., Fung, B. C. M., Iqbal, F., & Cheung, W. K. (2019). Learning stylometric representations for authorship analysis. *IEEE Transactions on Cybernetics*, *49*(1), 107–121. https://doi.org/10.1109/TCYB.2017.2766189.

Dubossarsky, H., De Deyne, S., & Hills, T. T. (2017). Quantifying the structure of free association networks across the life span. *Developmental Psychology*, *53*(8), 1560–1570. https://doi.org/10.1037/dev0000347. https://search.proquest.com/docview/1905844624/abstract/D03CF741ED964C15PQ/1.

Dussias, P. E., & Cramer Scaltz, T. R. (2008). Spanish–English L2 speakers' use of subcategorization bias information in the resolution of temporary ambiguity during second language reading. *Acta Psychologica*, *128*(3), 501–513. https://doi.org/10.1016/j.actpsy.2007.09.004.

Dussias, P. E., Marful, A., Gerfen, C., & Molina, M. T. B. (2010). Usage frequencies of complement-taking verbs in Spanish and English: Data from Spanish monolinguals and Spanish—English bilinguals. *Behavior Research Methods*, *42*(4), 1004–1011. https://doi.org/10.3758/BRM.42.4.1004.

Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior*, *20*(6), 641–655. https://doi.org/10.1016/S0022-5371(81)90220-6.

Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, *24*(2), 143–188.

Ellis, N. C. (2013). Second language acquisition. In G. Trousdale, & T. Hoffman (Eds.), *Oxford handbook of construction grammar* (pp. 365–378). Oxford: Oxford University Press. http://www-personal.umich.edu/ncellis/NickEllis/Publications_files/SecondLanguageAcquistionCxGPreprint.pdf.

Falandays, J. B., Brown-Schmidt, S., & Toscano, J. C. (2020). Long-lasting gradient activation of referents during spoken language processing. *Journal of Memory and Language*, *112*, 104088. https://doi.org/10.1016/j.jml.2020.104088.

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, *44*(4), 491–505. https://doi.org/10.1111/j.1469-8986.2007.00531.x.

Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language*, *41*(4), 469–495. https://doi.org/10.1006/jmla.1999.2660.

Federmeier, K. D., Kutas, M., & Schul, R. (2010). Age-related and individual differences in the use of prediction during language comprehension. *Brain and Language*, *115*(3), 149–161. https://doi.org/10.1016/j.bandl.2010.07.006.

Federmeier, K. D., & Laszlo, S. (2009). Chapter 1 Time for meaning. In *Psychology of learning and motivation: Vol. 51* (pp. 1–44). Elsevier. https://doi.org/10.1016/S0079-7421(09)51001-8.

Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Research*, *1146*, 75–84. https://doi.org/10.1016/j.brainres.2006.06.101.

Fedorenko, E., Behr, M. K., & Kanwisher, N. (2011). Functional specificity for high-level linguistic processing in the human brain. *Proceedings of the National Academy of Sciences*, *108*(39), 16428–16433. https://doi.org/10.1073/pnas.1112937108.

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, *47*(2), 164–203. https://doi.org/10.1016/S0010-0285(03)00005-7.

Ferreira, F., & Chantavarin, S. (2018). Integration and prediction in language processing: A synthesis of old and new. *Current Directions in Psychological Science*, *27*(6), 443–448. 0963721418794491. https://doi.org/10.1177/0963721418794491.

Ferreira, F., & Patson, N. D. (2007). The "Good enough" approach to language comprehension. *Language and Linguistics Compass*, *1*(1-2), 71–83. https://doi.org/10.1111/j.1749-818X.2007.00007.x.

Fine, A. B., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). Rapid Expectation Adaptation during Syntactic Comprehension. *PLoS ONE*, *8*(10), e77661. https://doi.org/10.1371/journal.pone.0077661.

Fischer-Baum, S., Charny, J., & McCloskey, M. (2011). Both-edges representation of letter position in reading. *Psychonomic Bulletin & Review*, *18*(6), 1083–1089. https://doi.org/10.3758/s13423-011-0160-3.

Fitz, H., & Chang, F. (2019). Language ERPs reflect learning through prediction error propagation. *Cognitive Psychology*, *111*, 15–52. https://doi.org/10.1016/j.cogpsych.2019.03.002.

Frank, M. C. (2016). (Tech. Rep.). PsyArXi. *A collaborative approach to infant research: Promoting reproducibility, best practices, and theory-building*. https://doi.org/10.31234/osf.io/27b43. https://psyarxiv.com/27b43/.

Frank, S. L., Otten, L. J., Galli, G., & Vigliocco, G. (2015). The ERP response to the amount of information conveyed by words in sentences. *Brain and Language*, *140*, 1–11. https://doi.org/10.1016/j.bandl.2014.10.006.

Fraundorf, S. H., & Jaeger, T. F. (2016). Readers generalize adaptation to newly-encountered dialectal structures to other unfamiliar structures. *Journal of Memory and Language*, *91*, 28–58. https://doi.org/10.1016/j.jml.2016.05.006.

Futrell, R., Gibson, E., & Levy, R. P. (2020). Lossy-Context Surprisal: An Information-Theoretic Model of Memory Effects in Sentence Processing. *Cognitive Science*, *44*(3), e12814. https://doi.org/10.1111/cogs.12814.

