frontiers | Research Topics

Variability in language predictions: Assessing the influence of speaker, text and experimental method

Edited by
Franziska Kretzschmar, Martine Grice, Phillip M. Alday,
Roel M. Willems and Ingmar Brilmayer

Published in
Frontiers in Communication
Frontiers in Psychology

ISSN 1664-8714
ISBN 978-2-8325-2639-2
DOI 10.3389/978-2-8325-2639-2

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Variability in language predictions: Assessing the influence of speaker, text and experimental method

Topic editors

Franziska Kretzschmar — Leibniz Institute for the German Language (IDS), Germany
Martine Grice — University of Cologne, Germany
Phillip M. Alday — Max Planck Institute for Psycholinguistics, Netherlands
Roel M. Willems — Radboud University, Netherlands
Ingmar Brilmayer — University of Cologne, Germany

Citation

Kretzschmar, F., Grice, M., Alday, P. M., Willems, R. M., Brilmayer, I., eds. (2023). *Variability in language predictions: Assessing the influence of speaker, text and experimental method*. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-8325-2639-2

Table of contents

frontiers | Frontiers in Communication

TYPE Editorial
PUBLISHED 22 May 2023
DOI 10.3389/fcomm.2023.1216399

OPEN ACCESS

EDITED AND REVIEWED BY
Xiaolin Zhou,
Peking University, China

*CORRESPONDENCE
Franziska Kretzschmar
kretzschmar@ids-mannheim.de

RECEIVED 03 May 2023
ACCEPTED 09 May 2023
PUBLISHED 22 May 2023

CITATION
Kretzschmar F, Alday PM, Grice M and Brilmayer I (2023) Editorial: Variability in language predictions: assessing the influence of speaker, text and experimental method. *Front. Commun.* 8:1216399.
doi: 10.3389/fcomm.2023.1216399

Editorial: Variability in language predictions: assessing the influence of speaker, text and experimental method

Franziska Kretzschmar[1]*, Phillip M. Alday[2], Martine Grice[3] and Ingmar Brilmayer[4]

[1]Department of Grammar, Leibniz Institute for the German Language, Mannheim, Germany, [2]Department of Psychology, Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands, [3]Institut für Linguistik–Phonetik, University of Cologne, Cologne, Germany, [4]Department of German Language and Literature I–Linguistics, University of Cologne, Cologne, Germany

KEYWORDS

prediction, language processing, variability, experimental linguistics, human cognition

Editorial on the Research Topic
Variability in language predictions: assessing the influence of speaker, text and experimental method

A central question in cognitive science is which mechanisms enable humans to filter relevant input from the environment, process it and then respond quickly and accurately. One important mechanism for information processing is prediction (or related concepts such as anticipation and expectation), which enables speculative information processing in advance of perception (cf. Friston, 2005, 2010; Clark, 2013). In language processing, the benefits of prediction typically appear as faster and more precise behavioral responses or altered neural responses (cf. Federmeier, 2007; Huettig, 2015; Tavano and Scharinger, 2015 for overviews).

However, the exact status and form of prediction in language processing remains controversial (e.g., Pickering and Garrod, 2013; Dell and Chang, 2014; Huettig, 2015; Bornkessel-Schlesewsky and Schlesewsky, 2019). For example, there is converging evidence that comprehenders predict lexical-conceptual units in sentences and beyond (e.g., Kutas and Hillyard, 1984; Altmann and Kamide, 1999; Metusalem et al., 2012; Hosemann et al., 2013). Form-based/sub-lexical information types (phonetic/phonological, orthographic, morphological), meanwhile, show varying effects depending on the specifics of the experimental protocol, e.g., the experimental method, participant sample or text characteristics (e.g., Balota et al., 1985; DeLong et al., 2005; Mishra et al., 2012; Freunberger and Roehm, 2016; Ito et al., 2016; Nieuwland, 2019).

We propose that the nature and strength of prediction in language are shaped by the same variables that influence language processing in general. This Research Topic presents a collection of articles that focus on the extent to which linguistic predictions depend on three main sources of variability in language processing: individual differences, variation in text type and modality, and differences in methodological approaches.

The first group of articles address the relationship between individual differences and prediction.

Hestvik et al. examine differences between children with typical development and children with Developmental Language Disorder (DLD) in a classical filler-gap ERP paradigm. They find that children with DLD do not show an early anterior negativity that that children with typical development do, indicating reduced prediction in DLD.

In their opinion article, Scholten et al. propose that differences in communicative behavior in individuals with autism spectrum disorder (ASD) may be explained by a reduced ability to predict upcoming information of all sorts. They review empirical findings showing that autistic individuals are "less surprised when their predictions are being violated" compared with controls.

Together, Hestvik et al. and Scholten et al. present evidence that inter-individual differences influence prediction strength, with weaker predictions in the two disorders investigated than assumed for the general population.

Theimann et al. investigate prediction strength and intra-individual differences in a sample of typically developing bilingual toddlers. Using the visual-world paradigm, they report that toddlers predict nouns faster following constraining verbs in their dominant language. This finding suggests that the effect of language dominance on prediction converges with the effects from other aspects of language experience in a typically developing participant sample (e.g., Mani and Huettig, 2014).

A second group of articles emphasize text-based and modality-based influences on predictions. Using corpus-based analyses, Egetenmeyer investigates tense-aspect-mood (TAM) forms in German and French football language. His analyses reveal that TAM forms in spoken football reports shift temporal reference across both languages compared with other genres. Moreover, listeners can use script knowledge to predict this shift, supporting experimental evidence for script knowledge as a basis for predictions (Metusalem et al., 2012).

Henrich and Scharinger tested whether metered speech influences the prediction of phonological stress. Using pseudowords in an omission mismatch negativity (oMMN) paradigm, they omitted a syllable of a trochee or iamb. Their results showed that omissions in the first syllable elicited larger and earlier oMMNs for trochees, i.e., the preferred foot in German, while omissions in the second syllable elicited larger oMMNs for iambs without a latency effect. Thus, predictive processing seems to play a particular role in metered speech, especially for the preferred foot type (cf. Wiese and Speyer, 2015).

Danner et al. and Krause and Kawamoto are both concerned with how movement is affected by prediction at turn transitions in dyadic communication. Danner et al. examined co-speech gestures in conversation and nursery rhymes using electromagnetic articulography. They found that brow and head movements are denser as speakers approach overlapping turn exchanges (as opposed to non-overlapping ones), with greater movement density on non-rhyme related speech content. Moreover, listeners generally produced more co-speech movement than speakers. Although the role of co-speech gesture in facilitating turn-end prediction is unclear, speakers' and listeners' co-speech movements pattern jointly in conversational interaction.

Krause and Kawamoto examined anticipatory postures for speech in the lip area before a turn transition using video motion tracking. The authors detected preparatory lip shapes indicating labiality (e.g., labial consonants and rounded vowels) before the acoustic onset of speech and considerably earlier than in less ecologically valid tasks. The authors propose that speakers can initiate articulation from ongoing prediction of the next speech opportunity and that planning and articulation can flexibly overlap in conversational contexts.

Finally, McConnell and Blumenthal-Dramé and Bornkessel-Schlesewsky et al. show that individual differences in language experience and the dynamics of variability in other language users (i.e., capacity and biology) can have profound impacts on language processing strategy, including the use of prediction.

McConnell and Blumenthal-Dramé examine the impact of language experience on processing of bigrams in a self-paced reading task in English, focusing on forward and backward transition probabilities estimated from a large corpus. They find that age and language experience influence the impact of transition probability on reading times, thus suggesting that prediction strategies vary strongly based on individual experience.

Bornkessel-Schlesewsky et al. used both electrophysiological and behavioral measures to study individual differences in listener adaption to speaker idiosyncrasies, thus capturing the impact of variability at two levels. They find that individuals with a steep aperiodic slope and low individual alpha frequency adapt most quickly to speaker idiosyncrasies as shown by changing N400 attunement over the course of the experiment.

Overall, the present collection of articles present further evidence for the importance of prediction and for the need to further investigate its interaction with varying experimental approaches.

Author contributions

FK wrote the first draft of the manuscript. PA, MG, and IB added and edited sections. All authors contributed to the Research Topic as Topic Editors, manuscript revision, read, and approved the submitted version.

Funding

Initial work on the Research Topic was supported by the Collaborative Research Center 1252 "Prominence in Language" [Project ID 281511265, funded by the German Research Foundation (DFG)] in the projects B07 "Agentivity as a key to prominence: Experimental approaches to argument alternations in German" and A01 "Intonation and attention orienting: Neurophysiological and behavioral correlates".

Acknowledgments

We thank the authors and reviewers who contributed to this Research Topic. We also thank Roel M. Willems for joining us as a Topic Editor.

Conflict of interest

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

Publisher's note

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

References

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

Balota, D. A., Pollatsek, A., and Rayner, K. (1985). The interaction of contextual constraints and parafoveal visual information in reading. *Cogn. Psychol.* 17, 364–390. doi: 10.1016/0010-0285(85)90013-1

Bornkessel-Schlesewsky, I., and Schlesewsky, M. (2019). Toward a neurobiologically plausible model of language-related, negative event-related potentials. *Front. Psychol.* 10, 298. doi: 10.3389/fpsyg.2019.00298

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* 36, 181–204. doi: 10.1017/S0140525X12000477

Dell, G. S., and Chang, F. (2014). The P-chain: Relating sentence production and its disorders to comprehension and acquisition. *Philos. Trans. R Soc. B.* 369, 20120394. doi: 10.1098/rstb.2012.0394

DeLong, K. A., Urbach, T. P., and Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nat. Neurosci.* 8, 1117–1121. doi: 10.1038/nn1504

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

Freunberger, D., and Roehm, D. (2016). Semantic prediction in language comprehension: evidence from brain potentials. *Lang. Cogn. Neurosci.* 31, 1193–1205. doi: 10.1080/23273798.2016.1205202

Friston, K. (2005). A theory of cortical responses. *Philos. Trans. R Soc. B.* 360, 815–836. doi: 10.1098/rstb.2005.1622

Friston, K. J. (2010). The free-energy principle: A unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138. doi: 10.1038/nrn2787

Hosemann, J., Herrmann, A., Steinbach, M., Bornkessel-Schlesewsky, I., and Schlesewsky, M. (2013). Lexical prediction via forward models: N400 evidence from German sign language. *Neuropsychologia* 51, 2224–2237. doi: 10.1016/j.neuropsychologia.2013.07.013

Huettig, F. (2015). Four central questions about prediction in language processing. *Brain Res.* 1626, 118–135. doi: 10.1016/j.brainres.2015.02.014

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

Kutas, M., and Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature* 307, 161–163. doi: 10.1038/307161a0

Mani, N., and Huettig, F. (2014). Word reading skill predicts anticipation of upcoming spoken language input: A study of children developing proficiency in reading. *J. Exper. Child Psychol.* 126, 264–279. doi: 10.1016/j.jecp.2014.05.004

Metusalem, R., Kutas, M., Urbach, T. P., Hare, M., McRae, K., and Elman, J. L. (2012). Generalized event knowledge activation during online sentence comprehension. *J. Memory Lang.* 66, 545–567. doi: 10.1016/j.jml.2012.01.001

Mishra, R. K., Singh, N., Pandey, A., and Huettig, F. (2012). Spoken language-mediated anticipatory eye movements are modulated by reading ability: Evidence from Indian low and high literates. *J. Eye Move. Res.* 5, 1–10. doi: 10.16910/jemr.5.1.3

Nieuwland, M. (2019). Do 'early' brain responses reveal word form prediction during language comprehension? A critical review. *Neurosci. Biobehav. Rev.* 96, 367–400. doi: 10.1016/j.neubiorev.2018.11.019

Pickering, M. J., and Garrod, S. (2013). An integrated theory of language production and comprehension. *Behav. Brain Sci.* 36, 329–347. doi: 10.1017/S0140525X1200 1495

Tavano, A., and Scharinger, M. (2015). Prediction in speech and language processing. *Cortex.* 68, 1–7. doi: 10.1016/j.cortex.2015.05.001

Wiese, R., and Speyer, A. (2015). Prosodic parallelism explaining morphophonological variation in German. *Linguistics* 53, 525–559. doi: 10.1515/ling-2015-0011

frontiers in Psychology

ORIGINAL RESEARCH
published: 13 July 2021
doi: 10.3389/fpsyg.2021.684248

Predicting One's Turn With Both Body and Mind: Anticipatory Speech Postures During Dyadic Conversation

Peter A. Krause[1,2]* and Alan H. Kawamoto[2]

[1] Department of Psychology, California State University Channel Islands, Camarillo, CA, United States, [2] Department of Psychology, University of California, Santa Cruz, Santa Cruz, CA, United States

In natural conversation, turns are handed off quickly, with the mean downtime commonly ranging from 7 to 423 ms. To achieve this, speakers plan their upcoming speech as their partner's turn unfolds, holding the audible utterance in abeyance until socially appropriate. The role played by prediction is debated, with some researchers claiming that speakers predict upcoming speech opportunities, and others claiming that speakers wait for detection of turn-final cues. The dynamics of articulatory triggering may speak to this debate. It is often assumed that the prepared utterance is held in a response buffer and then initiated all at once. This assumption is consistent with standard phonetic models in which articulatory actions must follow tightly prescribed patterns of coordination. This assumption has recently been challenged by single-word production experiments in which participants partly positioned their articulators to anticipate upcoming utterances, long before starting the acoustic response. The present study considered whether similar anticipatory postures arise when speakers in conversation await their next opportunity to speak. We analyzed a pre-existing audiovisual database of dyads engaging in unstructured conversation. Video motion tracking was used to determine speakers' lip areas over time. When utterance-initial syllables began with labial consonants or included rounded vowels, speakers produced distinctly smaller lip areas (compared to other utterances), prior to audible speech. This effect was moderated by the number of words in the upcoming utterance; postures arose up to 3,000 ms before acoustic onset for short utterances of 1–3 words. We discuss the implications for models of conversation and phonetic control.

Keywords: articulation, motor control, speech planning, timing prediction, turn-taking

Edited by:
Martine Grice,
University of Cologne, Germany

Reviewed by:
Francisco Torreira,
McGill University, Canada
Jelena Krivokapic,
University of Michigan, United States

****Correspondence:***
Peter A. Krause
peakraus@ucsc.edu

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *23 March 2021*
Accepted: *18 June 2021*
Published: *13 July 2021*

Citation:
Krause PA and Kawamoto AH (2021) Predicting One's Turn With Both Body and Mind: Anticipatory Speech Postures During Dyadic Conversation. Front. Psychol. 12:684248. doi: 10.3389/fpsyg.2021.684248

INTRODUCTION

Successful spoken communication requires navigating two overlapping sets of temporal constraints. On the one hand, there is what might be called phonological timing: how the flow of articulatory events gives rise to intelligible speech. Without proper phonological timing, the intended utterance "dab" might be distorted to "bad" (Browman and Goldstein, 1995). On the other hand, there is what might be called situational timing: how phonetic events are timed against the background grid of the environment, including others' speech. Situational timing is key to inter-speaker coordination. For example, inter-turn gaps at changes of floor are quite short, with mean gap time varying from 7 to 423 ms across several languages (Stivers et al., 2009). As we will outline below, most extant speech models assume that phonological and situational timing are governed by distinct cognitive mechanisms. We will argue that understanding inter-speaker coordination requires re-evaluating this assumption. Such coordination may arise when speakers apply situational timing mechanisms to aspects of the utterance traditionally viewed as the domain of phonological timing.

Traditional models assume that utterance initiation is controlled by an online decision mechanism sensitive to situational factors like a "go" signal (Sternberg et al., 1978), or, when adapted to the context of conversation, another speaker's communicative cues (e.g., Levinson and Torreira, 2015; Levinson, 2016). However, once initiated, an utterance's internal timing is assumed to follow a prefabricated motor plan. In Levelt et al.'s (1999) influential model, this plan is a programmatic gestural score produced by the phonetic encoding mechanism. In Articulatory Phonology with Task Dynamics (AP/TD), this plan comprises the parameterized constriction gestures (Saltzman and Munhall, 1989), the phase couplings between gestural planning oscillators (Saltzman and Byrd, 2000), and the π-gestures implementing prosodic adjustments at phrase boundaries (Byrd and Saltzman, 2003).

There is empirical support for the separability of speech planning and speech triggering, both in pure laboratory tasks and in conversation tasks. For example, delayed naming tasks (e.g., Sternberg et al., 1978) attempt to isolate the speech triggering process by informing the participant of what they will say ahead of time, and then providing a secondary "go" signal to cue speech onset. The assumption is that participants will withhold the articulatory response until the "go" signal. Contrariwise, speeded naming tasks (e.g., Meyer, 1990, 1991) attempt to isolate the planning process by asking participants to respond as quickly as possible after the content of the next utterance is revealed. The assumption is that participants will complete planning and then initiate articulation as quickly as possible afterward. In delayed naming, participants generally reserve acoustic onset until after the "go" signal, and produce shorter acoustic latencies compared to speeded naming tasks. Both phenomena fit with the claim that triggering has been isolated. Similarly, in conversational tasks, EEG evidence suggests that relevant speech planning begins long before a partner finishes their current utterance (Bögels et al., 2015). However, as mentioned above, inter-turn gaps are short; further, Levinson and Torreira (2015) indicate that acoustically overlapped speech comprises less than 5% of total conversation time. In aggregate, this suggests that speakers in conversation plan upcoming utterances and then acoustically withhold them while awaiting the next speech opportunity. Not only does the evidence converge to the conclusion that planning and triggering are separable, but it also implies possible parallels between delayed naming and conversational speech initiation. We will return to this point later.