Gambi, C., Gorrie, F., Pickering, M. J., & Rabagliati, H. (2018). The development of linguistic prediction: Predictions of sound and meaning in 2- to 5-year-olds. *Journal of Experimental Child Psychology*, *173*, 351–370. https://doi.org/10.1016/j.jecp.2018.04.012.

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, *6*(1), 110–125. https://doi.org/10.1037/0096-1523.6.1.110.

Gardner, B., Dix, S., Lawrence, R., Morgan, C., Sullivan, A., & Kurumada, C. (2021). Online pragmatic interpretations of scalar adjectives are affected by perceived speaker reliability. *PLOS ONE*, *16*(2), e0245130. https://doi.org/10.1371/journal.pone.0245130. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0245130.

Gibson, E., Bergen, L., & Piantadosi, S. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences*, *110*(20), 8051–8056. https://doi.org/10.1073/pnas.1216438110.

Gibson, E., Futrell, R., Piantadosi, S. P., Dautriche, I., Mahowald, K., Bergen, L., & Levy, R. (2019). How efficiency shapes human language. *Trends in Cognitive Sciences*, *23*(5), 389–407. https://doi.org/10.1016/j.tics.2019.02.003.

Gibson, E., Piantadosi, S. T., Brink, K., Bergen, L., Lim, E., & Saxe, R. (2013). A noisy-channel account of crosslinguistic word-order variation. *Psychological Science*, *24*(7), 1079–1088. https://doi.org/10.1177/0956797612463705.

Gibson, E., Tan, C., Futrell, R., Mahowald, K., Konieczny, L., Hemforth, B., & Fedorenko, E. (2017). Don't underestimate the benefits of being misunderstood. *Psychological Science*, *28*(6), 703–712. https://doi.org/10.1177/0956797617690277.

Goldberg, A. E. (2019). *Explain me this: Creativity, competition, and the partial productivity of constructions*. Princeton, New Jersey: Princeton University Press.

Grand, G., Blank, I. A., Pereira, F., & Fedorenko, E. (2018). Semantic projection: Recovering human knowledge of multiple, distinct object features from word embeddings. *arXiv:1802.01241 [cs]*. http://arxiv.org/abs/1802.01241.

Grice, H. P. (1975). Logic and conversation. In *Speech acts* (pp. 41–58). Brill. https://doi.org/10.1163/9789004368811_003.

Griffiths, T. L. (2020). Understanding human intelligence through human limitations. *Trends in Cognitive Sciences*, *24*(11), 873–883. https://doi.org/10.1016/j.tics.2020.09.001.

Grodner, D., & Sedivy, J. (2011). The effect of speaker-specific information on pragmatic inferences. In E. Gibson, & N. J. Pearlmutter (Eds.), *The processing and acquisition of reference* (pp. 239–272). The MIT Press. http://mitpress.universitypressscholarship.com/view/10.7551/mitpress/9780262015127.001.0001/upso-9780262015127-chapter-10.

Gwilliams, L., Linzen, T., Poeppel, D., & Marantz, A. (2018). In spoken word recognition, the future predicts the past. *Journal of Neuroscience*, *38*(35), 7585–7599. https://doi.org/10.1523/JNEUROSCI.0065-18.2018. https://www.jneurosci.org/content/38/35/7585.

Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, *304*(5669), 438–441. https://doi.org/10.1126/science.1095455.

Hale, J. (2001). A probabilistic earley parser as a psycholinguistic model. In *NAACL '01: Proceedings of the second meeting of the North American chapter of the association for computational linguistics on language technologies* (pp. 1–8). Association for Computational Linguistics. https://dl.acm.org/doi/10.3115/1073336.1073357.

Hamilton, W. L., Leskovec, J., & Jurafsky, D. (2016). Diachronic word embeddings reveal statistical laws of semantic change. In *Proceedings of the 54th annual meeting of the Association for Computational Linguistics (Vol. 1: Long Papers)* (pp. 1489–1501). Berlin, Germany: Association for Computational Linguistics. https://aclanthology.org/P16-1141.

Hanulíková, A., van Alphen, P. M., van Goch, M. M., & Weber, A. (2012). When one person's mistake is another's standard usage: The effect of foreign accent on syntactic processing. *Journal of Cognitive Neuroscience*, *24*(4), 878–887. https://doi.org/10.1162/jocn_a_00103.

Harrington Stack, C. M., James, A. N., & Watson, D. G. (2018). A failure to replicate rapid syntactic adaptation in comprehension. *Memory & Cognition*, *46*(6), 864–877. https://doi.org/10.3758/s13421-018-0808-6.

Hartshorne, J. K., & Germine, L. T. (2015). When does cognitive functioning peak? The asynchronous rise and fall of different cognitive abilities across the life span. *Psychological Science*, *26*(4), 433–443. https://doi.org/10.1177/0956797614567339.

Hawkins, R. D., Franke, M., Frank, M. C., Smith, K., Griffiths, T. L., & Goodman, N. D. (2021). From partners to populations: A hierarchical Bayesian account of coordination and convention. *arXiv:2104.05857 [cs]*. http://arxiv.org/abs/2104.05857.

Hawkins, R. D., Gweon, H., & Goodman, N. D. (2021). The division of labor in communication: Speakers help listeners account for asymmetries in visual perspective. *Cognitive Science*, *45*(3), e12926. https://doi.org/10.1111/cogs.12926.

Heaps, H. S. (1978). *Information retrieval, computational and theoretical aspects*. Academic Press.

Hill, F., Reichart, R., & Korhonen, A. (2016). Simlex-999: Evaluating semantic models with (genuine) similarity estimation. *Computational Linguistics*, *41*, 665–695.

Howes, D. H., & Solomon, R. L. (1951). Visual duration threshold as a function of word-probability. *Journal of Experimental Psychology*, *41*(6), 401–410. https://doi.org/10.1037/h0056020. https://search.proquest.com/docview/614251803/abstract/CB76903799D14A5APQ/1.

Huebner, P. A., & Willits, J. A. (2018). Structured semantic knowledge can emerge automatically from predicting word sequences in child-directed speech. *Frontiers in Psychology*, *9*, 133. https://doi.org/10.3389/fpsyg.2018.00133. https://www.frontiersin.org/articles/10.3389/fpsyg.2018.00133/full#B52.

Huettig, F., & Mani, N. (2016). Is prediction necessary to understand language? Probably not. *Language, Cognition and Neuroscience*, *31*(1), 19–31. https://doi.org/10.1080/23273798.2015.1072223.

Ito, A., Corley, M., Pickering, M. J., Martin, A. E., & Nieuwland, M. S. (2016). Predicting form and meaning: Evidence from brain potentials. *Journal of Memory and Language*, *86*, 157–171. https://doi.org/10.1016/j.jml.2015.10.007.

Ito, A., Martin, A. E., & Nieuwland, M. S. (2017). On predicting form and meaning in a second language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(4), 635–652. https://doi.org/10.1037/xlm0000315.

James, A. N., Fraundorf, S. H., Lee, E.-K., & Watson, D. G. (2018). Individual differences in syntactic processing: Is there evidence for reader-text interactions? *Journal of Memory and Language*, *102*, 155–181. https://doi.org/10.1016/j.jml.2018.05.006.

Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science*, *20*(2), 137–194. https://doi.org/10.1207/s15516709cog2002_1.

Kaan, E. (2014). Predictive sentence processing in L2 and L1: What is different? *Linguistic Approaches to Bilingualism*, *4*(2), 257–282. https://doi.org/10.1075/lab.4.2.05kaa. http://www.jbe-platform.com/content/journals/10.1075/lab.4.2.05kaa.

Kaan, E., & Chun, E. (2018). Priming and adaptation in native speakers and second-language learners. *Bilingualism: Language and Cognition*, *21*(2), 228–242. https://doi.org/10.1017/S1366728916001231. https://www.cambridge.org/core/journals/bilingualism-language-and-cognition/article/abs/priming-and-adaptation-in-native-speakers-and-secondlanguage-learners/9A45D19C3FD53DA92623E254DED8F78E.

Kamide, Y. (2012). Learning individual talkers' structural preferences. *Cognition*, *124*(1), 66–71. https://doi.org/10.1016/j.cognition.2012.03.001.

Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, *49*(1), 133–156. https://doi.org/10.1016/S0749-596X(03)00023-8.

Kaschak, M. P., & Glenberg, A. M. (2004). This construction needs learned. *Journal of Experimental Psychology: General*, *133*(3), 450–467. https://doi.org/10.1037/0096-3445.133.3.450.

Kemp, C., Xu, Y., & Regier, T. (2018). Semantic Typology and Efficient Communication. *Annual Review of Linguistics*, *4*(1), 109–128. https://doi.org/10.1146/annurev-linguistics-011817-045406.

Keshev, M., & Meltzer-Asscher, A. (2021). Noisy is better than rare: Comprehenders compromise subject-verb agreement to form more probable linguistic structures. *Cognitive Psychology*, *124*, 101359. https://doi.org/10.1016/j.cogpsych.2020.101359.

Kim, J. S., Elli, G. V., & Bedny, M. (2019). Knowledge of animal appearance among sighted and blind adults. *Proceedings of the National Academy of Sciences*, *116*(23), 11213–11222. https://doi.org/10.1073/pnas.1900952116.

Kirby, S. (2017). Culture and biology in the origins of linguistic structure. *Psychonomic Bulletin & Review*, *24*(1), 118–137. https://doi.org/10.3758/s13423-016-1166-7.

Kleinschmidt, D. (2020). *What constrains distributional learning in adults?*. https://doi.org/10.31234/osf.io/6yhbe. June.

Kleinschmidt, D. F., Fine, A. B., & Jaeger, T. F. (2012). A belief-updating model of adaptation and cue combination in syntactic comprehension. In *Proceedings of the annual meeting of the cognitive science society: Vol. 34* (p. 7).

Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, *122*(2), 148–203. https://doi.org/10.1037/a0038695.

Kraljic, T., & Samuel, A. G. (2011). Perceptual learning evidence for contextually-specific representations. *Cognition*, *121*(3), 459–465. https://doi.org/10.1016/j.cognition.2011.08.015.

Kraljic, T., Samuel, A. G., & Brennan, S. E. (2008). First impressions and last resorts: How listeners adjust to speaker variability. *Psychological Science*, *19*(4), 332–338. https://doi.org/10.1111/j.1467-9280.2008.02090.x.