The separability of speech planning and speech triggering does not, on its own, entail the strict encapsulation of phonological from situational timing. Instead, we propose that these ideas have been accidentally conflated, partly because of the classical "motor program" concept. Work in delayed naming has revealed evidence that speakers preferentially "chunk" their utterances during the final moments of preparation. Sternberg et al. (1978) found acoustic response latency following the "go" signal to be a linear function of the number of stress-bearing syllables in the utterance. This work was later replicated and extended by Wheeldon and Lahiri (1997). The finding has been offered as evidence that stress-bearing syllables are the "subprograms" of speech, terminology which certainly implies fixed movement timing. Arguably, however, this interpretation reflects a preexisting commitment to the computer metaphor, as much as it reflects the specific empirical evidence.

The notion of "soft" movement plans is already well-ensconced in the phonetics literature, in the form of AP/TD's articulatory gestures. In that theory, planned gestures do not uniquely determine the spatial trajectories of articulators. Each gesture corresponds to an articulatory synergy (e.g., Browman and Goldstein, 1989); if the motion of one articulator is impeded, other articulators in the synergy can move differently to compensate. This affords the flexibility to adapt to unexpected situational events (such as perturbation of jaw motion during a bilabial closure, e.g., Folkins and Abbs, 1975; Kelso et al., 1984). Specific kinematic trajectories are emergent from the intersection of the gesture with the (dynamic, evolving) embedding context. It seems possible, at least in principle, that a plan for phonological timing could similarly comprise constraints (e.g., on the ordering and/or permissible overlap of actions) rather than a rigid specification of the behavioral time course. [Note that this is admittedly not the case in AP/TD itself; timing in that theory is prescribed by gestures' stiffness parameters, combined with the stable phasing relationships of coupled planning oscillators. But other touchstones exist. See, for example, Jordan, 1986; Liu and Kawamoto, 2010; Tilsen, 2016, 2018. This narrative was recently reviewed in detail by Krause and Kawamoto (2020b)].

These issues are highly relevant to the triggering of speech in conversation. Conversational utterance timing is precise. This is true not only for canonical turns (as represented by extremely short inter-turn gaps), but also for backchannels, which tend to be acoustically initiated following similar syntactic and prosodic cues as floor transitions (e.g., Koiso et al., 1998; Ward and Tsukahara, 2000) and which have a proper timing that is both perceptible and non-random (Poppe et al., 2011). One possibility is that this precision is aided by mechanisms that predict opportunities for speech onset. Most of the relevant evidence comes from the turn-taking literature. De Ruiter et al. (2006) found that participants could predict the timing of turn ends from lexico-syntactic cues. Magyari and de Ruiter (2012) found that listeners partly predicted turn-end phrasings and suggested this prediction could be used as a proxy estimate of remaining turn length. Rühlemann and Gries (2020) gave evidence that speakers progressively slow speech rate over most of the turn, implying that listeners might use prosodic cues in turn-end prediction. However, contrary to the above, Bögels et al. (2015) reported that accurate turn-end detection required participants to hear turn-final intonational phrase boundaries.

Often overlooked is that the debated role of prediction in speech triggering is entangled with the issue of whether planning uniquely determines phonological timing. Before producing the earliest sounds of the utterance, speakers must first establish the initial constrictions in the vocal tract. Estimates and measurement practices vary, but Rastle et al. (2005) found the mean delay between articulatory and acoustic onset to range from 223 to 302 ms across syllable onset types. If speakers precisely time acoustic onset (e.g., by targeting no acoustic gap and no acoustic overlap at turn transitions, Sacks et al., 1974), they must work around this lead time. If this lead time is fixed by prior planning, a tricky problem arises. If a speaker waits to be certain a speech opportunity has arisen, they may initiate their utterance too late. The (un-compressible) lead time will then compound the delay preceding their acoustic response. If a speaker initiates their utterance from the predicted timing of a speech opportunity, then error in this prediction may lead them to start too early. The (un-expandable) lead time will then inexorably unfold to the point of an acoustic interruption. On this latter basis, Torreira et al. (2015) have argued that articulation is not initiated from predicted timing. Levinson and Torreira's (2015) and Levinson (2016) model of turn taking asserts that utterances are largely planned on a partner's turn but held in abeyance until the end of that turn is definitively detected.

To our knowledge, only one study has directly evaluated the late-initiation assumption using a conversational task. It therefore warrants specific consideration here. Torreira et al. (2015) analyzed breathing patterns of dyads during question-answer sequences. Specifically, they inspected the distribution of inbreath timings following the start of a question. This distribution was highly variable, but the mode fell 15 ms after question end. The authors reported this as evidence for late articulatory initiation. This study is an important first step in the area but has some critical limitations. One may question how well its restricted focus on question-answer sequences generalizes to both other kinds of turn exchanges and utterance types, such as backchannels, deliberately omitted from the turn-taking tradition. Further, we wonder whether the large variability in inbreath timings arose because a wider range of breathing strategies was in use than the study recognized. Finally, inbreath timing is not likely to index a fixed coordinative sequence for speech. For example, Mooshammer et al. (2019) found that acoustic response times were later for naming targets presented mid-inbreath, compared to ones presented mid-outbreath, suggesting speakers finished in-progress inhalations before initiating verbal responses. However, speakers also did not take new inbreaths, when presented with the target during exhalation.

Moreover, in naming research, articulatory kinematics often tell a different story from other measures. We earlier noted that conversational speech triggering invites comparison to the delayed naming paradigm. Classical findings in delayed naming, based on acoustic measures, appeared to indicate that speech was not initiated until the "go" signal. However, when delayed naming experiments have added articulatory measures, the results have suggested a different interpretation, one seemingly incompatible with the fixed-time-course narrative. Both Kawamoto et al. (2014) and Tilsen et al. (2016) presented participants with monosyllables to be read aloud upon "go" signal presentation, while measuring articulator positions using either video or structural MRI. The "go" signal followed stimulus presentation by a variable delay. Participants postured their vocal tracts to anticipate form-specific requirements of the utterance. They formed and maintained these postures during the unpredictable period separating stimulus onset from "go" signal, while nonetheless delaying the acoustic response until appropriate.

This suggests that speakers in conversation may have heretofore unrecognized degrees of freedom for coordinating acoustic onset timing. The silent interval during which initial constrictions are formed may in fact be compressible or expandable, even after movement has started. This leads us to the following general hypothesis motivating this study: We propose that speakers in conversation can initiate the earliest articulatory movements from predicted timing, at least under some conditions. We further suggest that they compensate for prediction error online, by slowing or speeding the articulatory time course as it unfolds. A second general hypothesis follows by implication from the first: The earlier that pre-acoustic articulation is initiated (with respect to the eventual acoustic onset), the lower the peak velocity of that motion.

This mechanism may not be equally utilized (or available) across all contexts. Laboratory work examining articulatory strategies for speech suggests they respond to many factors. Consider studies examining incrementality (i.e., speakers' choices to produce speech by small chunks as they are planned, versus waiting and then producing large chunks all at once). Propensity to incremental speech can reflect individual differences (Kawamoto et al., 2014), subtleties of task (such as whether instructions were to begin speaking as soon as possible or to speak as briefly as possible, Holbrook et al., 2019), and even language spoken (Swets et al., 2021). Speaker's use of predictive initiation with adjustment may therefore vary with several factors. These factors might include how early (with respect to the targeted moment of acoustic onset) the initial words of the utterance are planned, how much of the utterance can be held in working memory, and/or the speaker's willingness to produce those early words incrementally. Overall, then, it may be that particularly short, stereotyped utterances are the most likely to be prepared in this manner, when disregarding other contextual factors. Knudsen et al. (2020) have recently suggested that speakers use such "forgotten little words" to mitigate conversational costs. The present study makes use of this observation.

We should emphasize that this tendency for articulatory preparation to arise more in some contexts than others is a core theme of the present study. It is not intended *per se* as a refutation of Levinson and Torreira's (2015) proposed rule for articulatory initiation. It is more so intended as a fundamental reframing of the point, away from assertions of hard rules or typical cases, and toward consideration of the range of strategies available and their domains of application.

In the present study, we sought evidence for anticipatory speech postures in natural conversation. The study utilized data from the Cardiff Conversation Database (Aubrey et al., 2013), an audiovisual database of dyads engaging in unscripted conversations. Motion-tracked video of speakers' faces was used to assess changes in lip area over time. We looked for contrasts between utterances beginning with smaller lip area due to closure and/or rounding (the *labially constrained* condition) versus other oral configurations (the *labially unconstrained* condition). Further, we examined how these contrasts were moderated by the number of words in the utterance. We specifically predicted that lip areas for labially constrained vs. unconstrained utterances would be discriminable earlier, relative to acoustic onset, when utterances were very short (1–2 words long).

MATERIALS AND METHODS

This study used data drawn from the Cardiff Conversation Database (Aubrey et al., 2013). This database is available by request from https://ccdb.cs.cf.ac.uk/signup.html. The authors asked several dyads to engage in 5-min unscripted conversations while their faces were video-recorded (at 30 frames per second) and their speech audio-recorded. While the authors suggested possible conversation topics, these topics were not actively enforced. Full data coding and transcription has been completed for eight dyads; these are the dyads analyzed in the current study.

Participants

The eight analyzed dyads included six speakers. The following demographic information was provided by P. Rosin (personal communication, February 4, 2021; April 29, 2021). All speakers were Caucasian males. Ages ranged from 27 to 47 years ($M = 36.33$). Two speakers spoke English with a Welsh accent, one with a Scottish accent, one with a German accent, and two with an English accent (one having lived throughout Southern England, and one having grown up in Essex).

Each speaker participated in at least two dyads (equating to at least 10 min of recorded video footage). Details about the number of dyads each speaker participated in, plus additional information about their utterance distributions, appears in **Table 1**.

TABLE 1 | Speaker-specific information.

Speaker	Number of dyads	Mins. of recordings	Constrained utterances	Unconstrained utterances
P1	2	10	12	29
P2	3	15	12	50
P3	3	15	50	59
P4	2	10	19	23
P5	2	10	14	36
P6	3	15	19	29

Number of utterances refers only to retained data.

Elan Annotations

The database includes detailed, time-aligned behavioral annotations performed in the Elan software (Wittenburg et al., 2006) for all eight dyads. These annotations include temporal markings for the acoustic onsets and offsets of all verbal utterances, as well as full transcriptions of utterance content. Acoustic onsets for utterances were specifically marked at the first notable swell in audio intensity leading into an identifiable speech sound (P. Rosin, personal communication, April 29, 2021).

OpenFace Tracking Outputs

The database authors processed each facial video using OpenFace 2.0 (Baltrušaitis et al., 2018). The database contains the resulting output files. OpenFace detects the most prominent face in a video and tracks its pose in six degrees of freedom relative to camera origin, as well as tracking the three-dimensional positions of 128 key points on the face. We have previously described how to extract linguistically useful information about oral configuration from OpenFace output (Krause et al., 2020). OpenFace's positional estimates hew closest to the true values when OpenFace is provided with the camera's intrinsic lens parameters. OpenFace was not calibrated in this way before processing the facial videos in the database (P. Rosin, personal communication, February 3, 2021). However, since all statistical inference will be based on within-speaker comparisons, the lack of a pure correspondence to real-world units is incidental.

For readers unfamiliar with the OpenFace system, we provide **Figure 1** as illustration. The panels of **Figure 1** depict just those parameters of OpenFace that track the outer and inner lips (i.e., parameters 48–67). The dots colored in blue depict those parameters used in the lip-area computations described below. The panels specifically depict how OpenFace differentially tracks the lips when they are in different configurations (as in the *labially constrained* versus *labially unconstrained* utterance types, described in more detail below). To produce these plots, the first author spoke two example utterances on digital video that was later processed by OpenFace: the word "oodles" (a constrained utterance, left) and the word "apple" (an unconstrained utterance, right). The depicted tracking is from the moment of acoustic onset for both utterances. To facilitate comparison, both plots have been centered with respect to parameter 51, which marks the notch at the top of the outer lips.

Data Preparation

In total, the Python script described below identified 509 distinct utterances in the annotated Elan data.

Utterance Types and Inter-Utterance Gaps

A specialized Python script processed the Elan output for each speaker. For each annotated utterance, the script determined the most recently initiated prior utterance (even if this prior utterance was not yet concluded). If that recently initiated utterance was by the same speaker, the present utterance was labeled as a restart and marked to be dropped from the final dataset (our interest being in inter-speaker coordination). Utterances were also marked as restarts if the most recent

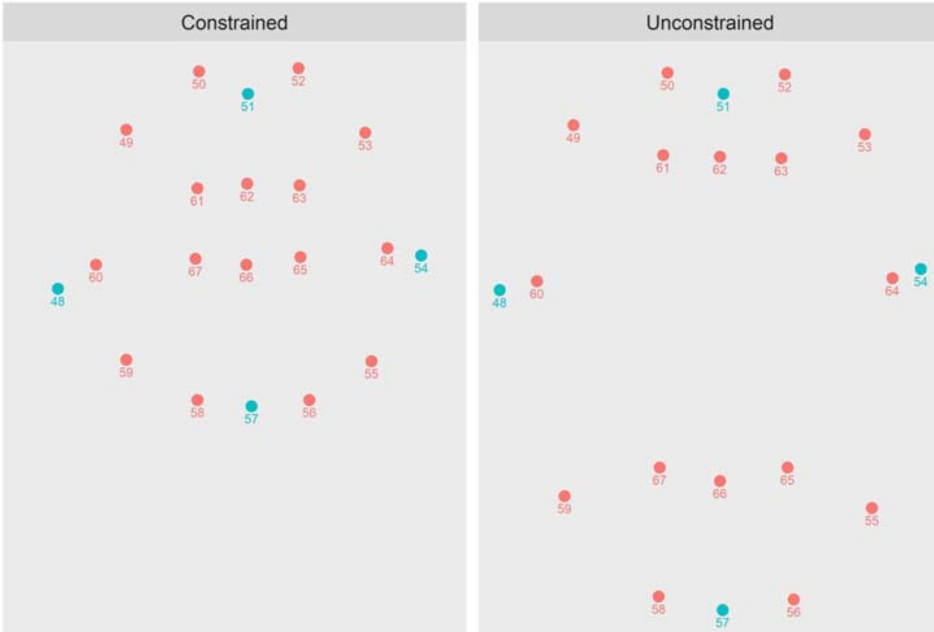


FIGURE 1 | Plots of OpenFace parameters 48–67, which track the outer and inner lips, as arrayed on two frames of real facial video, each capturing the moment of acoustic onset for a different word. **(Left)** The first author beginning the word "oodles," a constrained word. **(Right)** The first author beginning the word "apple," an unconstrained word.

initiation was from the other speaker but had occurred within 500 ms of a prior initiation from the current speaker. Otherwise, utterances were marked as responses, and an inter-utterance gap time was computed by subtracting the onset time of the current utterance from the offset time of the most recently initiated prior utterance.

Similarly to Heldner and Edlund (2010), responses were further labeled as gaps (if the current utterance followed by previous one by a positive gap time), between-overlaps (if the current utterance started before the prior utterance completed, but finished afterward), and within-overlaps (if the current utterance lay completely within the acoustic boundaries of the prior utterance).

Lip Area Trajectories

By referencing the OpenFace outputs, the Python script computed lip area at each of 90 frames (3,000 ms) preceding the acoustic onset of each annotated turn. The script used the estimated *x*- and *y*-coordinates of four key points: The left-hand and right-hand corners of the lips (OpenFace parameters 48 and 54, respectively), and the external points at the center-top and center-bottom of the lips (OpenFace parameters 51 and 57, respectively). Area was computed as described by Liu et al. (accepted).

The strategy creates a perimeter of line segments running clockwise around the points, each conceived as the hypotenuse of a right triangle. Lip area is the sum of the areas of all four right triangles, plus the area of a residual central rectangle. Our formulas assume that one labels the left corner as (X_1,Y_1) and then increments the X- and Y-values while moving clockwise around the set. The following formula produces the areas of each right triangle *i* (which has one corner at (X_i,Y_i):

$$A_i = \frac{|X_i - X_{i+1}| \times |Y_i - Y_{i+1}|}{2}$$

The following formula produces the area of the residual central rectangle *j*:

$$A_j = |X_1 - X_3| \times |Y_2 - Y_4|$$

Identification of Key Predictor Values

The Python script also counted the number of words in the transcription for each turn and extracted the first transcribed word for reference.

We manually coded each turn as either *labially constrained* (in which case we would expect a comparatively small lip area at acoustic onset) or as *labially unconstrained*. We made this assessment based on the first syllable of the first annotated word. Specifically, we considered both the initial consonant (if applicable) and the nuclear vowel. Turns with initial consonants were classified as labially constrained if the consonant was bilabial, labiodental, or rounded, i.e., a member of the set /b, f, m, p, ɹ, v, w/. Regardless of initial consonant, turns were classified as labially constrained if their first nuclear vowel was rounded, i.e., a member of the set /ɔ, o, ʊ, u/. Turns not fitting either of these criteria were classified as labially unconstrained.

RESULTS

Utterance Types and Inter-Utterance Gap Times

Restarts and the initial utterances of a conversation were expunged from the data, leaving 352 utterances in the set. Gap utterances comprised 47% of retained data, between-overlaps 22%, and within-overlaps 30%.

Because the nominally intended start times for gap utterances and between-overlaps are relatively clear (i.e., the acoustic offset of the prior utterance) their coordination can be further characterized by considering their inter-utterance gap times (i.e., their floor transfer offsets, Levinson and Torreira, 2015). Figure 2 depicts a histogram of the floor transfer offsets in the final data. The mean floor transfer offset was 258.37 ms (SD = 914.96).

The coordination of within-overlaps cannot be so easily characterized, since these are typically backchannels or other short utterances whose timing targets are a syntactic or prosodic closure within an utterance that is acoustically continued. However, spot-checking of the transcript suggested that within-overlaps were coordinated reasonably.