Kuperberg, G. R., Brothers, T., & Wlotko, E. W. (2020). A tale of two positivities and the N400: Distinct neural signatures are evoked by confirmed and violated predictions at different levels of representation. *Journal of Cognitive Neuroscience*, *32*(1), 12–35. https://doi.org/10.1162/jocn_a_01465.

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, *31*(1), 32–59. https://doi.org/10.1080/23273798.2015.1102299.

Kuperberg, G. R., Kreher, D. A., Sitnikova, T., Caplan, D. N., & Holcomb, P. J. (2007). The role of animacy and thematic relationships in processing active English sentences: Evidence from event-related potentials. *Brain and Language*, *100*(3), 223–237. https://doi.org/10.1016/j.bandl.2005.12.006.

Kurumada, C., Brown, M., & Tanenhaus, M. K. (2018). Effects of distributional information on categorization of prosodic contours. *Psychonomic Bulletin & Review*, *25*(3), 1153–1160. https://doi.org/10.3758/s13423-017-1332-6.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*(1), 621–647. https://doi.org/10.1146/annurev.psych.093008.131123.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*(5947), 161. https://doi.org/10.1038/307161a0. https://www.nature.com/articles/307161a0.

Labov, W. (1969). Contraction, deletion, and inherent variability of the English copula. *Language*, *45*(4), 715–762. https://doi.org/10.2307/412333.

Lahar, C. J., Tun, P. A., & Wingfield, A. (2004). Sentence–Final word completion norms for young, middle-aged, and older adults. *The Journals of Gerontology: Series B*, *59*(1), 7–P10. https://doi.org/10.1093/geronb/59.1.P7. https://academic.oup.com/psychsocgerontology/article/59/1/P7/645382.

Laszlo, S., & Federmeier, K. D. (2009). A beautiful day in the neighborhood: An event-related potential study of lexical relationships and prediction in context. *Journal of Memory and Language*, *61*(3), 326–338. https://doi.org/10.1016/j.jml.2009.06.004.

Leckey, M., & Federmeier, K. D. (2020). The P3b and P600(s): Positive contributions to language comprehension. *Psychophysiology*, *57*(1), e13351. https://doi.org/10.1111/psyp.13351.

Lerner, Y., Honey, C. J., Silbert, L. J., & Hasson, U. (2011). Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. *Journal of Neuroscience*, *31*(8), 2906–2915. https://doi.org/10.1523/JNEUROSCI.3684-10.2011. http://www.jneurosci.org/content/31/8/2906.

Levy, R. (2008a). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177. https://doi.org/10.1016/j.cognition.2007.05.006.

Levy, R. (2008b). A noisy-channel model of rational human sentence comprehension under uncertain input. In *Proceedings of the conference on empirical methods in natural language processing (EMNLP '08)* (pp. 234–243). USA: Association for Computational Linguistics. https://dl.acm.org/doi/10.5555/1613715.1613749.

Levy, R., Bicknell, K., Slattery, T., & Rayner, K. (2009). Eye movement evidence that readers maintain and act on uncertainty about past linguistic input. *Proceedings of the National Academy of Sciences*, *106*(50), 21086–21090. https://doi.org/10.1073/pnas.0907664106.

Lewis, M., Zettersten, M., & Lupyan, G. (2019). Distributional semantics as a source of visual knowledge. *Proceedings of the National Academy of Sciences*, *116*(39), 19237–19238. https://doi.org/10.1073/pnas.1910148116.

Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, *29*(3), 375–419. https://doi.org/10.1207/s15516709cog0000_25.

Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences*, *10*(10), 447–454. https://doi.org/10.1016/j.tics.2006.08.007.

Lieberman, E., Michel, J.-B., Jackson, J., Tang, T., & Nowak, M. A. (2007). Quantifying the evolutionary dynamics of language. *Nature*, *449*, 713–716. https://doi.org/10.1038/nature06137.

Liu, L., Burchill, Z., Tanenhaus, M. K., & Jaeger, T. F. (2017). Failure to replicate talker-specic syntactic adaptation. In *Proceedings of the cognitive science society* (p. 6).

Liu, L., & Jaeger, T. F. (2018). Inferring causes during speech perception. *Cognition*, *174*, 55–70. https://doi.org/10.1016/j.cognition.2018.01.003.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and hearing*, *19*(1), 1–36. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3467695/.

MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology*, *4*, 226. https://www.frontiersin.org/article/10.3389/fpsyg.2013.00226.

MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: Comment on just and carpenter (1992) and waters and caplan (1996). *Psychological Review*, *109*(1), 35–54. https://doi.org/10.1037/0033-295X.109.1.35.

MacWhinney, B. (1995). Language-specific prediction in foreign language learning. *Language Testing*, *12*(3), 292–320. https://doi.org/10.1177/026553229501200303.

MacWhinney, B. (1997). Second language acquisition and the competition model. In J. Kroll, & A. De Groot (Eds.), *Tutorials in bilingualism* (pp. 49–67).

Mahowald, K., Fedorenko, E., Piantadosi, S. T., & Gibson, E. (2013). Info/information theory: Speakers choose shorter words in predictive contexts. *Cognition*, *126*(2), 313–318. https://doi.org/10.1016/j.cognition.2012.09.010.

Marslen-Wilson, W. D. (1975). Sentence perception as an interactive parallel process. *Science*, *189*(4198), 226–228. https://doi.org/10.1126/science.189.4198.226. https://science.sciencemag.org/content/189/4198/226.

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, *10*(1), 29–63. https://doi.org/10.1016/0010-0285(78)90018-X.

Maye, J., Aslin, R. N., & Tanenhaus, M. K. (2008). The weckud wetch of the wast: Lexical adaptation to a novel accent. *Cognitive Science*, *32*(3), 543–562. https://doi.org/10.1080/03640210802035357.

McElree, B., Foraker, S., & Dyer, L. (2003). Memory structures that subserve sentence comprehension. *Journal of Memory and Language*, *48*(1), 67–91. https://doi.org/10.1016/S0749-596X(02)00515-6.

Mcgurk, H., & Macdonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746–748. https://doi.org/10.1038/264746a0. https://www.nature.com/articles/264746a0.

McMurray, B., Clayards, M. A., Tanenhaus, M. K., & Aslin, R. N. (2008). Tracking the time course of phonetic cue integration during spoken word recognition. *Psychonomic Bulletin & Review*, *15*(6), 1064–1071. https://doi.org/10.3758/PBR.15.6.1064.

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2009). Within-category VOT affects recovery from "lexical" garden-paths: Evidence against phoneme-level inhibition. *Journal of Memory and Language*, *60*(1), 65–91. https://doi.org/10.1016/j.jml.2008.07.002.

McQueen, J. M., & Huettig, F. (2012). Changing only the probability that spoken words will be distorted changes how they are recognized. *The Journal of the Acoustical Society of America*, *131*(1), 509–517. https://doi.org/10.1121/1.3664087.

Metzing, C., & Brennan, S. E. (2003). When conceptual pacts are broken: Partner-specific effects on the comprehension of referring expressions. *Journal of Memory and Language*, *49*(2), 201–213. https://doi.org/10.1016/S0749-596X(03)00028-7.

Meylan, S., & Gahl, S. (2014). The divergent lexicon: Lexical overlap decreases with age in a large corpus of conversational speech. In *Proceedings of the annual meeting of the cognitive science society: Vol. 36*. https://escholarship.org/uc/item/4s3757m5.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems*, *26*. https://papers.nips.cc/paper/2013/hash/9aa42b31882ec039965f3c4923ce901b-Abstract.html.

Millar, R. M., & Trask, L. (2015). *Trask's historical linguistics*. Routledge. February.

Mirault, J., Snell, J., & Grainger, J. (2018). You that read wrong again! A Transposed-word effect in grammaticality judgments. *Psychological Science*, *29*(12), 1922–1929. https://doi.org/10.1177/0956797618806296.

Montag, J. L., Jones, M. N., & Smith, L. B. (2015). The words children hear: Picture books and the statistics for language learning. *Psychological science*, *26*(9), 1489–1496. https://doi.org/10.1177/0956797615594361.

Montag, J. L., & MacDonald, M. C. (2015). Text exposure predicts spoken production of complex sentences in 8- and 12-year-old children and adults. *Journal of Experimental Psychology: General*, *144*(2), 447–468. https://doi.org/10.1037/xge0000054.

Moore, M., & Gordon, P. C. (2015). Reading ability and print exposure: Item response theory analysis of the author recognition test. *Behavior Research Methods*, *47*(4), 1095–1109. https://doi.org/10.3758/s13428-014-0534-3.

Mosteller, F., & Wallace, D. L. (1963). Inference in an authorship problem. *Journal of the American Statistical Association*, *58*(302), 275. https://doi.org/10.2307/2283270. https://www.jstor.org/stable/2283270?origin=crossref.

Nieuwland, M. S. (2020). *How 'rational' is semantic prediction? A discussion and reanalysis of Delaney-Busch, Morgan, Lau & Kuperberg (2019)*. https://doi.org/10.31234/osf.io/sm7fq. https://psyarxiv.com/sm7fq/.

Nieuwland, M. S., & Van Berkum, J. J. A. (2006). When peanuts fall in love: N400 evidence for the power of discourse. *Journal of Cognitive Neuroscience*, *18*(7), 1098–1111. https://doi.org/10.1162/jocn.2006.18.7.1098.

Norman-Haignere, S., Kanwisher, N., & McDermott, J. (2015). Distinct cortical pathways for music and speech revealed by hypothesis-free voxel decomposition. *Neuron*, *88*(6), 1281–1296. https://doi.org/10.1016/j.neuron.2015.11.035.

O'Connor, R. E., & Forster, K. I. (1981). Criterion bias and search sequence bias in word recognition. *Memory & Cognition*, *9*, 78–82.

Payne, B. R., & Federmeier, K. D. (2018). Contextual constraints on lexico-semantic processing in aging: Evidence from single-word event-related brain potentials. *Brain Research*, *1687*, 117–128. https://doi.org/10.1016/j.brainres.2018.02.021.

Payne, B. R., Lee, C.-L., & Federmeier, K. D. (2015). Revisiting the incremental effects of context on word processing: Evidence from single-word event-related brain potentials. *Psychophysiology*, *52*(11), 1456–1469. https://doi.org/10.1111/psyp.12515.

Payne, B. R., & Silcox, J. W. (2019). Aging, context processing, and comprehension. In K. D. Federmeier (Ed.), *Vol. 71. Psychology of learning and motivation* (pp. 215–264). Elsevier.