For example, the most extreme within-overlap marked in the dataset arises in a conversation between P3 and P4. At this point, P4 has just been asked his favorite biography. Comparison of the Elan annotations with the acoustic waveforms reveals this rough pattern of coordination (bolded text indicates the specific point of overlap):

> **P4**: The, the one on Copeland one Aaron Copeland one was a good one, **uhh**, a large volume. . .
> **P3**: **Okay**.

P4's turn continues for some time after without an acoustic break, resulting in the inter-utterance gap time for P3's "Okay" being computed as −99 s, despite it having been realized shortly after the resolution of P4's statement "Aaron Copeland was a good one."

All within-overlaps were retained in the final dataset so as not to compromise statistical power for detecting anticipatory postures.

Lip Area Trajectories

Statistical analysis of lip area was carried out in R version 4.1.0 ("Camp Pontanezen"), using the "lme4," "lmerTest," "interactions," "effects," and "bootMer" packages. We fit linear mixed-effects models to the data at 15-frame (500-ms) intervals, starting at 3,000 ms prior to acoustic onset, and ending at acoustic onset. This resulted in 7 fit models. The dependent variable in each case was lip area, with predictor variables of interest being utterance word count (log10-transformed to correct for extreme positive skewness), labial constraint (constrained vs. unconstrained), and their two-way interaction. We used Speaker ID and the first word of the utterance as clustering variables. We determined the random effects structure for each model via a backward selection procedure that started with the maximal structure (Barr et al., 2013) and then simplified until lme4 returned no warnings about convergence failure or model singularity. To facilitate hypothesis testing, we estimated denominator degrees of freedom via Satterthwaite's method (see Luke, 2017 for a justification).

Our labial constraint predictor was dummy-coded "0" for constrained utterances and "1" for unconstrained utterances. Therefore, positive slope values would indicate larger lip area for unconstrained than constrained utterances, signaling the likely presence of speech postures. This gives the slopes of the word count × labial constraint interaction terms a straightforward interpretation. Negative values for these slopes would indicate

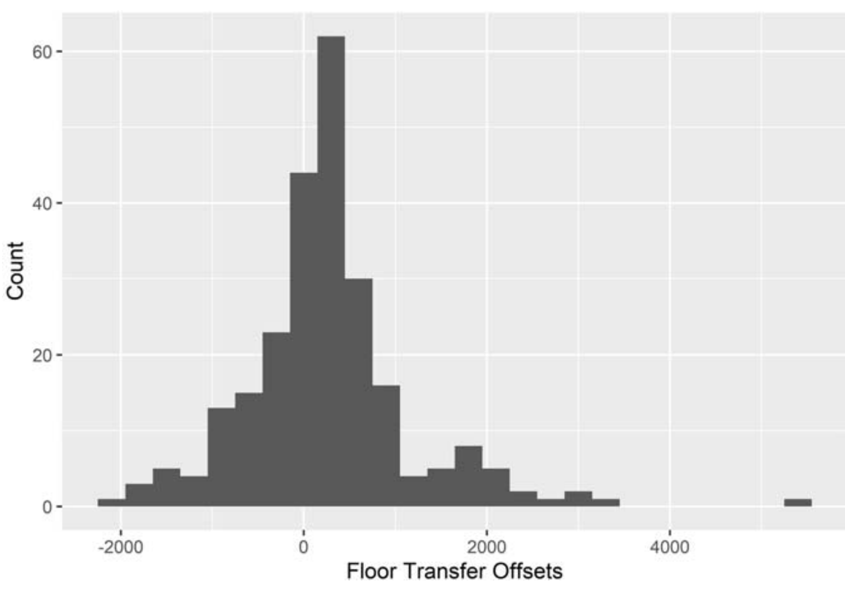


FIGURE 2 | A histogram depicting the floor transfer offsets (i.e., the inter-utterance gaps for gap utterances and between-overlaps) in the final dataset.

decreasing probability and/or magnitude of speech postures with increasing word count, consistent with our hypothesis. When the interaction term was statistically reliable, we followed up by computing Johnson-Neyman intervals revealing the specific ranges of word count over which the labial constraint predictor differed from 0.

Table 2 reports the statistical results in detail, including the random effects structures of the final fitted models. Where we report the results of Johnson-Neyman intervals, we give ranges of actual word counts (as opposed to their log-10 transformations), and report only those ranges where the slope of labial constraint was both positive and within the observed span of the data.

The pattern of results in Table 2 is consistent with our prediction that anticipatory postures would emerge earlier (resulting in earlier distinctions between labially constrained vs. unconstrained turns) for shorter utterances. The word count × labial constraint interaction was statistically reliable from −3,000 ms through −1,000 ms, with Johnson-Neyman intervals suggesting that postures were most likely for utterances of three words or fewer at the earliest time points. At the end of the time course this interaction (unsurprisingly) disappears, leaving only a reliable main effect of labial constraint at 0 ms.

Figure 3 presents a more visual illustration of these effects. For each of the seven models it plots the predicted lip area values (with bootstrapped 95% confidence intervals) for labially constrained and unconstrained utterances of two and eight words in length. (These values were chosen because two words is decidedly within the span at which the earliest anticipatory postures arose, while eight words is decidedly outside it, without being substantially larger).

Maximum Lip Movement Speed

In the Introduction, we suggested that speakers might increase articulatory movement speed to partly offset delays from planning difficulty, meaning faster movement speeds would be expected for utterances of more words.

We analyzed lip movement speeds for all utterances. We did this by computing the change in area at each of the final 15 steps of the analysis window (when lips should be approaching final configuration targets) and multiplying each discrete change by 30 so that the units were given as mm^2/s. We determined the maximum lip movement speed by selecting the resulting value with the largest absolute magnitude. We then fit a linear mixed-effects model as described above, with this absolute magnitude as the dependent variable. Log10-transformed word count, labial constraint, and their two-way interaction were the predictors. The final fitted model included random intercepts for Speaker ID and first word, and a random slope for labial constraint at the Speaker ID level. The only reliable effect was a main effect of word count ($\beta = 455.22$, SE = 185.15), $t(64.73) = 2.50$, $p = 0.02$. This effect suggests that, as predicted, lip movement speeds increase as word count increases. Figure 4 depicts the regression line for word count (with 95% confidence ribbon).

Content of One-Word Utterances

The results suggest that the earliest anticipatory speech postures arose for utterances of three or fewer words. Although

TABLE 2 | Reports of linear mixed models fit to lip area.

Time point (ms)	Model details	
−3,000	Random effects	(1 | speaker) + (1 | first word)
	Constraint × log10(word count)	β = −64.48, SE = 24.99, *t*(109.35) = −2.58*
	Main effect of constraint	β = 68.73, SE = 25.96, *t*(20.81) = 2.65*
	Johnson-Neyman interval for probable speech postures	[0, 3.09]
−2,500	Random effects	(log10(word count) | speaker)
	Constraint × log10(word count)	β = −61.86, SE = 24.14, *t*(291.77) = −2.56*
	Main effect of constraint	β = 66.33, SE = 23.62, *t*(251.79) = 2.81**
	Johnson-Neyman interval for probable speech postures	[0, 3.72]
−2,000	Random effects	(log10(word count) | speaker) + (1 | first word)
	Constraint × log10(word count)	β = −58.07, *SE* = 25.09, *t*(113.91) = −2.32*
	Main effect of constraint	β = 71.78, SE = 25.83, *t*(20.80) = 2.78*
	Johnson-Neyman interval for probable speech postures	[0, 4.47]
−1,500	Random effects	(log10(word count) | speaker)
	Constraint × log10(word count)	β = −90.33, *SE* = 25.56, *t*(299.03) = −3.53***
	Main effect of constraint	β = 105.23, SE = 24.96, *t*(253.49) = 4.22***
	Johnson-Neyman interval for probable speech postures	[0, 6.45]
−1,000	Random effects	(log10(word count) | speaker) + (1 | first word)
	Constraint × log10(word count)	β = −70.37, SE = 27.22, *t*(191.49) = −2.59*
	Main effect of constraint	β = 82.24, SE = 29.40, *t*(51.24) = 2.80**
	Johnson-Neyman interval for probable speech postures	[0, 3.89]
−500	Random effects	(log10(word count) | speaker) + (1 | first word)
	Constraint × log10(word count)	β = −42.84, *SE* = 28.79, *t*(277.68) = −1.49
	Main effect of constraint	β = 62.23, *SE* = 32.81, *t*(97.14) = 1.90†
	Johnson-Neyman interval for probable speech postures	N/A
0	Random effects	(log10(word count) | speaker) + (1 | first word)
	Constraint × log10(word count)	β = −35.16, SE = 24.35, *t*(267.60) = −1.44
	Main effect of constraint	β = 97.83, SE = 27.63, *t*(89.50) = 3.54***
	Johnson-Neyman interval for probable speech postures	N/A

Random effects were first extracted as numeric vectors. Double-bar notation indicates uncorrelated random effects. Although Word Count was log10-transformed, Johnson-Neyman intervals are given here as number of words, for ease of interpretation.
†$p < 0.1$.
*$p < 0.05$.
**$p < 0.01$.
***$p < 0.001$.

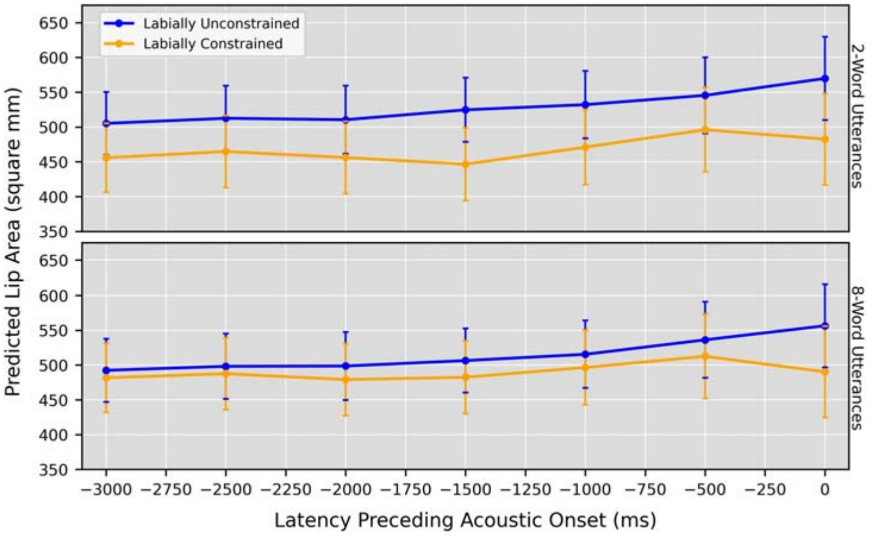


FIGURE 3 | Predicted lip area values produced by the linear mixed models, when setting the word count predictor to 2 words and 8 words. Models were fit to junctures at 15-frame (500-ms) intervals, starting at 90 frames (3,000 ms) preceding acoustic onset. Error bars: Bootstrapped 95% CI.

our hypothesis proceeded from the assumption that shorter utterances are easier to plan, there are multiple confounded reasons that this might be. For example, the smaller number of words might *per se* lower planning complexity, but the communicative content itself might also be simpler or higher in frequency. For this reason, some readers may wish to get a sense of the content of these shorter utterances. Table 3 presents a frequency-ordered list of every type of one-word utterance in the final dataset (one-word utterances representing 62% of all utterances of three or fewer words).

DISCUSSION

The main finding of the present study is that speakers awaiting their turn in natural conversation sometimes produce anticipatory oral postures well in advance of starting their acoustic utterance. In this study, statistical detection of these postures was facilitated when utterances were especially short (1–3 words). This may suggest that the articulatory planning and/or control of these postures is made easier when the overall utterance is low in complexity. As such, although this study extends the findings of a previous study of utterance timing in conversation (Torreira et al., 2015), it may not directly contradict that prior study's finding that utterances were

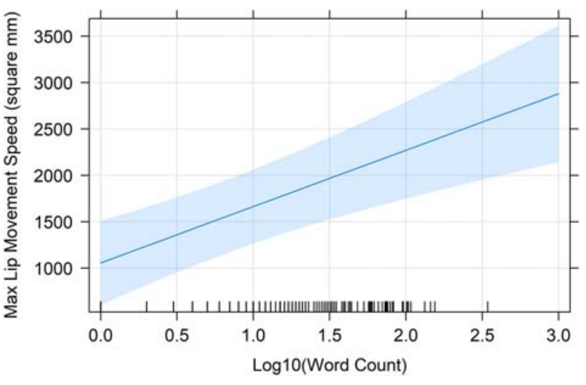


FIGURE 4 | The regression line (with 95% confidence band) for the change of maximum lip movement speed with log10-transformed word count, as predicted by the mixed-effects model.

TABLE 3 | Content and frequency of 1-word utterances.

Word	Count	Labial constraint
Yeah	37	Unconstrained
Alright	13	Constrained
Right	9	Constrained
No	8	Constrained
Yep	5	Unconstrained
Yes	4	Unconstrained
Really	3	Constrained
So	2	Constrained
Excellent	1	Unconstrained
Mmhmm	1	Constrained
Next	1	Unconstrained
Nice	1	Unconstrained
Oh	1	Constrained
Ok	1	Constrained
Thanks	1	Unconstrained
that's	1	Unconstrained
Very	1	Constrained
Well	1	Constrained
What	1	Constrained
Which	1	Constrained

initiated late, since the utterances in the earlier study were of higher complexity.

Prior studies have reported similar effects for isolated word production tasks (Kawamoto et al., 2014; Drake and Corley, 2015; Tilsen et al., 2016; Krause and Kawamoto, 2019, 2020a; Tilsen, 2020). We believe, however, that this is the first report of such effects in an ecological conversation task. In addition, while Tilsen (2020) recently found that articulatory postures could arise as long as 500 ms before canonical articulatory onset, the current study is remarkable in finding that very short utterances could see postures arise as long as 3,000 ms before acoustic onset.

Limitations

The non-experimental nature of our approach leaves the causal underpinnings of these postures underdetermined. We assumed that the short utterances would be faster to plan, leading to more cases of delay between phonological availability and acoustic response opportunity. However, the observed pattern may instead arise from some correlate of utterance length.

The analysis reflects 352 utterances produced by six male Caucasian speakers of English; generality may be limited. However, because anticipatory posturing has not previously been reported in an ecological task, the results are intrinsically important. They provide a valuable existence proof for the narrative of flexible articulatory control outlined in the Introduction, whereby speakers can independently manipulate utterance initiation and the time course of articulation. Further, the results suggest fertile possibilities for experimental follow-up.

Theoretical Implications

We propose that, under some conditions, speakers can initiate articulation from ongoing prediction of the next speech opportunity, while using online control to finesse the moment of acoustic onset. Outstanding questions at this juncture include why postures are more likely under some conditions than others, and whether the postures are strategically functional.

Why Does the Emergence of Postures Vary?

As noted earlier, Torreira et al. (2015) concluded that articulation was initiated just after a partner's utterance ended. Although this may reflect the specific dependent measure used (inbreaths, as indexed via inductive plethysmography), it may also be that the specific utterances analyzed did not facilitate early articulation. In the present study, anticipatory posturing only verifiably arose for utterances of 1–3 words in length. Although we predicted that postures should be variable, in accordance with the strategic flexibility observed in past articulatory studies, it remains uncertain exactly how this variability is structured.

We presumed that number of words in the utterance indexed planning complexity. If we are correct, then the phonology of more complex utterances might become available later, relative to the targeted moment of acoustic onset, leaving a shorter span inside which anticipatory postures could emerge and reduce movement speeds. Admittedly, however, "planning complexity" is ambiguous here. It could be that having fewer words to plan places less strain on the phonological system. It could also be that utterances that perform certain communicative functions or comprise certain high-frequency phrases are planned more easily, and that these utterances incidentally tend to be shorter. Further, the emergence of postures for certain kinds of utterances might reflect, not simpler planning, but some difference in either conventional timing or the need to visually signal intent (see below).

It might also be that the number of words in an utterance somehow moderates the readiness with which phonological plans are conferred into action. In their classic delayed-naming study, Sternberg et al. (1978) found that longer prepared utterances had longer acoustic latencies following the go signal. Although the present study is one of several to show that articulatory motion can be partly de-coupled from acoustic onset, perhaps this de-coupling becomes more difficult to manage as the upcoming utterance grows in length.

Are Postures Strategically Functional?

Possibility 1: intention leaks into articulation

One possibility is that the emergence of postures is not strategic but is instead a passive, incidental consequence of how the planning and motor systems are coupled. We base this possibility on Tilsen's (2019) explanation of the anticipatory postures observed in laboratory tasks. In this account, the postures arise when partly activated but unselected speech gestures cascade their influences into current articulatory targets, across a partly permeable threshold.

This proposal can account for anticipatory speech postures in which the phonologically relevant constrictions are only partly formed (which Tilsen et al., 2016, found to be common). However, not all anticipatory postures are partial. In video recordings of their delayed naming task, Kawamoto et al. (2014) observed speakers who both closed their lips and accumulated intra-oral pressure when preparing /p/-initial utterances. (A comparable observation is not viable for the present study, owing to the small number of spontaneously produced utterances happening to start with bilabial plosives).

Possibility 2: reduction of movement costs

Starting early (i.e., lengthening posture duration) may minimize movement costs, as would be predicted by optimal control theories of speech behavior (e.g., Nelson et al., 1984). We observed that maximum lip movement speed was relatively lower over the ranges of word counts at which anticipatory postures arose. For a frictionless system, peak velocity is an estimate of the integral of force applied per unit mass with respect to time (Nelson, 1983). Such force integrals are one candidate approach for quantifying the energy and/or effort costs of motor function (see, e.g., Turk and Shattuck-Hufnagel, 2020).