Peelle, J. E. (2019). Language and aging. In G. I. de Zubicaray, & N. O. Schiller (Eds.), *The Oxford handbook of neurolinguistics* (pp. 296–316). Oxford University Press.

Pennington, J., Socher, R., & Manning, C. (2014). GloVe: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 1532–1543). Doha, Qatar: Association for Computational Linguistics. https://www.aclweb.org/anthology/D14-1162.

Peter, M., Chang, F., Pine, J. M., Blything, R., & Rowland, C. F. (2015). When and how do children develop knowledge of verb argument structure? Evidence from verb bias effects in a structural priming task. *Journal of Memory and Language*, *81*, 1–15. https://doi.org/10.1016/j.jml.2014.12.002.

Piantadosi, S. T., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences*, *108*(9), 3526–3529. https://doi.org/10.1073/pnas.1012551108.

Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin*, *144*(10), 1002–1044. https://doi.org/10.1037/bul0000158.

Politzer-Ahles, S., Lee, K. K., & Shen, L. (2020). Ganong effects for frequency may not be robust. *The Journal of the Acoustical Society of America*, *147*(1), EL37–EL42. https://doi.org/10.1121/10.0000562.

Poppels, T., & Levy, R. P. (2016). Structure-sensitive noise inference: Comprehenders expect exchange errors. In *Proceedings of the 38th annual conference of the cognitive science society* (pp. 378–383).

Potter, M. C., & Lombardi, L. (1990). Regeneration in the short-term recall of sentences. *Journal of Memory and Language*, *29*(6), 633–654. https://doi.org/10.1016/0749-596X(90)90042-X.

Rabagliati, H., Gambi, C., & Pickering, M. J. (2016). Learning to predict or predicting to learn? *Language, Cognition and Neuroscience*, *31*(1), 94–105. https://doi.org/10.1080/23273798.2015.1077979.

Rabovsky, M., Hansen, S. S., & McClelland, J. L. (2018). Modelling the N400 brain potential as change in a probabilistic representation of meaning. *Nature Human Behaviour*, *2*(9), 693. https://doi.org/10.1038/s41562-018-0406-4. https://www.nature.com/articles/s41562-018-0406-4.

Ramscar, M., Hendrix, P., Shaoul, C., Milin, P., & Baayen, H. (2014). The myth of cognitive decline: Non-linear dynamics of lifelong learning. *Topics in Cognitive Science*, *6*(1), 5–42. https://doi.org/10.1111/tops.12078.

Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology*, *81*(2), 275–280. https://doi.org/10.1037/h0027768.

Rodd, J. M., Cai, Z. G., Betts, H. N., Hanby, B., Hutchinson, C., & Adler, A. (2016). The impact of recent and long-term experience on access to word meanings: Evidence from large-scale internet-based experiments. *Journal of Memory and Language*, *87*, 16–37. https://doi.org/10.1016/j.jml.2015.10.006.

Rubin, D. C., & Schulkind, M. D. (1997). The distribution of autobiographical memories across the lifespan. *Memory & Cognition*, *25*(6), 859–866. https://doi.org/10.3758/BF03211330.

Ryskin, R., & Brown-Schmidt, S. (2017). The malleability of linguistic representations poses a challenge to the priming-based experimental approach [Commentary]. *Behavioral and Brain Sciences*, *40*, E309. https://doi.org/10.1017/S0140525X17000577. https://search.proquest.com/openview/f13544818665452562d6555e78e4f21c/1?pq-origsite=gscholar&cbl=47829.

Ryskin, R., Futrell, R., Kiran, S., & Gibson, E. (2018). Comprehenders model the nature of noise in the environment. *Cognition*, *181*, 141–150. https://doi.org/10.1016/j.cognition.2018.08.018.

Ryskin, R., Kurumada, C., & Brown-Schmidt, S. (2019). Information integration in modulation of pragmatic inferences during online language comprehension. *Cognitive Science*, *43*(8), e12769. https://doi.org/10.1111/cogs.12769.

Ryskin, R., Levy, R. P., & 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. https://doi.org/10.1016/j.neuropsychologia.2019.107258.

Ryskin, R., Ng, S., Mimnaugh, K., Brown-Schmidt, S., & Federmeier, K. D. (2020). Talker-specific predictions during language processing. *Language, Cognition and Neuroscience*, *35*(6), 797–812. https://doi.org/10.1080/23273798.2019.1630654. https://www.tandfonline.com/doi/citedby/10.1080/23273798.2019.1630654.

Ryskin, R., Qi, Z., Covington, N. V., Duff, M., & Brown-Schmidt, S. (2018). Knowledge and learning of verb biases in amnesia. *Brain and Language*, *180-182*, 62–83. https://doi.org/10.1016/j.bandl.2018.04.003.

Ryskin, R., Qi, Z., Duff, M. C., & Brown-Schmidt, S. (2017). Verb biases are shaped through lifelong learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(5), 781–794. https://doi.org/10.1037/xlm0000341.

Ryskin, R., Stearns, L., Bergen, L., Eddy, M., Fedorenko, E., & Gibson, E. (2021). An ERP index of real-time error correction within a noisy-channel framework of human communication. *Neuropsychologia*, *158*, 107855. https://doi.org/10.1016/j.neuropsychologia.2021.107855.

Ryskin, R., Stevenson, S., & Heller, D. (2020). Probabilistic weighting of perspectives in dyadic communication. In *Proceedings of the 42nd annual conference of the cognitive science society, Toronto, CA* (pp. 252–258).

Ryskin, R., Wang, R. F., & Brown-Schmidt, S. (2016). Listeners use speaker identity to access representations of spatial perspective during online language comprehension. *Cognition*, *147*, 75–84. https://doi.org/10.1016/j.cognition.2015.11.011.

Salthouse, T. A. (2000). Methodological assumptions in cognitive aging research. In F. I. M. Craik, & T. A. Salthouse (Eds.), *Vol. 2. The handbook of aging and cognition* (pp. 467–498). Lawrence Erlbaum Associates Publishers.

Salverda, A. P., Kleinschmidt, D., & Tanenhaus, M. K. (2014). Immediate effects of anticipatory coarticulation in spoken-word recognition. *Journal of Memory and Language*, *71*(1), 145–163. https://doi.org/10.1016/j.jml.2013.11.002.

Samuel, A. G. (1981). Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General*, *110*, 474–494.

Schuster, S., & Degen, J. (2020). I know what you're probably going to say: Listener adaptation to variable use of uncertainty expressions. *Cognition*, *203*, 104285. https://doi.org/10.1016/j.cognition.2020.104285.

Sedivy, J. (2005). Evaluating explanations for referential context effects: Evidence for Gricean mechanisms in online language interpretation. In M. Tanenhaus, & J. C. Trueswell (Eds.), *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions* (pp. 345–364). MIT Press.

Sedivy, J., Tanenhaus, M., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition*, *71*(2), 109–147. https://doi.org/10.1016/S0010-0277(99)00025-6.

Shrestha, P., Sierra, S., González, F., Montes, M., Rosso, P., & Solorio, T. (2017). Convolutional neural networks for authorship attribution of short texts. In *Proceedings of the 15th conference of the European chapter of the association for computational linguistics: Volume 2, short papers* (pp. 669–674). Valencia, Spain: Association for Computational Linguistics. https://www.aclweb.org/anthology/E17-2106.

Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, *128*(3), 302–319. https://doi.org/10.1016/j.cognition.2013.02.013.

Snedeker, J., & Trueswell, J. C. (2004). The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive Psychology*, *49*(3), 238–299. https://doi.org/10.1016/j.cogpsych.2004.03.001.

Staub, A., Dodge, S., & Cohen, A. L. (2019). Failure to detect function word repetitions and omissions in reading: Are eye movements to blame? *Psychonomic Bulletin & Review*, *26*, 340–346. https://doi.org/10.3758/s13423-018-1492-z.

Szmrecsanyi, B., Grafmiller, J., Bresnan, J., Rosenbach, A., Tagliamonte, S., & Todd, S. (2017). Spoken syntax in a comparative perspective: The dative and genitive alternation in varieties of English. *Glossa: A Journal of General Linguistics*, *2*(1), 86. https://doi.org/10.5334/gjgl.310. http://www.glossa-journal.org//articles/abstract/10.5334/gjgl.310/.

Szostak, C. M., & Pitt, M. A. (2013). The prolonged influence of subsequent context on spoken word recognition. *Attention, Perception, & Psychophysics*, *75*(7), 1533–1546. https://doi.org/10.3758/s13414-013-0492-3.

Tagliamonte, S., Smith, J., & Lawrence, H. (2005). No taming the vernacular! Insights from the relatives in northern Britain. *Language Variation and Change*, *17*(1), 75–112. https://doi.org/10.1017/S0954394505050040. https://search.proquest.com/docview/218158463/abstract/523384DB46884DACPQ/1.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632–1634.

Tooley, K. M., Swaab, T. Y., Boudewyn, M. A., Zirnstein, M., & Traxler, M. J. (2014). Evidence for priming across intervening sentences during on-line sentence comprehension. *Language, Cognition and Neuroscience*, *29*(3), 289–311. https://doi.org/10.1080/01690965.2013.770892.

Toscano, J. C., & McMurray, B. (2012). Cue-integration and context effects in speech: Evidence against speaking-rate normalization. *Attention, Perception, & Psychophysics*, *74*(6), 1284–1301. https://doi.org/10.3758/s13414-012-0306-z.

Toscano, J. C., & McMurray, B. (2015). The time-course of speaking rate compensation: Effects of sentential rate and vowel length on voicing judgments. *Language, Cognition and Neuroscience*, *30*(5), 529–543. https://doi.org/10.1080/23273798.2014.946427.

Troyer, M., & Kutas, M. (2020). To catch a Snitch: Brain potentials reveal variability in the functional organization of (fictional) world knowledge during reading. *Journal of Memory and Language*, *113*, 104111. https://doi.org/10.1016/j.jml.2020.104111.

Trude, A. M., & Brown-Schmidt, S. (2012). Talker-specific perceptual adaptation during online speech perception. *Language and Cognitive Processes*, *27*(7-8), 979–1001.

Trueswell, J. C., Tanenhaus, M. K., & Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *19*(3), 528–553.

van de Meerendonk, N., Indefrey, P., Chwilla, D. J., & Kolk, H. (2011). Monitoring in language perception: Electrophysiological and hemodynamic responses to spelling violations. *NeuroImage*, *54*(3), 2350–2363. https://doi.org/10.1016/j.neuroimage.2010.10.022.