Possibility 3: social signaling

Testing of Levinson and Torreira's (2015) model has largely emphasized listeners' abilities to predict and prepare for floor yielding. However, the anticipatory postures observed in this study may reflect the listener's agency in effecting speech opportunities. It is possible that listeners deliberately use them to indicate readiness to speak. Irrespective of whether the behavior is deliberate, it may be perceived as a social signal by the current speaker. It might also assist the current speaker in managing attention, such that the incoming utterance is a less surprising event.

This proposal has some precedence. Kendon (1967) reported that as a speaker approached a possible floor transition, they tended to shift their gaze to the listener; this gaze continued for a while all the way through the transition. Listeners sometimes visually signaled a readiness to take the floor prior to the gaze shift. Similarly, Bavelas et al. (2002) found that brief windows of mutual gaze between speaker and listener often preceded backchannels. Kendrick and Holler (2017) presented evidence that listeners' gaze patterns could influence the end of speakers' turns. Averted gazes signaled that listeners were planning dispreferred responses; in some cases these gazes yielded last-minute repairs by the speaker intended to eliminate the dispreferred response.

Concluding Remarks

The current study presented preliminary evidence that speech planning and articulation may flexibly overlap in natural conversation. This suggests that articulation may at times be initiated based on the predicted timing of speech opportunities, without obligating an acoustic interruption. Future work will be necessary to determine exactly what mechanisms speakers use to regulate the time course of articulation, how much deliberate strategy is involved, and whether this phenomenon carries social/pragmatic implications.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://doi.org/10.17605/OSF.IO/JKR96.

AUTHOR CONTRIBUTIONS

PK carried out the data encoding and analysis and wrote the entire manuscript. Both authors jointly conceived and designed this study.

FUNDING

Publication fees for the present article were funded in part by the Department of Psychology at California State University Channel Islands.

REFERENCES

Aubrey, A. J., Marshall, D., Rosin, P. L., and Vandeventer, J. (2013). "Cardiff Conversation Database (CCDb): a database of natural dyadic conversations," in *Proceedings of the V & L Net Workshop on Language for Vision, 2013*, Portland, OR.

Baltrušaitis, T., Zadeh, A., Lim, Y. C., and Morency, L. (2018). "OpenFace 2.0: facial Behavior Analysis Toolkit," in *Proceedings of the 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)(FG)*, Xi'an, 59–66. doi: 10.1109/FG.2018.00019

Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. *J. Mem. Lang.* 68, 255–278.

Bavelas, J., Coates, L., and Johnson, T. (2002). Listener responses as a collaborative process: the role of gaze. *J. Commun.* 52, 566–580. doi: 10.1111/j.1460-2466.2002.tb02562.x

Bögels, S., Magyari, L., and Levinson, S. C. (2015). Neural signatures of response planning occur midway through an incoming question in conversation. *Sci. Rep.* 5:12881. doi: 10.1038/srep12881

Browman, C. P., and Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology* 6, 201–251. doi: 10.1017/S0952675700001019

Browman, C. P., and Goldstein, L. (1995). "Dynamics and articulatory phonology," in *Mind as Motion: Explorations in the Dynamics of Cognition*, eds R. F. Port and T. van Gelder (Cambridge, MA: The MIT Press), 175–193.

Byrd, D., and Saltzman, E. (2003). The elastic phrase: modeling the dynamics of boundary-adjacent lengthening. *J. Phon.* 31, 149–180. doi: 10.1016/S0095-4470(02)00085-2

De Ruiter, J. P., Mitterer, H., and Enfield, N. J. (2006). Projecting the end of a speaker's turn: a cognitive cornerstone of conversation. *Language* 82, 515–535. doi: 10.1353/lan.2006.0130

Drake, E., and Corley, M. (2015). Articulatory imaging implicates prediction during spoken language comprehension. *Mem. Cogn.* 43, 1136–1147. doi: 10.3758/s13421-015-0530-6

Folkins, J. W., and Abbs, J. H. (1975). Lip and jaw motor control during speech: responses to resistive loading of the jaw. *J. Speech Hear. Res.* 18, 207–220. doi: 10.1044/jshr.1801.207

Heldner, M., and Edlund, J. (2010). Pauses, gaps and overlaps in conversations. *J. Phon.* 38, 555–568. doi: 10.1016/j.wocn.2010.08.002

Holbrook, B. B., Kawamoto, A. H., and Liu, Q. (2019). Task demands and segment priming effects in the naming task. *J. Exp. Psychol. Learn. Mem. Cogn.* 45, 807–821. doi: 10.1037/xlm0000631

Jordan, M. I. (1986). *Serial Order: A Parallel Distributed Processing Approach (Tech. Rep. No. 8604)*. San Diego, CA: University of California, Institute for Cognitive Science.

Kawamoto, A. H., Liu, Q., Lee, R. J., and Grebe, P. R. (2014). The segment as the minimal planning unit in speech production: evidence from absolute response latencies. *Q. J. Exp. Psychol.* 67, 2340–2359. doi: 10.1080/17470218.2014.927892

Kelso, J. S., Tuller, B., Vatikiotis-Bateson, E., and Fowler, C. A. (1984). Functionally specific articulatory cooperation following jaw perturbations during speech: evidence for coordinative structures. *J. Exp. Psychol. Hum. Percept. Perform.* 10, 812–832. doi: 10.1037//0096-1523.10 6.812

Kendon, A. (1967). Some functions of gaze-direction in social interaction. *Acta Psychol.* 26, 22–63. doi: 10.1016/0001-6918(67)90005-4

Kendrick, K., and Holler, J. (2017). Gaze direction signals response preference in conversation. *Res. Lang. Soc. Interact.* 50, 12–32. doi: 10.1080/08351813.2017.1262120

Knudsen, B., Creemers, A., and Meyer, A. (2020). Forgotten little words: how backchannels and particles may facilitate speech planning in conversation? *Front. Psychol.* 11:593671. doi: 10.3389/fpsyg.2020.593671

Koiso, H., Yasuo, H., Tutiya, S., Ichikawa, A., and Den, Y. (1998). An analysis of turn-taking and backchannels based on prosodic and syntactic features in Japanese map task dialogs. *Lang. Speech* 41, 295–321. doi: 10.1177/002383099804100404

Krause, P. A., and Kawamoto, A. H. (2019). Anticipatory mechanisms influence articulation in the form preparation task. *J. Exp. Psychol. Hum. Percept. Perform.* 45, 319–335. doi: 10.1037/xhp0000610

Krause, P. A., and Kawamoto, A. H. (2020a). Nuclear vowel priming and anticipatory oral postures: evidence for parallel phonological planning? *Lang. Cogn. Neurosci.* 35, 106–123.

Krause, P. A., and Kawamoto, A. H. (2020b). On the timing and coordination of articulatory movements: historical perspectives and current theoretical challenges. *Lang. Linguist. Compass* 14:e12373. doi: 10.1111/lnc3.12373

Krause, P. A., Kay, C. A., and Kawamoto, A. H. (2020). Automatic motion tracking of lips using digital video and OpenFace 2.0. *Lab. Phonol. J. Assoc. Lab. Phonol.* 11, 9, 1–16. doi: 10.5334/labphon.232

Levelt, W. J. M., Roelofs, A., and Meyer, A. S. (1999). A theory of lexical access in speech production. *Behav. Brain Sci.* 22, 1–38. doi: 10.1017/S0140525X99001776

Levinson, S. C. (2016). Turn-taking in human communication–origins and implications for language processing. *Trends Cogn. Sci.* 20, 6–14. doi: 10.1016/j.tics.2015.10.010

Levinson, S. C., and Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Front. Psychol.* 6:731. doi: 10.3389/fpsyg.2015.00731

Liu, Q., Holbrook, B. H., Kawamoto, A. H., and Krause, P. A. (accepted). Verbal reaction times based on tracking lip movement. *Ment. Lex.*

Liu, Q., and Kawamoto, A. H. (2010). "Simulating the elimination and enhancement of the plosivity effect in reading aloud," in *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*, eds S. Ohlsson and R. Catrambone (Austin, TX: Cognitive Science Society), 2284–2289.

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behav. Res. Methods* 49, 1494–1502. doi: 10.3758/s13428-016-0809-y

Magyari, L., and de Ruiter, J. P. (2012). Prediction of turn-ends based on anticipation of upcoming words. *Front. Psychol.* 3:376. doi: 10.3389/fpsyg.2012.00376

Meyer, A. S. (1990). The time course of phonological encoding in language production: the encoding of successive syllables of a word. *J. Mem. Lang.* 29, 525–545. doi: 10.1016/0749-596X(90)90050-A

Meyer, A. S. (1991). The time course of phonological encoding in language production: phonological encoding inside a syllable. *J. Mem. Lang.* 30, 69–89. doi: 10.1016/0749-596X(91)90011-8

Mooshammer, C., Rasskazova, O., Zöllner, A., and Fuchs, S. (2019). "Effect of breathing on reaction time in a simple naming experiment: Evidence from a pilot experiment," in *Paper Presented at the International Seminar on the Foundations of Speech*, Sønderborg.

Nelson, W. L. (1983). Physical principles of economies of skilled movements. *Biol. Cybern.* 46, 135–147. doi: 10.1007/bf00339982

Nelson, W. L., Perkell, J. S., and Westbury, J. R. (1984). Mandible movements during increasingly rapid articulations of single syllables: preliminary observations. *J. Acoust. Soc. Am.* 75, 945–951. doi: 10.1121/1.390559

Poppe, R., Truong, K., and Heylen, D. (2011). "Backchannels: Quantity, type, and timing matters," in *Proceedings of the Intelligent Virtual Agents–11th International Conference, IVA 2011*, Reykjavik.

Rastle, K., Croot, K. P., Harrington, J. M., and Coltheart, M. (2005). Characterizing the motor execution stage of speech production: consonantal effects on delayed naming latency and onset duration. *J. Exp. Psychol. Hum. Percept. Perform.* 31, 1083–1095. doi: 10.1037/0096-1523.31.5.1083

Rühlemann, C., and Gries, S. T. (2020). Speakers advance-project turn completion by slowing down: a multifactorial corpus analysis. *J. Phon.* 80:100976. doi: 10.1016/j.wocn.2020.100976

Sacks, H., Schegloff, E. A., and Jefferson, G. (1974). A simple systematic for the organisation of turn taking in conversation. *Language* 50, 696–735. doi: 10.2307/412243

Saltzman, E., and Byrd, D. (2000). Task-dynamics of gestural timing: phase windows and multifrequency rhythms. *Hum. Mov. Sci.* 19, 499–526. doi: 10.1016/S0167-9457(00)00030-0

Saltzman, E. L., and Munhall, K. (1989). A dynamical approach to gestural patterning in speech production. *Ecol. Psychol.* 1, 333–382. doi: 10.1207/s15326969eco0104_2

Sternberg, S., Monsell, S., Knoll, R. L., and Wright, C. E. (1978). "The latency and duration of rapid movement: Comparisons of speech and typewriting," in *Information Processing in Motor Control and Learning*, ed. G. E. Stelmach (New York, NY: Academic Press), 117–152.

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., et al. (2009). Universals and cultural variation in turn-taking in conversation. *Proc. Natl. Acad. Sci.* 106, 10587–10592. doi: 10.1073/pnas.0903616106

Swets, B., Fuchs, S., Krivokapić, J., and Petrone, C. (2021). A cross-linguistic study of individual differences in speech planning. *Front. Psychol.* 12:655516. doi: 10.3389/fpsyg.2021.655516

Tilsen, S. (2016). Selection and coordination: the articulatory basis for the emergence of phonological structure. *J. Phon.* 55, 53–77. doi: 10.1016/j.wocn.2015.11.005

Tilsen, S. (2018). "Three mechanisms for modeling articulation: Selection, coordination, and intention," in *Cornell Working Papers in Phonetics and Phonology, 2018* (Ithaca, NY: Cornell University).

Tilsen, S. (2019). Motoric mechanisms for the emergence of non-local phonological patterns. *Front. Psychol.* 10:2143. doi: 10.3389/fpsyg.2019.02143

Tilsen, S. (2020). Detecting anticipatory information in speech with signal chopping. *J. Phon.* 82, 1–27.

Tilsen, S., Spincemaille, P., Xu, B., Doerschuk, P., Luh, W.-M., Feldman, E., et al. (2016). Anticipatory posturing of the vocal tract reveals dissociation of speech movement plans from linguistic units. *PLoS One* 11:e0146813. doi: 10.1016/j.wocn.2012.08.005

Torreira, F., Bögels, S., and Levinson, S. C. (2015). Breathing for answering: the time course of response planning in conversation. *Front. Psychol.* 6:284. doi: 10.3389/fpsyg.2015.00284

Turk, A., and Shattuck-Hufnagel, S. (2020). *Speech Timing: Implications for Theories of Phonology, Phonetics, and Speech Motor Control*. Oxford: Oxford University Press.

Ward, N., and Tsukahara, W. (2000). Prosodic features which cue back-channel responses in English and Japanese. *J. Pragmat.* 32, 1177–1207. doi: 10.1016/S0378-2166(99)00109-5

Wheeldon, L., and Lahiri, A. (1997). Prosodic units in speech production. *J. Mem. Lang.* 37, 356–381. doi: 10.1006/jmla.1997.2517

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., and Sloetjes, H. (2006). "ELAN: a Professional Framework for Multimodality Research," in *Proceedings of the LREC 2006, 5th International Conference on Language Resources and Evaluation*, Genoa.

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

frontiers in Communication

ORIGINAL RESEARCH
published: 13 September 2021
doi: 10.3389/fcomm.2021.730914

Genre Determining Prediction: Non-Standard TAM Marking in Football Language

Jakob Egetenmeyer*

SFB 1252/Department of Romance Studies, University of Cologne, Cologne, Germany

German and French football language display tense-aspect-mood (TAM) forms which differ from the TAM use in other genres. In German football talk, the present indicative may replace the pluperfect subjunctive. In French reports of football matches, the imperfective past may occur instead of a perfective past tense-aspect form. We argue that the two phenomena share a functional core and are licensed in the same way, which is a direct result of the genre they occur in. More precisely, football match reports adhere to a precise script and specific events are temporally determined in terms of objective time. This allows speakers to exploit a secondary function of TAM forms, namely, they shift the temporal perspective. We argue that it is on the grounds of the genre that comprehenders predict the deviating forms and are also able to decode them. In various corpus studies, we explore the functioning of these phenomena in order to gain insights into their distribution, grammaticalization and their functioning in discourse. Relevant factors are Aktionsart properties, rhetorical relations and their interaction with other TAM forms. This allows us to discuss coping mechanisms on the part of the comprehender. We broaden our understanding of the phenomena, which have only been partly covered for French and up to now seem to have been ignored in German.

Keywords: football language, TAM, genre, prediction, temporal perspective, French imparfait, German present indicative, temporal discourse structure

OPEN ACCESS

Edited by:
Franziska Kretzschmar,
Leibniz Institute for the German Language (IDS), Germany

Reviewed by:
Gerda Haßler,
University of Potsdam, Germany
Laure Lansari,
Université de Paris, France
Björn Rothstein,
Ruhr University Bochum, Germany

****Correspondence:***
Jakob Egetenmeyer
j.egetenmeyer@uni-koeln.de

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Communication

Received: *25 June 2021*
Accepted: *19 August 2021*
Published: *13 September 2021*

Citation:
Egetenmeyer J (2021) Genre Determining Prediction: Non-Standard TAM Marking in Football Language. Front. Commun. 6:730914. doi: 10.3389/fcomm.2021.730914

INTRODUCTION

German and French football language features tense-aspect-mood (TAM) forms which deviate from the TAM use in similar structural environments in other genres. We set out to investigate these phenomena in depth using various online corpora. Both phenomena are verbal in nature. They also share the property that the form used lacks a relevant feature which, by contrast, a competing form would express. However, they differ with respect to the relevant referential domain and therefore also in terms of the paradigmatic position of the forms in the language system. On the one hand, in German football language, the present indicative may take the place of the pluperfect subjunctive, which impacts the world reference coordinate. On the other hand, French football reports contain uses of the imperfective past tense-aspect form (*imparfait*) expressing sequences of events which would be realized by a perfective past in other genres. The deviation is thus located on the level of temporal reference. Although one might suspect erroneous interpretations in terms of truth values (for German) and temporal sequentiality (for French), the deviating forms do not seem to pose difficulties for comprehenders. The phenomena share strong functional parallels. Therefore, a combined analysis is highly fruitful. A particular benefit concerns our understanding of their licensing, which, according to our account, is rooted in the genre they appear in.

The analysis of the two languages indicates that it is specifically the genre which enables speakers to interpret the conflicting forms correctly. We argue that due to the properties of the genre, the precise semantic and script-based expectations outweigh the effect of the deviating TAM forms. Football reports are special in several respects. There is a reduced set of typical events, they adhere to a well-entrenched script, and the events referred to may also be temporally determined in terms of an objective time line. We argue that these properties allow the speakers to exploit secondary functions of TAM forms. They access the temporal coordinates and shift the temporal perspective (see de Saussure and Sthioul, 1999). Thus, the hypothesis is that in the case of football language, deviating TAM forms are predicted on the grounds of genre.

There has been some work on tense in English football language (see Walker, 2008), and also on the imperfectives occurring in French football language (see Labeau, 2004; Labeau, 2007; Egetenmeyer, In press). By contrast, the German phenomenon does not seem to have been covered in the literature. Furthermore, as we show in the present contribution, the French imperfective may in fact appear as the only inflected verb form in entire newspaper articles on football. Our investigation deepens the understanding of the different phenomena in German and French. On the grounds of several corpus studies with different kinds of data, we carve out the relevant properties and determine how the phenomena function within discourse. In order to achieve this aim, we address five research questions. The first two focus on the quality of the phenomena in the linguistic system. 1) In what kind of texts and contexts do the deviating TAM forms appear and what can we say about their frequency? 2) What is the status of the forms within the given linguistic system? The other three questions relate to the speaker's intentions and the role of the genre for the decoding of the message on the part of the comprehender. 3) For what reason do speakers apply deviant TAM forms? 4) How do comprehenders cope with the deviating TAM forms and how do they resolve the missing information? 5) Finally, what do these phenomena tell us about predictive language processing?