Van Berkum, J. J. A., van den Brink, D., Tesink, C. M. J. Y., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience*, *20*(4), 580–591. https://doi.org/10.1162/jocn.2008.20054.

Van Petten, C., & Kutas, M. (1991). Influences of semantic and syntactic context on open- and closed-class words. *Memory & Cognition*, *19*(1), 95–112. https://doi.org/10.3758/BF03198500.

Verhaeghen, P. (2003). Aging and vocabulary score: A meta-analysis. *Psychology and Aging*, *18*(2), 332–339. https://doi.org/10.1037/0882-7974.18.2.332.

Verhaeghen, P., & Salthouse, T. A. (1997). Meta-analyses of age-cognition relations in adulthood: Estimates of linear and nonlinear age effects and structural models. *Psychological Bulletin*, *122*(3), 231–249. https://doi.org/10.1037/0033-2909.122.3.231.

Verhagen, V., Mos, M., Backus, A., & Schilperoord, J. (2018). Predictive language processing revealing usage-based variation. *Language and Cognition*, *10*(2), 329–373. https://doi.org/10.1017/langcog.2018.4. https://www.cambridge.org/core/journals/language-and-cognition/article/predictive-language-processing-revealing-usagebased-variation/876B7947FB92BA6A6F7A7E65446B85A8.

von der Malsburg, T., Poppels, T., & Levy, R. P. (2020). Implicit gender bias in linguistic descriptions for expected events: The cases of the 2016 United States and 2017 United Kingdom elections. *Psychological Science*, *31*(2), 115–128. https://doi.org/10.1177/0956797619890619.

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science*, *167*(3917), 392–393. https://doi.org/10.1126/science.167.3917.392. https://science.sciencemag.org/content/167/3917/392.

Wells, J., Christiansen, M., Race, D., Acheson, D., & Macdonald, M. (2009). Experience and sentence processing: Statistical learning and relative clause comprehension. *Cognitive Psychology*, *58*(2), 250–271. https://doi.org/10.1016/j.cogpsych.2008.08.002.

Wheeler, D. D. (1970). Processes in word recognition. *Cognitive Psychology*, *1*(1), 59–85. https://doi.org/10.1016/0010-0285(70)90005-8.

Wilkes-Gibbs, D., & Clark, H. H. (1992). Coordinating beliefs in conversation. *Journal of Memory and Language*, *31*(2), 183–194. https://doi.org/10.1016/0749-596X(92)90010-U.

Wingfield, A., Aberdeen, J. S., & Stine, E. A. L. (1991). Word onset gating and linguistic context in spoken word recognition by young and elderly adults. *Journal of Gerontology*, *46*(3), 127–P129. https://doi.org/10.1093/geronj/46.3.P127.

Wlotko, E. W., & Federmeier, K. D. (2012). Age-related changes in the impact of contextual strength on multiple aspects of sentence comprehension: Context use in young and older adults. *Psychophysiology*, *49*(6), 770–785. https://doi.org/10.1111/j.1469-8986.2012.01366.x.

Wolk, C., Bresnan, J., Rosenbach, A., & Szmrecsanyi, B. (2013). Dative and genitive variability in late modern English: Exploring cross-constructional variation and change. *Diachronica*, *30*(3), 382–419. https://doi.org/10.1075/dia.30.3.04wol. https://www.jbe-platform.com/content/journals/10.1075/dia.30.3.04wol.

Wulff, D. U., De Deyne, S., Jones, M. N., & Mata, R. (2019). New perspectives on the aging lexicon. *Trends in Cognitive Sciences*, *23*(8), 686–698. https://doi.org/10.1016/j.tics.2019.05.003.

Ylinen, S., Bosseler, A., Junttila, K., & Huotilainen, M. (2017). Predictive coding accelerates word recognition and learning in the early stages of language development. *Developmental Science*, *20*(6), e12472. https://doi.org/10.1111/desc.12472.

Yoon, S. O., Benjamin, A. S., & Brown-Schmidt, S. (2016). The historical context in conversation: Lexical differentiation and memory for the discourse history. *Cognition*, *154*, 102–117. https://doi.org/10.1016/j.cognition.2016.05.011.

Yoon, S. O., Benjamin, A. S., & Brown-Schmidt, S. (2021). Referential form and memory for the discourse history. *Cognitive Science*, *45*(4), e12964. https://doi.org/10.1111/cogs.12964.

Yoon, S. O., & Brown-Schmidt, S. (2014). Adjusting conceptual pacts in three-party conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(4), 919–937. https://doi.org/10.1037/a0036161.

Yoon, S. O., Koh, S., & Brown-Schmidt, S. (2012). Influence of perspective and goals on reference production in conversation. *Psychonomic Bulletin & Review*, *19*(4), 699–707. https://doi.org/10.3758/s13423-012-0262-6.

Zaslavsky, N., Kemp, C., Regier, T., & Tishby, N. (2018). Efficient compression in color naming and its evolution. *Proceedings of the National Academy of Sciences*, *115*(31), 7937–7942. https://doi.org/10.1073/pnas.1800521115.

Zellou, G., & Dahan, D. (2019). Listeners maintain phonological uncertainty over time and across words: The case of vowel nasality in English. *Journal of Phonetics*, *76*, 100910. https://doi.org/10.1016/j.wocn.2019.06.001.