We proceed as follows. In the next section, we introduce the *Theoretical Background*. We show in what way the forms in question deviate from the standard TAM use (*German Tense-Aspect-Mood Forms: Present Indicative Substituting Pluperfect Subjunctive* and *French Tense-Aspect-Mood forms: Imperfective Substituting Perfective Past*). We discuss the commonalities of the uses (*Commonalities: Football Frame and Shifted Perspective Time*) and the role of predictions in their decoding (*The Role of Predictions and What We Can Learn from the Data*). In *Materials and Methods*, we describe the corpora used in this study and how we approach them. First, we focus on the German data (*Collection of German Data in Specialized and Non-Specialized Corpora*), then on the French (*Collection of French Data in Non-Specialized Corpora*). Then the results are presented. We begin with what we found on the Aktionsart of the verbs involved (*Aktionsart Properties*) and continue with properties pertaining to the discourse level (*Discursive Properties*). In *Discussion*, we discuss our results and summarize what we have found regarding the research questions mentioned above. Furthermore, we draw up observations on possible further research.

THEORETICAL BACKGROUND

This section covers the relevant theoretical background necessary to discuss the data of interest properly. This is important for the following reasons. We account for linguistic phenomena in two languages, German and French. The two languages pertain to different language families and the functioning of TAM categories is not identical. The two phenomena are different in nature, concerning mood choice in German and tense-aspect choice in French. However, we approach both of them with the same interest, namely, the role of genre for decoding them. Importantly, they are deviations from the expected forms. As comprehenders understand them without difficulties, we argue that they actually predict such forms to occur in football language. The object of study, especially the German TAM forms of interest, and the way we address the data is to some extent new; for instance, TAM research does not normally include the factor of genre, and therefore needs theoretical backing. Thus, finally, apart from the linguistic phenomena and insights into their commonalities, we also introduce predictive language processing in this section.

The subsections are ordered in the following way. *German Tense-Aspect-Mood Forms: Present Indicative Substituting Pluperfect Subjunctive* presents the German phenomenon of interest. *French Tense-Aspect-Mood forms: Imperfective Substituting Perfective Past* continues with the French counterpart. With regard to both languages, we specify the linguistic variety in which the phenomena in question appear. In the first two subsections, we will mention important characteristics. *Commonalities: Football Frame and Shifted Perspective Time* discusses the common core of the functioning of the German and French phenomena. It shows that they are both licensed by the specific properties of the genre. Thus, the section also motivates why we account for the two very different phenomena in a parallel fashion. *The Role of Predictions and What We Can Learn from the Data* replenishes the theoretical background with insights into the factor of prediction.

German Tense-Aspect-Mood Forms: Present Indicative Substituting Pluperfect Subjunctive

The formal deviation we are interested in with respect to German occurs in conditional clauses with past reference and a counterfactual or irrealis reading. In standard German, this is expressed by means of the pluperfect subjunctive, that is, "das Perfekttempus des Konjunktivs II" (Duden, 2009, p. 517, "the perfect tense of the subjunctive II"). (1) is drawn from the examples presented in Duden (2009, p. 518). Both *wäre festgebunden gewesen* ("had been tied down") and *hätte durchbohrt* ("would have pierced") are pluperfect subjunctive forms. Example (2), with the conjunction *wenn* ("if"), is a semantically close alternative.

1) *Wäre er festgebunden gewesen, hätte ihn die Stange sicher durchbohrt.* (Duden, 2009, p. 518, p. 518).
'Had he been tied down, the pole would surely have pierced him.'
2) *Wenn er festgebunden gewesen wäre, (dann) hätte ihn die Stange sicher durchbohrt.*
'If he had been tied down, (then) the pole would surely have pierced him.'

While not all uses of the forms of the German subjunctive are stable and substitutes may be found, especially in spoken discourse (see Duden, 2009, p. 516 and for instance, Gallmann, 2007 who, however, focuses on morphological reasoning), the TAM use in counterfactual conditional clauses has a distinctive value (see Duden, 2009, p. 518–519, see below). Thus, in order to express counterfactuality, the use of the pluperfect subjunctive is considered "unverzichtbar" (Duden, 2009, p. 540, "indispensable"). However, contrary to this expected standard, present indicative verb forms can be found in counterfactual conditional clauses in German football language (see example (4) below).

Thus, this first puzzle forms part of the realm of modality. According to the basic definition of Palmer (2001, p. 1), "[m]odality is concerned with the status of the proposition that describes the event." Although the definition has been said to be imprecise (see Salkie, 2009, p. 79), it may serve to emphasize a crucial point, as it does not specify how this status is brought about. In German conditional clauses, the protasis may or may not be introduced by the conjunction *wenn* ("if") (see examples (1) and (2) above). The alternatives *falls* ("if") and *sofern* ("provided that") are not compatible with the counterfactual reading (see Zifonun et al., 1997, p. 2280) and might be used as a test battery. When there is no conjunction, as in example (1), the subordinate clause is realized as a verb-first clause (see Zifonun et al., 1997, p. 2281). The apodosis may but does not have to involve the adverbial *dann* ("then") (see example (2)). Although, in general, certain cases of syncretism between indicative and conjunctive exist (see Zifonun et al., 1997, p. 1739–1743), the verbal forms occurring in the protasis and the apodosis indicate the propositional status in a largely unequivocal fashion (see Zifonun et al., 1997, p. 1745–1746). This is a more general phenomenon, which is not restricted to German; for instance, Portner (2009, p. 221–247) discusses the interplay between modality and tense-aspect forms for English. As Zifonun et al. (1997, p. 1745) put it, when there is an indicative in the protasis this may yield a hypothetical reading, but counterfactuality is generally ruled out. As noted, when counterfactuality is attributed to the condition, a pluperfect subjunctive is used in the subordinate clause. This also indicates past reference. Although in this case the main clause often features the same TAM form (see Duden, 2009, p. 518), it may also contain a past subjunctive (called *Konjunktiv II* or *Konjunktiv Präteritum*, see Fabricius-Hansen, 1999, p. 131) (see also Zifonun et al., 1997, p. 1745–1746). However, this alters the temporal reference. Declerck (2011, p. 28) calls this phenomenon ""[m]odal backshifting" (or "formal distancing")." Importantly, its functioning differs from the backshifting of verb forms found in indirect speech, and therefore should not be confused with it (see Declerck, 2011, p. 28; Duden, 2009, p. 516–541). Zifonun et al. (1997, p. 1746) present the following example.

3) *Wenn die Sängerin gelächelt hätte, {wären wir glücklich/wären wir glücklich gewesen}.* (Zifonun et al., 1997, p. 1746, adapted)
'If the singer had smiled, we {would be/would have been} happy.'

If we focus on the apodosis in (3), the first variant yields a co-temporal reference with regard to the moment of speech, while the preferred reading of the second possibility is one of past reference (see Leirbukt, 2008, discussed below). Still, Zifonun et al. (1997, p. 1747–1748) mention the less typical possibility that a counterfactual reading may arise with mixed forms in which one half of the structure contains a past subjunctive form while the other shows an indicative. However, as becomes apparent in the examples cited in Zifonun et al. (1997, p. 1747), a strong contextual determination is necessary and the structure appears to be marked.

In his introductory section, Leirbukt (2008, p. 1–6) presents the various possible ways of expressing potentiality and counterfactuality in German. As he shows, the distinction is not independent of the temporal localization (see Leirbukt, 2008, p. 1–6). Lewis (1979) takes a philosophical stance and discusses the role the divergence of an invariant past, as opposed to an undetermined and therefore flexible future, has on counterfactuals. Now, as already noted, when the German pluperfect subjunctive occurs in a past context, a counterfactual reading is typically realized; however, it is not the only possibility, as Leirbukt (2008, p. 27 with further references) shows, although he focusses on non-past temporal reference (see Leirbukt, 2008, p. 6).

However, we assume that in general, the argument cannot be inverted. As noted, in order to express a counterfactual reading in the past, the pluperfect subjunctive should be necessary. But football language teaches us otherwise, as example (4) shows. Up to this point, in our inquiry into the research literature, we have neither found reference to the present indicative substituting the pluperfect subjunctive in general, nor to its highly interesting use in football language. Furthermore, we have not come across this phenomenon in French, for which we analyze another phenomenon, as presented in the following section (see Becker, 2014 for a description of mood in Romance languages).

4) *Latza [. . .] dachte nach dem 1:1 in der Domstadt an die vergebene Großchance des eingewechselten Teamkollegen Robin Quaison in der letzten der 97 Minuten: "Wenn er den* ***macht****,* ***heulen*** *hier 50.000 rum – und wir* ***freuen*** *uns. Schade."* (FR 1)
'After the 1:1 in Cologne, Latza thought about the missed big chance of the substitute teammate Robin Quaison in the last of the 97 minutes: "If he had converted (lit.: converts) that one, 50,000 people would have cried (lit.: cry) with

disappointment—and we would have been (lit.: are) happy. Too bad."'

As the example shows, a clearly counterfactual proposition is conveyed by means of verbs marked by the present indicative in both the protasis and the apodosis of the conditional clause. To our knowledge, beyond football language this is ruled out. Crucially, the use shows a tendency towards being an oral phenomenon. However, as shown by the above example and example (5) below, it is brought into the written form within direct quotes. Furthermore, in our analysis, we found cases from live tickers, a written (but close-to-speech) variety (see *Discursive Properties*). In example (4), the occurrence shows two further markers of genre and sociolinguistic status, namely, the structure in the protasis (*den machen*, "convert that one") is a typical expression in the realm of football, with a noticeable marker for the language of proximity in the terms of Koch and Oesterreicher (2011). Furthermore, the apodosis features the verb *rumheulen* ("whine"), which is marked as colloquial. According to our data, example (4) may be seen as a quite typical instance of the phenomenon. Among the factors to discuss are the following. The protasis features a telic event expression (*den machen*, "convert that one"). The apodosis expresses an activity (*rumheulen*, "whine") (and an additional state, *freuen*, "are happy"). They show a rhetorical relation of consequence (see Asher and Lascarides, 2003, p. 169), which may be considered less typical.

However, apart from the genre restriction, the phenomenon is versatile. Most importantly, it has two different syntactic instantiations paralleling the introductory examples (1) and (2). The second type is exemplified in example (5), where the protasis lacking the conjunction *wenn* ("if") is realized as a verb-first sentence (see Zifonun et al., 1997, p. 2281). Following the apodosis, another consequence of the situation expressed by the conditional clause is expressed. Interestingly, the speaker switches to the pluperfect subjunctive (*wären gelaufen*, "would have gone"). The switch underlines that the speaker is well aware of the counterfactual meaning of his own words. Apparently, the standard TAM form expressing this kind of world reference is available as an alternative.

5) *[D]er Innenverteidiger, der sich noch immer über seine vergebene Kopfballchance im letzten WM-Gruppenspiel gegen Südkorea ärgert[, sagte]: "**Mache** ich das Tor, **kommen** wir gegen Südkorea weiter, dann wären viele Dinge sicherlich anders gelaufen."* (Spiegel 1).

'The central defender, who is still upset about his missed header chance in the final World Cup group match against South Korea, said: "If I had scored (lit.: score), we would have succeeded (lit.: succeed) against South Korea to the next round, then many things would probably have gone differently."'

However, as our final example (6) indicates, a speaker may also continue with the present indicative when expressing a further consequence. This example comes from a live commentary. It underlines two important properties of this use especially clearly. First, it indicates grammatically (*wär' gewesen*, "would have been") and lexically (*Theorie*, *Theorie*, *Theorie*, "theory, theory, theory") that reference is made to a counterfactual situation. Second, it excludes a generalizing meaning, as reference is made to a specific event in which a player (Volland) did not get the pass he was supposed to get. As the feature of specific reference is an important property of the structure investigated here, it is also given in examples (4) and (5). However, in (4) and (5) the referential status is not determined by the same speaker, but rather by the author of the article, while the structure of interest is part of a direct quote of an interviewee. By contrast, in example (6), there is only one speaker. As noted, in the sentence following the conditional structure, the speaker goes on to speak of a further consequence the successful pass would have had and maintains the present tense.

(6) *Und das wär' das Tor gewesen. Theorie, Theorie, Theorie, noch ist nicht Schluss. **Kommt** der Ball, **kann** Volland den machen. Dann heißt der Gegner nicht England, sondern Schweiz. Aber die Uhr tickt noch* (ZDF 1).

'And that would have been the goal. Theory, theory, theory, it's not over yet. If the ball had gotten (lit.: gets) to him, Volland could have scored (lit.: can score). Then the opponent would not have been (lit.: is not) England, but Switzerland. But the clock is still ticking.'

French Tense-Aspect-Mood Forms: Imperfective Substituting Perfective Past

In classic literary French, two main past tense forms are used which express the aspectual distinction between perfective (*passé simple*) and imperfective (*imparfait*). Diachronically speaking, the *passé simple* has been extensively substituted in oral discourse and, to a certain extent, also in written discourse by the compound past (*passé composé*) (see Verkuyl et al., 2004, p. 253, 265–266). However, the opposition with the imperfective past is maintained (see Molendijk et al., 2004, p. 298; refinements can be found, however, in Facques, 2002). Thus, typically, series of past events are expressed by the simple past or the compound past, while the imperfective past tense-aspect form is used for co-occurring or background eventualities (see Weinrich, 1964; Kamp and Rohrer, 1983). However, in certain contexts, the imperfective past may be used instead of its perfective counterpart. This phenomenon is often called the *imparfait narratif* ("narrative imperfect"; see Gosselin, 1999; Bres, 2005; and others). Importantly, such occurrences are quite restricted in terms of syntactic, contextual and genre-related terms (see Caudal, submitted; Egetenmeyer, In press). Apart from literary texts and newspaper articles on politics (see Egetenmeyer, In press), we also find football reports among the genres which feature such imperfective tense-aspect uses (see Labeau, 2007, p. 220 who quotes Herzog, 1981, p. 67 as having noted the distribution quite early). Importantly, as Egetenmeyer, In press underlines, its use in football reports is peculiar: It is the only genre which seems to allow for a full substitution of non-imperfective tense-aspect forms with the imperfective (however, see Facques, 2002, p. 115 for an example coming from a non-football related newspaper article which also shows a rather strong tendency to avoid non-imperfective tense-aspect forms). An important differentiating property is that while the typical narrative imperfect is normally embedded under an adverbial expression (see Vetters, 1996, p. 128), as shown in example (7), in football reports explicit reference to times may be dropped entirely.

7) *Quinze jours plus tard, lady Burbury qui résidait en compagnie de son époux dans leur domaine de Burbury, **s'éprenait** d'un jeune pasteur des environs, venu déjeuner au château.* (Aymé, 1968, p. 38, cited after Tasmowski-De Ryck, 1985, p. 60, p. 60).
'A fortnight later, Lady Burbury, who was residing with her husband at their Burbury estate, fell in love with a young clergyman from the area who had come to lunch at the castle.'

In (7), the adverbial expression *quinze jours plus tard* ("a fortnight later") introduces a (relative) time point at which the event of falling in love (*s'éprendre*) is realized. As Egetenmeyer (In press) shows, this factor, along with other properties, restricts the use in comparison with the one found in football reports. Interestingly, whole reports of football matches may be written using the *imparfait* where otherwise perfective (or non-imperfective) tense-aspect forms would be used. When comparing the situation with Spanish, which shows many parallels in the use of the corresponding imperfective past (see Escandell-Vidal, submitted with further references), we find that the usage in question is not paralleled there. For instance, as Quintero Ramírez and Carvajal Carvajal (2017, p. 229) indicate, in Mexican Spanish newspaper reports the imperfect past tense is not used to express sequences of eventualities.

Example (8), taken from Egetenmeyer (In press), is the beginning of a newspaper article reporting a football match. All five finite verb forms would be expected to be realized as compound past forms due to the expression of sequences of events. However, they are all marked by the imperfective, as are most of the other finite verbs in the rest of the article (see Egetenmeyer, In press).

8) [1] *Le Blésois Gonçalves **était** le premier à se mettre en action (10^{e}),* [2] *mais sa frappe **passait** juste à côté.* [3] *Les locaux **répondaient** de suite, avec une bonne tête de Maelbrancke,* [4] *mais le défenseur Radet **sauvait** sur sa ligne.* [5] *La réponse blésoise ne se **faisait** pas attendre [...].* (Sketch Engine: *La Nouvelle République*, 22.08.2016)
'[1] The Blesoisian Gonçalves was the first one to get into gear (10th), [2] but his shot just missed. [3] The locals responded immediately with a good header by Maelbrancke, [4] but the defender Radet saved on the line. [5] The Blesoisian answer was not long in coming.'

As we will see, in such structures, there is a high proportion of verbs lexically expressing boundedness (see also Bres, 1999, p. 5), which we assume facilitates processing. As noted above, the specialty of this usage is that no temporal determination is necessary in order to license it (see Egetenmeyer, In press). This contrasts with other similar uses, such as the typical narrative imperfect, which tends to co-occur with a temporal sentence adverbial under which it is embedded (see Egetenmeyer, In press), but also uses appearing in relative clauses (see Caudal, submitted, who analyzes the examples in Bres, 2005). Those types of uses show a direct or indirect temporal determination of the expressed eventuality. It should be mentioned, however, that the above example does in fact contain an explicit (relative) temporal indication, namely, "10^{e} [minute]" ("10th minute"). In terms of discourse structure, it is relevant to note that the indication is realized as an insertion and is therefore not syntactically integrated into the sentence. As our data show, such an indication is not necessary for the usage. However, as discussed in the following subsection, it makes a principle explicit which we assume to be relevant for the occurrence of imperfective tense-aspect forms in football reports, namely that it shows that a football match is measured in terms of an objective time. In addition to this principle, possible temporal indications and temporal adverbs, as well as verb meaning (in terms of and beyond Aktionsart), support the decoding of temporal relationships when distinctive grammatical means are not available. It should be noted that not all imperfective forms occurring in football reports have the same function (see *Discursive Properties*). For instance, the imperfective verbs may lack narrative features altogether. By contrast, there may also be instances of the typical narrative imperfect, although this is not frequent.

In example (9), a temporal specification is only given in the case of a decisive event within the match (see clause [5]). The adverbial expression (*la 21^{e} (minute)*, "the 21st minute") is syntactically integrated into the discourse. Furthermore, clause [4] includes an adverb (*puis*, "then"), which typically expresses a sequence and may therefore be considered a relevant tool in contexts lacking grammatical markers of sequentiality as in the phenomenon at hand. However, in the example, it does not strictly express a sequence but has additive meaning. In terms of rhetorical relations, clauses [2] to [4] are an elaboration with regard to [1] (see Jasinskaja and Karagjosova, 2020, for subordinating rhetorical relations). Although the order of mention is meaningful, only the events expressed in [2] and [3] are temporally adjacent, not those of [3] and [4]. Thus, the example underlines the temporal flexibility of the imperfect in football reports.

9) [1] *Les Malouins se **montraient** les plus entreprenants,* [2] *Desmenez **tirait** un corner dangereux* [3] *qui **contraignait** Shungu, le portier visiteur à s'imposer* [4] *puis un coup franc de Desmenez **mettait** en diffficulté Shungu.* [5] *Il **fallait** attendre la 21^{e} pour voir une première offensive des visiteurs grâce à son attaquant Orhand* [6] *qui **obligeait** Favris, le gardien local à se détendre.* (Emolex: *Ouest-France*, 19.03.2007).
'[1] The Malouins showed themselves as more enterprising, [2] Desmenez shot a dangerous corner [3] that forced Shungu, the visiting keeper, to come forward [4] and then a free kick from Desmenez challenged Shungu. [5] It was necessary to wait until the 21st minute to see a first offensive of the visitors thanks to their forward Orhand [6] who forced Favris, the local keeper, to reach out.'

Finally, when the report is reduced to the most important events in very brief match presentations, the temporal specifications may be indicated regularly, as in example (10). In this example, the temporal indications are again introduced within the inserted brackets and are not syntactically integrated into the discourse.

10) [1] *Pour ses grands débuts sur le banc de Montpellier, Frédéric Hantz* ***était*** *servi.* [2] *Après un débordement de Martin, Yatabaré* ***ouvrait*** *la marque à bout portant (15^e^).* [3] *Martin* ***marquait*** *ensuite le but du break sur penalty (41^e^)* [4] *et Dabo* ***clouait*** *enfin le spectacle avec un doublé face à un Ajaccio méconnaissable (53^e^ et 58^e^)* (Sketch Engine: *Foot01.com*, 369844001).

‘[1] For his big debut on the Montpellier bench, Frederic Hantz was served. [2] After a cross attack by Martin, Yatabaré opened the score from close range (15th). [3] Martin then scored the breakthrough goal from the penalty spot (41st) [4] and Dabo finally closed the show with a double against an unrecognizable Ajaccio (53rd and 58th).’

Commonalities: Football Frame and Shifted Perspective Time

The two preceding subsections introduced the basic characteristics of the phenomena of interest. While the phenomena show certain parallels, they are also different in important respects. They share the basic property of pertaining to the verbal domain. In both languages, the TAM marking deviates from what would be expected in a different genre. Simply put, the marking would not withstand a normative stance. By contrast, an important difference is that the German TAM marking deviates in the realm of world reference, while the French counterpart shows its deviation with regard to temporal reference. Finally, they share two decisive properties, which also motivate their joint treatment. First, an important licensing factor for their realization, which is directly connected with the factor of genre, is given by the frame or the script of football matches. Second, the functioning of both phenomena can be explained as a shift in perspective time. In the following, we go into the details of these last two ideas.

A football match is conventionalized and functions according to a specific set of rules, of which at least the basic ones are known to the general public in the speech communities relevant for this paper (see also the interesting properties ascribed to football reports in newspapers by Engel and Labeau, 2005, p. 204–205, with reference to Grevisse, 1997, some of which, however, would need sociological verification; in addition, see Hennig, 2000, p. 43–44, for live football reports on television and radio). Therefore, (at least) the central information block of football language may be taken to show relevant features covered by accounts of scripts (see Schank and Abelson, 1977, p. 36–68) and frames (Fillmore, 1977; Fillmore, 2006). There are three further related properties of football matches which distinguish them, for instance, from the often-cited restaurant script. First, the non-generalized events of the football match are directly related to an objective time line. In a related matter, Engel and Labeau (2005, p. 215, with further references) remark that in football reports the events are often presented chronologically. Second, the matches of interest to a large group of people are televised. The large group of passive participants (i.e., viewers) and the factor of television broadcast further objectivizes the match, as there are many witnesses and the match can be (and is, in fact) recorded to be watched again. Importantly, the objective temporal determination of specific events is not circumstantial, but plays a decisive role for the match. For instance, it may have consequences for tactical planning. Furthermore, due to its role within the match, it has a high informative value in football reports and many other situations where football language is used. These very specific temporal properties have an influence on what needs to be conveyed in a relevant speech situation. We assume that it is due to such properties that the rigidity of certain components of the linguistic system may be attenuated. As a consequence, other functions may be exploited. As we will see, both languages make use of this principle in a similar way, although they diverge in terms of what exactly is modified.

In terms of discourse structural functioning (see Becker and Egetenmeyer, 2018 for our conception of temporal discourse structure), the two phenomena adhere to the same principle, namely, they show a shift in temporal perspective. For the French narrative imperfect, this has been discussed in a similar vein by Berthonneau and Kleiber (1999), de Saussure and Sthioul (1999) and Schrott (2012). Further relevant discussions come from the realm of free indirect discourse (see, for instance, Banfield, 1982; Ehrlich, 1990; and Eckardt, 2014; distinctions between different forms of perspective taking are presented in Hinterwimmer, 2017). Labeau (2006) mentions perspective specifically in the context of television talk, a category to which the genre of football reports partly pertains. According to Becker and Egetenmeyer, 2018, p. 37, with reference to Guéron, 2015, p. 278), the perspective time “is the point in time where the text-internal origo is situated.” While, in a standard narration of successive events expressed by means of verbs marked with the perfective past, the perspective time is anchored to the speech time, the eventive use of the *imparfait* in football reports is accompanied and licensed by a shift in perspective to the past. In terms of de Saussure and Sthioul (1999, p. 6), the perspective time is included in the run time of the event, which, however, may be determined more precisely as the location time corresponding to the event (see Becker and Egetenmeyer, 2018). With regard to football language, we have to keep in mind a further component which seems to be missing in the above-mentioned publications, namely, that this perspective time needs to be continuously updated as the events are narrated one after the other by means of verbs marked by the *imparfait*. A similar principle is described by De Swart (2007, p. 2282) with regard to the special use of the French present perfect in Camus’ *L’Étranger*, which, according to her, is mirrored by the adverbials used in the text. Schrott (2012) emphasizes the underlying perception of the perspectivizing entity with the narrative *imparfait*. Envisioned in this way, this use of the *imparfait* could be understood as a means to bring the report closer to the speaker / hearer and thereby to render it livelier (see, however, Labeau, 2007, p. 220, who notes that the literary narrative *imparfait* with a temporal adverbial tends to render a passage rather clumsy). In this way, an (intended) immediacy of the experience is conveyed (see also below). The actualization of the secondary (competing) perspectivizing function in the footballer's context might also be applied to the English narrative present perfect, described in the football context by Walker (2008), and even to the use of the present perfect in Australian police reports (see Ritz, 2010). These two publications, however, do not mention this interpretation.

Although the German phenomenon we are interested in does not pertain to the realm of tense-aspect but to the modal domain, it may also be interpreted in this way. In this interpretation, the speaker shifts the temporal perspective to a past reference time, which, as we saw above, may correlate with a distinct objective time. From this past perspective time, the expressed event is posterior; that is, it is a kind of future in the past. Correspondingly, conceptualized from this perspective time, the realization of the event is still possible. The present indicative is then the corresponding TAM choice. Thus, even more clearly than in French, the effect of an immediacy of the experience in operationalized. An important clue to substantiate our hypothesis is, as we will see, that the collected instances all pertain to oral or close-to-speech varieties (see *Discursive Properties*) (see Wüest, 1993, p. 231 for the varying strength of correlation between temporal perspectivization and text types). We have already noted above that the phenomena at hand make use of secondary functions of the linguistic forms. In oral speech, exploiting the potential of flexibility of language is even more common (see also Labeau, 2006, p. 18–19).

The Role of Predictions and What We Can Learn from the Data

When processing language, comprehenders partly resort to prestored knowledge in order to predict what is to come next (see Kuperberg and Jaeger, 2015 for a theoretical overview). Thus, prestored knowledge has an important function in communication. As Kuperberg (2013, p. 14) underlines, the "benefit of a predictive language processing architecture is comprehension efficiency." It may also be calculated, as shown, for instance, by Levy (2008), who focusses on surprisal (that is, basically, when predictions are not met). The research in the realm of predictions frequently discusses verbal properties and also touches upon the role of genre (see below). Importantly, as we will indicate below, we suggest an alternative way of approaching the phenomena, which takes genre to be a predictor of deviating TAM forms. Addressed in this way, football language is an especially revealing case. The analysis of corpus data containing the phenomena in question is also a basis from which future experimental research may profit.

As discussed in the preceding sub-sections, the analyzed structures deviate from otherwise expected ones. This is especially interesting against the backdrop of predictive language processing. First, we might be inclined to ask how speakers would deal with deviating TAM forms if they interpreted them as errors. Hanulíková et al. (2012) indicate that when confronted with speakers with non-perfect acquisition of an L2, hearers do not show any reaction to syntactic errors (no P600 effect). Kuperberg (2013, p. 17) calls this "predictive error-based learning." One might be inclined to think that the same is happening in the case of the deviating TAM forms in football language. There are two main arguments against this view. First, as already indicated, the phenomena of interest are not extremely rare. Second, journalists and reporters are not completely free in their linguistic choices. We may assume that their employers would refuse to accept an overly individual or defective style of speaking or writing. By contrast, the use of a specific diastratical marking in order to indicate pertinence to a group (see Koch and Oesterreicher, 2011), such as the group of football fans, may be permissible or even desirable in order to reach a high number of listeners or readers. However, idiosyncratic markers should never be too strong, in the sense that the content of the utterance still needs to be comprehensible to the general public. This is another argument favoring licensing effects, as discussed in *Commonalities: Football Frame and Shifted Perspective Time*. Again, if a hearer/reader were to incur processing difficulties every time she/he is confronted with a non-standard TAM form, the form would quickly be banned from the genre it appears in. By contrast, as the phenomena in question are frequent in football speech, they cannot be expected to be overly costly in terms of the comprehenders' processing. Rather, we assume a genre-based prediction that such TAM forms will occur.

What do we know about the role of TAM forms in predictive language? In the research literature, verbal properties are shown to play a crucial role in the predictive processing of language (see, for instance, Kuperberg and Jaeger, 2015, p. 6–7, and the references therein). Among the best studied verb-related factors are its argument properties, such as its selection restrictions (see Altmann and Kamide, 1999). However, tense and aspect have also been investigated (see Kuperberg and Jaeger, 2015, p. 7; Arai and Keller, 2013, p. 5; both with further references). For instance, Altmann and Kamide (2007) show that the tense-aspect morphology is relevant for predicting the verb's object. Furthermore, Philipp et al. (2017) discuss the interaction of semantic roles with telicity and event structure with respect to processing costs. Graf et al. (2017) analyze the interaction between verbal (telicity) and nominal (agentivity) features. As a final example, Dery and Koenig (2015) specifically focus on temporal updates. By contrast, mood and modality do not seem to have aroused much interest. However, genre, our second category, has also been studied (see Kuperberg and Jaeger, 2015, p. 16). For instance, according to Fine et al. (2013, p. 15), comprehenders are sensitive to genre and other more individual properties of the input, and they "continuously adapt their syntactic expectations" as they are exposed to new linguistic input. Further evidence can be found in Squires (2019). She discusses insights from several studies underlining that the knowledge of the listener about the speaker in terms of dialect or sociolect positively influences how possible expectation violations are processed (see Squires, 2019, p. 2–4, with further references). More specifically, she presents three experiments concerning the role the genre of pop songs has on the evaluation of non-standard morphosyntactic forms by listeners (see Squires, 2019, p. 8–23). She concludes "that speech genre can serve as an expectation-shifting sociolinguistic cue during sentence processing" (Squires, 2019, p. 23). However, according to Squires, the listeners' expectations are rather vague with respect to what phenomena might occur and how they deviate from standard language (see Squires, 2019, p. 23).

An important part of the research literature focusses on syntactic or morphological markers as predictors. By contrast, Fine et al. (2013) take syntactic structure as the predicted component. Similarly, we analyze TAM forms as predicted elements. To be precise, we argue that they are predicted on the grounds of the genre they occur in. The relevance of this idea is backed by Kuperberg and Jaeger (2015, p. 4, with reference to Anderson, 1990), who state that a comprehender will "use all her stored probabilistic knowledge, in combination with the preceding context, to process th[e] input." However, we do not analyze prediction locally, that is, in a sentence-based way, but globally in the sense of the classification of a whole discourse in terms of the subject it covers. We find a similar principle in the study by Schumacher and Avrutin (2011), who analyze the role of a certain discourse type for the processing of article-less noun phrases. The discourse type that they are interested in, a classification for which they use the category term "register" (see Schumacher and Avrutin, 2011, p. 306, footnote 3, for the way they determine it), is that of newspaper headlines. The authors present two studies involving NPs without an article, where the participants are only made aware of the pertinence of the critical items to a headline in one study, not in the other (Schumacher and Avrutin, 2011, p. 307). They show "that awareness of a particular register, and the expectations associated with it, has an impact on the readers' processing patterns" (Schumacher and Avrutin, 2011, p. 318). On these grounds, we assume that an analysis in our terms is promising.

MATERIALS AND METHODS

As its definition is content-based, football language is a broad concept linguistically (see Burkhardt, 2006 for a distinction of three lexically motivated sub-types). Football-related content may be presented in very different situational (for instance, with regard to the aim of the speech or writing event), social (with respect to the speaker/hearer constellation) and medial (concerning the medium) contexts. Typical situations for football language are football news reports on the radio, live commentaries on television (see Hennig, 2000, p. 43–44, for properties distinguishing the two), newspaper articles or printed interviews, but also fans talking among themselves. Thus, the concept of football language needs to capture a possible diversification in terms of diaphasics, diastratics and diamedial realization (see Koch and Oesterreicher, 2011). We assume an oral predominance with regard to the general use of football language, which however, may be taken to be a general property of linguistic data (see Sinclair, 2005, *Discussion*, who classifies the distribution as a general problem for corpora). However, due to practical issues we restrict ourselves to written data at this point, which may include direct quotes and spoken interviews published in a written format.

An important part of the insights generated comes from the data collection process and the challenges we encountered when retrieving the linguistic material from the corpora. Therefore, the present section goes into further details regarding the data analyzed. It is especially dedicated to the corpus studies we conducted, the factors we considered and how we solved the issues that arose.

In order to understand the phenomena in depth, we collected data from pertinent corpora. In both languages, the linguistic material was annotated upon retrieval. Although the phenomena analyzed, in their core, both boil down to deviant TAM forms, they do not pertain to directly parallel verbal paradigms. Due to the systemic differences, different factors need to be accounted for in their retrieval and in the ensuing analysis. On the one hand, in German, the syntactic structure is relevant, in the sense that the occurrences are combinations of a subordinate and a main clause. However, as the annotations of the corpora we used do not consider syntactic structure, we can only make use of it in the corpus queries in cases where an explicit subordination marker is realized or a specific word order is used (verb-first). By contrast, an implicit subordination is much more difficult to find. On the other hand, in French what is of most interest is basically the re-occurrence of a tense-aspect form in a sequence of sentences. But the information on the quantity of a form within the textual entity it occurs in (the sentence, the paragraph or a text in its entirety) cannot be directly retrieved from a standard online corpus.

An issue beyond the linguistic means of expression is the availability of specialized corpus data, which differed between the two languages. With respect to German, we had specialized corpora at our disposal (see *Collection of German Data in Specialized and Non-Specialized Corpora*). Interestingly however, relevant examples were not easy to find there, which might be considered a sign of low frequency. By contrast, we did not have specialized French corpora. Therefore, in order to collect the data, we used specific verbs and collocations (see *Collection of French Data in Non-Specialized Corpora*). Furthermore, as we intended to analyze whole articles in French, we extended the passages retrieved step by step using the corresponding database or through queries on the general internet.

Collection of German Data in Specialized and Non-Specialized Corpora

As already noted, the German phenomenon does not seem to have been described in the literature. Therefore, we started off with an unstructured study using the general internet (via www.google.de) in order to determine relevant factors for its occurrence. They concerned the syntactic structure and, more importantly, the relevant verbal types (see below). For the main study, we had corpora at our disposal which specifically contained German football language. Interestingly however, we only found few instances of the phenomenon of interest. Therefore, we additionally used unspecialized corpus data to balance our findings. Finally, we conducted a counter-study in order to test whether there are similar phenomena beyond football language. We analyzed sub-corpora contained in Cosmas II, a

TABLE 1 | Overview of the German corpus study.

Corpus	Specialized corpus data			Non-specialized newspaper corpus data		Counter-study
	KSP - Fußball-Spielberichte, kicker.de, 2006–2016	**KIC - Fußball-Liveticker, kicker.de, 2006–2016**	**SID - Fußball-Liveticker, Sport-Informations-Dienst, 2010–2016**	**KSA - Kölner-Stadtanzeiger, 2000–2019**	**FNP - Frankfurter Neue Presse, 2000–2020**	**ZEIT - Die Zeit, 1953–2020**
Type of data	Match reports	Live ticker	Live ticker	Regional newspaper	Regional newspaper	National weekly newspaper
Number of texts (rounded)	3,000	3,000	1,800	2 million	2.6 million	388,000
Number of tokens (rounded)	1.9 million	3.5 million	3.8 million	600 million	700 million	343 million
Number of verbs investigated	9	9	9	4	4	4
Number of examples examined	423	474	380	730	697	800
Number of relevant hits	0	1	5	3	3	0
Linguistic production of relevant hits	—	Report	Report	Direct speech	Direct speech	—

database issued by the Leibniz-Institut für Deutsche Sprache (IDS) (https://www2.ids-mannheim.de/cosmas2/). It comprises 570 sub-corpora with a total of 56.5 billion tokens (https://www2.ids-mannheim.de/cosmas2/uebersicht.html, accessed: June 16, 2021). The sub-corpora are organized into eighteen different archives (https://www2.ids-mannheim.de/cosmas2/projekt/referenz/archive.html, accessed: June 16, 2021). **Table 1** gives an overview of the studies conducted.

In a first step, we analyzed three sub-corpora of football language, which are contained in the corpus "W - Archiv der geschriebenen Sprache." The first one is "KSP - Fußball-Spielberichte, kicker.de, 2006–2016." As the name indicates, it consists of a collection of data from the German football journal *Kicker*. The other two corpora consist of data from football live tickers. More precisely, the second corpus, "KIC - Fußball-Liveticker, kicker.de, 2006–2016" collects live ticker data coming from the very same journal. The third sub-corpus, "SID - Fußball-Liveticker, Sport-Informations-Dienst, 2010–2016," contains live ticker data from the biggest German sports news agency (see https://www.journalistenkolleg.de/service/organisationen/sid-sport-informations-dienst, accessed: August 6, 2021). "KSP" and "KIC" each amount to 3,000 texts. However, "KSP" contains 1.9 million tokens, while "KIC" comprises nearly 3.5 million. "SID" contains approximately 1,800 texts with close to 3.8 million tokens. Already in the unstructured pre-study, we found a tendency for an oral predominance. However, there are two reasons why the three sub-corpora appear to be promising despite the fact that they present written texts. First, they may contain direct quotes. Second, the live tickers (KIC and SID) may be said to be rather close to the pole of a language of immediacy on Koch and Oesterreicher's (2012) scale of linguistic conception.

The corpus research with regard to the three sub-corpora had to be carried out in two steps. The first step exploited an important advantage of Cosmas II, namely that the entries are lemmatized and the corpus allows for relatively complex queries. As discussed in *German Tense-Aspect-Mood Forms: Present Indicative Substituting Pluperfect Subjunctive*, the forms we are interested in are simple present tense forms (or share the form of the present indicative). As a consequence, we had to deal with a considerable amount of noise. Therefore, in the second step, we went through the hits manually and retrieved the relevant examples in order to analyze them more in depth.

In formulating the corpus queries we took into consideration the two different syntactic possibilities, namely, 1) clauses introduced by an explicit conjunction *wenn* ("if"), or 2) clauses displaying verb-first word order. For the case of an overt conjunction, we used the query that the verb should follow within a range of five words from the conjunction. We intended the verbal inflection not to be restricted. Thus, we used the query, "wenn /+w5 &verb," where we filled the verb slot with a specific lexical item (see below). The second query type specified a verb-first clause (see *German Tense-Aspect-Mood Forms: Present Indicative Substituting Pluperfect Subjunctive*). In Cosmas II, this can be spelled out as "&verb /w0 <sa>," which determines that the verb should occur as the beginning of a new sentence.

In the previously conducted unstructured analysis using the search engine google (www.google.de), we had intended to find out what verbs may occur in such contexts. There, we used simple co-occurrence patterns of different verbs in the present tense and denotations of individuals typically involved in football activities like *Torwart* ("goalkeeper") and *Schiedsrichter* ("referee"). The findings indicated that at least two factors were relevant. First, we only found verbs expressing typical actions in football matches, especially if they were compatible with rather colloquial football language. Due to its seeming frequency, an appropriate example is *machen* ("make") combined with the short demonstrative pronoun (*den*) in the structure *den machen* ("to score a goal (as part of a specific opportunity)"). Second, all the verbs we found were telic.

On these grounds we chose nine different verbs for our structured study of the specialized corpora, among them *flanken* ("to center"), *halten* ("to stop (a ball); to save a penalty") and *verwandeln* ("to convert (for instance, a penalty)"). All of these verbs are typical of football contexts. In the uses we looked for, they are all telic or ingressive and most have a punctual reading. We considered all of them in both syntactic configurations. This led to a total of six relevant hits. The amount of noise was considerable. In "KSP," there were 423 hits, of which, however, none was relevant. Of the 474 hits in "KIC" only one was relevant. By contrast, in "SID," there were 380 hits, including the remaining five relevant ones.

In a second step, we chose the sub-corpora of two local newspapers with a relatively high circulation, namely the *Kölner Stadt-Anzeiger* and the *Frankfurter Neue Presse*. They both belong to newspaper groups which are often listed among Germany's top ten regional newspaper groups (see, for instance, for the second quarter of 2019 https://meedia.de/2019/07/22/die-auflagen-bilanz-der-groessten-83-regionalzeitungen-kaum-noch-titel-mit-einem-minus-unter-2/, accessed: June 16, 2021). Both sub-corpora are contained in "W2 - Archiv der geschriebenen Sprache." "KSA - Kölner-Stadtanzeiger, 2000–2019" contains 20 years of the newspaper, amounting to over two million texts with more than six hundred million tokens. "FNP - Frankfurter Neue Presse, 2000–2020" covers 21 years, close to 2.6 million texts and over seven hundred million tokens. Again, we used both query types presented above, but we reduced the target verbs to four. We chose verbs which had a high probability, in terms of their lexical content, of occurring in football contexts. They express scoring (*machen*, "to make," see above, *treffen*, "to hit," *verwandeln*, "to convert") or saving a goal (*halten*, "to save") and are all compatible with a colloquial style. We restricted the study to the first one hundred occurrences per query. However, the structures involving *verwandeln* ("to convert") are less frequent and both sub-corpora yielded fewer than one hundred cases per query. The resulting 1,427 hits are comparable to the amount of data from the specialized corpora. In the same vein, the hits were examined manually in order to retrieve the relevant cases. Again, there was a high proportion of noise. However, although the corpus data was not reduced to football content in this case, six of the hits are relevant in our terms. They are distributed evenly, with three examples per sub-corpus. Interestingly, all six instances present direct speech. This matched our expectation based on the preliminary unstructured studies. By contrast, the hits from the live tickers (KIC, SID) were not instances of direct speech; however, given the nature of live tickers, they represent a close-to-speech variety (see *Discursive Properties*).

Finally, we also conducted a minor counter-study. The aim was to minimize the possibility that the phenomenon could be more pervasive and not restricted to football language. We used a set of four verbs which do not express typical football actions. However, they have comparable properties to the verbs in our football set, namely that they are telic and, due to their semantic content, may easily refer to decisive events within the texts where they appear (see *Discussion*). We investigated *beenden* ("to bring to an end"), *entscheiden* ("to decide"), *sterben* ("to die") and *unterbrechen* ("to interrupt"), and again tested both structures. We used the sub-corpus "ZEIT - Die Zeit, 1953–2020," included in "W - Archiv der geschriebenen Sprache" of Cosmas II. *Die Zeit* is the weekly newspaper with the second-highest circulation in Germany (see https://meedia.de/2020/01/17/die-auflagen-bilanz-der-tages-und-wochenzeitungen-bild-und-welt-verlieren-erneut-mehr-als-10-die-zeit-legt-dank-massivem-digital-plus-zu/, accessed: August 6, 2021). We chose the data of this national weekly in order to allow for a wide range of subjects and more varied linguistic structures. The sub-corpus, which covers 68 years, contains close to 388 thousand texts and 343 million tokens. As before, we restricted our study to the first one hundred hits per item, totaling another eight hundred hits. We examined them manually. Importantly however, none of them instantiated the structure of interest.

Collection of French Data in Non-Specialized Corpora

In contrast to the German data, for French we did not have specific corpora of football language at our disposal. Therefore, we slightly adapted our procedure. We maintained lexical meaning as a central component and also used verbs expressing typical football actions. Furthermore, we made use of collocations (see Lehecka, 2015). We retrieved the relevant hits and their context from the corpora. Furthermore, we augmented the context as much as possible by using queries with strings from the preceding or the following context, either within the database or on the general internet.

The newspaper data investigated with respect to French stem from the two databases Emolex and Sketch Engine. We collected relatively long strings of context as we intended to find out more about the discursive functioning of the narrative imperfect. More specifically, our aim was three-fold. First, we collected sequential data in order to analyze the discourse context. In this respect, we investigated, for instance, the positioning of the form in question within a paragraph and what other tense-aspect forms it co-occurs with. Second, we analyzed the interaction of different imperfect uses in co-occurring contexts. Third, we were interested in the pervasiveness of the form in the text type at hand. While some of the insights from the first two aims are presented in Egetenmeyer (In press), the last issue is most relevant for the present paper.

We began with an analysis of the newspaper data contained in Emolex (see Diwersy et al., 2014) (the original URL, http://emolex.u-grenoble3.fr, is no longer available and has been changed to http://phraseotext.univ-grenoble-alpes.fr/emoBase/, accessed: June 16, 2021). It is a monolingual press corpus containing data from national and regional newspapers from the years 2007–2008. Kern and Grutschus (2014, p. 188) illustrate the two groups with *Le Monde* from 2008 and *Libération* from 2007, contrasted with *Ouest-France* from 2007 to 2008. In total, the corpus contains over 112 million words from close to 300,000 texts (see http://phraseotext.univ-grenoble-alpes.fr/emoConc/emoConc.new.php). Its data coverage is thus especially good. An important advantage of the corpus is that it contains the

TABLE 2 | Aktionsart distribution of the German hits.

Aktionsart	Verb in protasis	Verb in apodosis
Achievement	11	2
Accomplishment	1	4
Activity	—	1
State	—	5
Total	**12**	**12**

whole articles and does not inhibit their retrieval piece by piece in their entirety.

We conducted two different studies. The first one concerned nine verbal types or verbal phrases, where the verbs involved were marked for third person singular, *imparfait d'indicatif*. In terms of lexical context, the verbs were likely to occur in a football context. Among them were *centrer* ("to center"), *dévier* ("to deflect") and the collocation *marquer* + *but* ("to score + goal"). As *centrer* ("to center") presented a considerably higher number of instances than the other verbs used in the queries, we restricted our analysis in this case to the first 50 occurrences with football content. We found relevant instances for five of the verbs and retrieved forty-four relevant examples from a total of forty-two different newspaper articles. As indicated above, we were able to retrieve the whole texts in all cases.

Interestingly, the verb choice in combination with the inflection yielded a low percentage of noise in this corpus study. Approximately 83% of all hits occurred in a football context. Among these hits, the verbs contained in our queries may be classified in nearly 42% of cases as narrative uses of the imperfect in the sense defined in Egetenmeyer (In press) for football language. That is, they were part of strings of verbs marked by the *imparfait* which expressed sequences of events. Other instances showed a descriptive or background reading and the like, and were therefore not included in our data collection.

In order to test whether we could replicate our successful first study with another database, we issued further queries in the database Sketch Engine, and more specifically, within the sub-corpus "Timestamped JSI web corpus 2014–2017 French." The corpus is very large. Therefore, we restricted the data to the year 2016, leaving us with over one billion tokens. We looked for the two verbs *contrer* ("to counter") and *dribbler* ("to dribble"/"to pass someone dribbling") and the verbal collocation *marquer* + *but* ("to score + goal"), whose components had to occur within a range of five words. Due to the high amounts of hits in all queries, we limited our focus to the first fifty items with a football context and where the noun *but* occurred in the singular. Thereby, we retrieved thirty relevant instances from twenty-nine different texts. In twenty-two cases we were also able to retrieve the entire article, either from the database or from the free internet, using queries in Google. Again, we analyzed the data retrieved in more depth.

RESULTS

In the previous section we discussed the ways in which we collected data on the linguistic phenomena presented in *Theoretical Background*. These phenomena are formally different, although they are based on a similar conceptual principle. As the phenomena pertain to different languages, different databases and corpora had to be used. The databases differ with respect to content-related specificity and with regard to the corpus query language, that is, what kinds of queries are possible. Therefore, we adapted our queries and we were able to retrieve relevant data in all cases. We then analyzed the data retrieved in depth.

It is important to recall the paramount aim of the contribution. We intend to collect evidence for the role genre plays in building up predictions for TAM marking. To achieve this goal, we need to understand the phenomena better. This guides the presentation of the results. In the following two sub-sections, we make reference to both German and French. Furthermore, both sub-sections combine quantitative and qualitative insights. The first sub-section considers Aktionsart properties of the verbs involved in the deviant TAM marking. The second sub-section discusses how the structures are embedded into the discourse they occur in.

Aktionsart Properties

Our interest concerns morphological marking on the verb, tense-aspect and mood marking, which deviate from a basic prescriptive marking. As indicated in *Commonalities: Football Frame and Shifted Perspective Time*, the phenomena in both languages show an important parallel in their temporal anchoring, resulting from the temporal perspective involved. Therefore, we can determine the most direct and pervasive potential factor of the forms analyzed as being the part of the verbs' lexical meaning which is relevant in terms of temporal structure, namely, the Aktionsart of the verbs.

As part of the queries, we controlled for the Aktionsart properties. In the German queries, all verbs we searched for were telic in the reading in question. The examples we retrieved consist of eleven achievements and one accomplishment. In addition, the apodosis of the structure also contains a verb. There, we found a certain variance in terms of Aktionsart. The data retrieved show two achievements, four accomplishments, one activity and five states in the apodosis (see **Table 2**). This variance is also relevant for the relationships holding within the structures (see *Discursive Properties*).

Although in our French queries one verb was atelic, we were only able to retrieve instances with telic verbs. In the most extensive study, the one realized with Emolex, we were able to retrieve examples containing five different verbs of which, in their typical reading, three express achievements, one an

TABLE 3 | Aktionsart distribution of the French hits in Emolex.

Aktionsart Properties	Verbs	Relevant hits
Achievement	*centrer* ("to center")	19
	dévier ("to deflect")	9
	marquer + *but* ("to score + goal")	8
Accomplishment	*dribbler* ("to dribble")	4
Ingressive process	*contrer* ("to counter")	4
Total		**44**

TABLE 4 | Aktionsart distribution of *imparfait* verbs in the 25 texts with the highest proportion of *imparfait* verbs (data from Emolex).

Aktionsart	Sum	Proportion
Achievement	227	45.49%
Accomplishment	69	13.83%
Ingressive process	20	4.01%
(Subtotal telic verbs)	(316)	(63.33%)
Activity	55	11.02%
State	128	25.65%
(Subtotal atelic verbs)	(183)	(36.67%)
Total	**499**	**100%**

accomplishment and one an ingressive process. Of the achievement verbs, *centrer* ("to center") yielded most relevant examples (19), followed by *dévier* ("to deflect") (9) and *marquer* + *but* ("to score + goal") (8). The queries with the accomplishment verb *dribbler* ("to dribble" in the sense of "to pass someone by dribbling") and the ingressive process *contrer* ("to counter") both yielded four relevant occurrences each (see **Table 3**).

As noted in *Collection of French Data in Non-Specialized Corpora*, in this study, we retrieved whole articles and analyzed them further. The forty-four hits pertained to forty-two articles. From these forty-two articles, we selected those where at least 75% of all inflected verbs, not counting direct discourse if it occurred, were marked by the *imparfait*. In these cases, the *imparfait* forms typically occur in long sequences in which other forms intervene only very rarely. The non-imperfective forms rather tend to occur at the beginning or at the end of the articles. This augments the probability of narrative uses in sequence. The data set contained two articles which did not involve any other finite form than the *imparfait* and a further twenty-three articles showed 75% or more *imparfait* markings. Four further articles were close to the threshold, but we excluded them from the further step, together with the articles with fewer *imparfait* verbs than this. As part of this step, we analyzed the Aktionsart of all the *imparfait* verbs contained in the subset of twenty-five articles. In cases of doubt on the classification of the Aktionsart, we took a bearing on Lehmann (1991). We thereby intended to test the finding of Bres (1999, p. 5) that narrative imperfects frequently occur with lexically bounded verbs. However, it is important to note that not all of these *imparfait* verbs show a narrative use. Although a very high share actually occurs in chains of *imparfait* verbs, some may express intervening descriptions, fulfilling the function of stage setting or other non-narrative functions. Habituals are also possible. However, as the sheer proportion of verbs marked by the *imparfait* is unprecedented in French texts, it is reasonable to maintain this broad focus and not to select specific uses at this point.

In the twenty-five texts identified, we found a total of 599 inflected verbs. Four hundred and ninety-nine of these verbs are marked by the *imparfait* (83.3%). **Table 4** gives an overview of the distribution of *imparfait* verbs. Of these verbs, 316 are telic (63.3%) and 183 are atelic (36.7%). By far the largest group is that of achievement verbs, with 227 (45.5% of all verbs marked by the *imparfait*). This is precisely what we may expect from a football report, namely that events in sequence are narrated. The group with the second largest share is that of states, with 128 verbs (25.7%). Again, given the way we counted the verbs, this may be expected. Many of these verbs contribute background information against which the importance of the narrated events may be evaluated. Furthermore, there are 69 accomplishment verbs, 20 verbs expressing an ingressive process, which we also counted as telic, and 55 activity verbs.

As noted with respect to the data coming from Sketch Engine, we were not able to retrieve the entire article in all cases. Therefore, we did not repeat the extensive study of the Aktionsart of all *imparfait* verbs in the articles, as we did with the data retrieved from Emolex. However, we analyzed the data with respect to preceding, following and intervening non-*imparfait* tense-aspect forms (see *Discursive Properties*).

Discursive Properties

In the preceding sub-section, we have already presented insights from the French articles which we retrieved as a whole. This was a first glance with respect to the more global perspective. However, the present subsection is dedicated to insights concerning the relational structure of the forms within their co-text. There are four main discursive properties which we derived from the data. With respect to German, we analyzed the function that the structure plays in the discourse context and the rhetorical relation which holds between its components. With regard to French, we investigated what tense-aspect forms occurred before and after long strings of verbs marked by the *imparfait*. Furthermore, we examined whether another form intervened and if so, which form that was. These insights may also be understood as indications of the discursive anchoring of the structures in question.

We expected the German present indicative substituting a pluperfect subjunctive to occur in oral or close-to-speech varieties. Interestingly, none of the cases we found in the corpora with specific football language are direct discourse (six instances), while in the non-specific newspaper corpora all six hits occur as part of direct speech. As noted with respect to the specialized corpora, we retrieved all six instances from corpora presenting live ticker data ("KIC," "SID") (see also **Table 1** in *Collection of German Data in Specialized and Non-Specialized Corpora* for the distribution), which is a variety with characteristics of a language of proximity. Thus, the distribution of the data is as expected.

Furthermore, we analyzed the rhetorical relation between the two clauses comprising the German structure (see Kehler, 2002; Asher and Lascarides, 2003 and others). This is summarized in **Table 5**. Interestingly, relationships involving temporal sequentiality predominate. Five instances show a strict

TABLE 5 | Rhetorical relations in the German data.

Rhetorical relations	Numbers
Occasion	5
Narration	5
Result	1
Elaboration	1
Total	**12**

contiguity between the two eventualities expressed, which may be best classified as cases of OCCASION in terms of Kehler (2011, p. 1970) (see example 11). And, if we were to make such a fine-grained distinction, five further instances express a less direct temporal sequence that may be classified as NARRATION in the sense of Asher and Lascarides (2003, p. 162) (see example 12). One example shows a RESULT relation (see example 13) and one an ELABORATION relation (see example 14).

11) *Der Peruaner kommt an den Fünfer gerauscht, verpasst den Ball aber ganz knapp, der durch seine Beine geht. Wenn er den* ***trifft****, dann* ***zappelt*** *das Leder auch im Netz.* (Cosmas II: SID/B16.00096).
‘The Peruvian rushes to the goal area, but just misses the ball, which passes through his legs. If he had hit (lit.: hits) it, the leather would have wriggled (lit.: wriggles) in the net.’
12) [1] *Nach 70 Minuten hätte Nhu-Phan Nguyen den VfL erneut in Führung bringen können,* [2] *scheiterte jedoch [. . .] an FCB-Torhüter Kevin Kraus.* [3] “*Womöglich war das der Knackpunkt.* [4] ***Verwandelt*** *er,* [5] *dann* ***gewinnen*** *wir das Spiel,* [6] *da bin ich mir sicher”,* [7] *erklärte Brunetto* (Cosmas II: KSA14/SEP.08133).
‘[1] After 70 min, Nhu-Phan Nguyen could have given VfL the lead again, [2] but failed to beat FCB goalkeeper Kevin Kraus. [3] “Possibly that was the crux. [4] If he had converted (lit.: converts), [5] then we would have won (lit.: win) the game, [6] I'm sure,” [7] explained Brunetto.’
13) *Alles oder nichts: Galvez steigt mit viel Risiko ins Tackling gegen Arnold ein und spielt den Ball.* ***Trifft*** *er das Leder nicht,* ***muss*** *er wohl frühzeitig zum Duschen.* (Cosmas II: KIC/B15.00264)
‘All or nothing: Galvez makes a risky tackle on Arnold and plays the ball. If he had (lit.: does) not hit the leather, he would probably have had (lit.: probably has) to take an early shower.’
14) [1] *Schäfer faustet einen Groß-Freistoß genau vor die Füße von Lex,* [2] *der aus etwa 16 Metern Maß nimmt* [3] *und mit Gewalt drauf hält.* [4] ***Geht*** *er rein,* [5] ***schießt*** *er das Tor des Monats* – [6] *doch so geht die Kugel über den Querbalken*. (Cosmas II: SID/Z15.00097)
‘[1] With his fist, Schäfer diverts a free kick by Groß right at the feet of Lex, [2] who takes aim from about 16 meters [3] and shoots forcefully. [4] If the ball had gone (lit.: goes) in, [5] he had scored (lit.: scores) the goal of the month – [6] but this way the ball flies over the crossbar.’

While example (12) shows direct discourse (see also the examples in *German Tense-Aspect-Mood Forms: Present Indicative Substituting Pluperfect Subjunctive*), examples (11), (13) and (14) pertain to the written medium. Live tickers have a relatively strong tendency towards the present tense (see Hennig, 2000, p. 62–63 for a similar preference in live football reports). This is due to the way content is presented, namely, the utterance is supposed to be realized (or somewhat more realistically, to start) at the very moment the events referred to are seen. However, as examples (11) and (13) show, the conditional clauses refer to situations which might have become the case at a previous point in time (*den trifft*, “hits it”; *trifft das Leder nicht* “does not hit the leather”) but, as the speaker knows at the moment of speech, were not realized in that way (*verpasst den Ball*, “misses the ball”; *spielt den Ball*, “hits the ball”). The conditional clause is thus counterfactual. Similarly, in example (12), the interviewee (Brunetto) refers to a specific past moment described in clauses [1] and [2]. He considers that at a time point anterior to this, a win would still have had been possible. However, the condition of a goal (sentence [4]) was not met (see sentence [2]).

Example (14) is a bit more subtle and underlines what we argued for in *Commonalities: Football Frame and Shifted Perspective Time*, namely, the conceptual proximity between the true futurate present tense and the present indicative taking the place of a pluperfect subjunctive. Again, the live ticker seems to present the information as if it were in objective real-time. At the time of the shot, the goal is still possible. Thus, uttered strictly at this time, a present tense form may express potentialis or future reference. However, even if we abstract away from the fact that an utterance (in written form) is impossible in objective real-time (perception and taking notes simply takes too much time) and take the relevant temporal measurement to be a subjective time flow, we may still classify the second sentence of the example ([4], [5]) as irrealis. This is so because of the evaluation in clause [5] (*schießt er das Tor des Monats*, “he scores the goal of the month”), which can only be ascribed after the realization of the shot. And at that point in time, it is already known to the speaker that the attempt has been in vain.

In the French data, we analyzed the tense-aspect forms occurring before the sequences of verbs marked by the *imparfait*. This part of the study included all examples retrieved. We analyzed both the data from Emolex and the data from Sketch Engine. The sequences of verbs, typically one or more paragraphs, consisted of or at least contained the narrative uses of the *imparfait* which had been the focus of the data collection.

As noted, in the data retrieved from Emolex, in two cases two relevant verbs are contained in the same article. As, in both of these cases, the two verbs also pertain to the same sequence of *imparfait* verbs, we do not count two separate instances (which would amount to four); this leaves us with 42 relevant texts. **Table 6** presents the frequencies of the different TAM forms. Twelve cases actually show an *imparfait* as first verb. Most frequently, the preceding verb shows a past tense with indicative mood (23 cases, 54.8% of all cases and 76.7% of the cases showing a verb other than an *imparfait* before the sequence). This may be expected. However, within this group the distribution does not adhere to a clear principle. There are eight verbs marked by the *plus-que-parfait* (the pluperfect), six marked by the *passé simple* (simple past marked for perfectivity) and nine marked by the *passé compose* (compound past). The remaining preceding verbs comprise four verbs marked by the present tense, one by the present conditional and two by the past conditional. With respect to the tense-aspect forms following the sequence of verbs marked by the *imparfait*, there is no clear tendency. The largest share is composed of the cases where no verb with a different tense-aspect form follows (19 cases). Apart

TABLE 6 | TAM forms related to the *imparfait* chains in Emolex (for a total of 42 texts).

	None/ only imparfait	Plus-que-parfait	Passé simple	Passé composé	Présent	Futur simple	Futur antérieur	Cond. I	Cond. II	Subj. présent
Preceding TAM form	12	8	6	9	4	0	0	1	2	0
First intervening TAM form	26	5	0	0	4	0	1	0	5	1
Following TAM form	19	3	4	4	9	1	0	0	2	0

TABLE 7 | TAM forms related to the *imparfait* chains in Sketch Engine (for a total of 29 texts).

	None/only imparfait	Plus-que-parfait	Passé simple	Passé composé	Présent	Futur simple
Preceding TAM form	6	2	2	7	11	1
First intervening TAM form	15	5	0	1	7	1
Following TAM form	7	2	1	9	9	1

from this group, there are three verbs marked by the pluperfect, four marked by the *passé simple* and another four by the *passé composé*. Nine verbs are marked by the present tense, two by the past conditional and one by the simple future. Finally, we analyzed the persistence of the *imparfait* within the largest sequence of *imparfait* forms in the articles. In our data from the Emolex corpus, the *imparfait* was predominantly persistent, as 26 "chains" of *imparfait* verbs were not interrupted at all (61.9%). In the 16 cases with an interruption (38.1%), we analyzed the verb forms intervening in the chains. More specifically, we focused on the first interrupting verb form. The variance is conspicuously reduced. There are only five different TAM forms, of which only three occur more often than once. Five verbs are marked by the pluperfect, four by the present tense, five by the past conditional and one each by the *futur antérieur* (future perfect) and the present subjunctive. In this data set, the interruption of the chain of *imparfait* verbs is mainly realized by only one verb. By contrast, there are only three cases where the interruption comprises more than one verb, and in one further case, the chain of *imparfait* verbs is interrupted twice. As stated, we only counted the first intervening tense-aspect form. Furthermore, it is interesting to note that many of the intervening tense-aspect forms are typical indicators of a perspective shift, namely, the pluperfect and the present tense (see Becker et al., 2021). This is another indication favoring our hypothesis of a special perspective in the case of the footballer's *imparfait* (see *Commonalities: Football Frame and Shifted Perspective Time*), as interruptions of the chains correlate with a shift in perspective (see also Sthioul, 2000 for the interplay of tense-aspect forms and perspective taking).

The linguistic material which we retrieved from Sketch Engine of 29 texts partly confirms our findings from the above-described set of data. As noted, we were only able to collect the entire articles in 22 cases, which needs to be taken into consideration in the analysis and the results. We repeated our analysis of tense-aspect forms preceding and following the main sequences of *imparfait* verbs ("chains"). The frequencies of the various TAM forms are presented in **Table 7**. With respect to the preceding tense-aspect forms, the tendency found in Emolex of a predominance of past tense-aspect forms did not repeat itself. Six articles start with the *imparfait* immediately. In one of these cases, however, the preceding paragraph is missing. Of the remaining 23 instances, less than half show a past tense (11 cases, 47.8%), so that non-past tenses are slightly dominant (12 instances, 52.2%). The first sub-set consists of two verbs marked by the pluperfect, two by the *passé simple* and seven by the compound past. In the second sub-set, the present tense dominates with 11 forms. The remaining verb is marked by the simple future. Similarly to the data from Emolex, the data from Sketch Engine also lacks a clear tendency in the realm of the tense-aspect form following the *imparfait* chain. However, the variance is reduced. In this case, seven instances lack a following verb not marked by the *imparfait*. However, in three of these cases a following paragraph is missing. Of the remaining 22 verbs, 12 are marked for past tense (54.5%) and ten show a non-past tense-aspect marking (45.5%). More specifically, two verbs are marked by the pluperfect, one is marked by the *passé simple* and nine are marked for the compound past. Again, the present tense is predominant in the non-past array as nine verbs show a present indicative marking and only one is marked by the simple future. Relatively speaking, the *imparfait* chain is interrupted more frequently in this data set, namely in 14 cases (48.28%), as opposed to 15 examples without interruption (51.72%). Interestingly enough, among the interruption cases, the two tense-aspect forms with high perspective shifting potential, the pluperfect and the present tense, are again predominant. The pluperfect occurs five times and the present tense seven times. Furthermore, there is one instance of a compound past and one case of simple future. However, it is important to point out a restriction on these last results. In ten articles, the *imparfait* chain is interrupted more than once. We analyzed only the first intervening verb. In contrast to the first data set, the first interruption always consists of only one non-*imparfait* verb.

DISCUSSION

In the paper at hand, we reported on several corpus studies of football language concerning TAM forms which differ from the expected standard forms. In German football talk, the present indicative may replace the pluperfect subjunctive. In French football reports, the imperfective past may replace the perfective past. On the basis of our results, we are able to address the basic research questions listed in the *Introduction*. 1) With regard to frequency and distribution, we collected evidence for the established status of the uses as they occurred in different corpora. The French TAM form appears to be more frequent than the German counterpart. The French footballer's *imparfait* can be found across longer strings of text. In addition, the German TAM deviation has been shown to occur in oral and close-to-speech varieties, while the French one pertains to written language. As the German phenomenon had not been researched previously, we also included a counter-study which indicates that it does not extend beyond football (or sports-related) language. 2) The grammaticalization status of these phenomena was underlined in terms of their lexical variability and also with regard to their discourse semantic variability. For both languages, we showed that a wide range of Aktionsart features is compatible with the phenomena, although only telic verbs were used in the German corpus queries. When analyzing the whole newspaper articles (or the retrievable part), French also displayed a predominance of telic verbs. This is in accordance with Bres (1999, p. 5) and others who analyze the literary use of the narrative imperfect (see also below). Furthermore, for German, we analyzed the rhetorical relations within the structure, which also show a certain variability. The detailed analyses were necessary as the previous research literature is still poor in at least the following two regards. First, it does not seem to have accounted for the German present indicative replacing the pluperfect subjunctive in football talk at all. Second, with respect to French football language, the literature is focused on relatively short excerpts or discusses the distribution of tense-aspect forms in football reports in general (see for instance work by Labeau, 2004; Labeau, 2007). By contrast, we included longer strings and also addressed entire football articles as a whole. Furthermore, the combined analysis of the two phenomena is new.

The other research questions go beyond the quantitative analysis of the data. (iii) Why do speakers use such deviant TAM forms and, relatedly, (iv) what linguistic and non-linguistic clues serve as coping mechanisms on the part of the comprehender? As discussed in *Commonalities: Football Frame and Shifted Perspective Time*, there are good arguments that the forms analyzed function on the grounds of a perspective shift. We can only hypothesize about the reasons why speakers do so. However, given the content referred to, we may assume that the speaker has the intention to induce interest in the listener. Therefore, it is probable that the speaker intends to convey a conceptual proximity with respect to the content and operationalizes the immediacy of the experience (see Walker, 2008, p. 300 for arguments against a so-called hot news reading of the present perfect in English football language). Importantly, as we saw in *Commonalities: Football Frame and Shifted Perspective Time*, this "flexibility" in the use of TAM forms seems to be licensed by the linguistic and extra-linguistic properties of the genre. More specifically, the most central subject, football matches, adheres to a specific script or frame with typical events following a well-entrenched course of action and which are in accordance with a set of rules. Furthermore, the events may be determined temporally in terms of objective time. The speaker may assume that the comprehender has all this knowledge. Therefore, the temporal (French) and world reference indices (German) may easily be retrieved contextually. As a consequence, the speaker may resort to secondary functions of the TAM forms. Apart from scripts and frames and objective time, the comprehender may rely on linguistic means to decipher the message. This is especially relevant in French, where the lack of perfectivity is, at least partly, compensated for by lexical boundedness (see also Bres, 1999, p. 5). We saw that the deviation in French may occur across whole newspaper articles. By contrast, due to its properties, the German phenomenon seems to be restricted mainly to referring to single events in the extra-linguistic world. It should be noted that our data indicates a very specific quality of these events. The events often seem to be decisive for the match in question. Thus, they are prominent within the discourse situation.

Finally, the paramount research question was (v) what insights can be gained from the phenomena in predictive language processing? We described the deviant TAM forms in two languages as a function of genre. As we have seen, the specific properties of the genre license the deviant uses in the first place and also convey the necessary parameters to decipher the intended message. Schumacher and Avrutin (2011) present a similar approach, although the phenomenon they discuss pertains to the realm of reference to individuals and not to events as in our case. Thus, the case of deviant TAM forms in German and French football language shows the extensive predictive potential of genre.

To conclude, it should be mentioned again that our studies were concerned with corpus data. They allowed us to explore the whole range of properties of the phenomena discussed. However, in a further step, it should be investigated whether our insights and our hypotheses concerning processing bear closer examination. Therefore, they should be tested with experimental methods. Apart from a more general analysis, following, for instance, the example by Schumacher and Avrutin (2011) or Squires (2019), in order to confirm the role of genre in the processing of the deviant TAM forms, it would be especially interesting to test the role of the linguistic properties we determined. This might concern the preference for telic Aktionsart categories in French or, for example, the range of rhetorical relations in the German structures. The findings would also be of relevance for the functioning of these categories beyond the language of football. Furthermore, the description of the German present indicative replacing the pluperfect subjunctive should be broadened. Of special interest is the grammaticalization path. It could be tested whether genericity might have functioned as a gateway, as this was a factor that led to some confusion for our assistants. Furthermore, in similar respects, it may be enlightening to compare non-broadcast interview data with its written presentation in newspapers in order to examine the frequency with which the footballer's present is "corrected" by means of a pluperfect subjunctive, and which newspapers do so. This might answer the question of why we did not find any examples in the sub-corpus of the football journal *Kicker* ("KSP"). Finally, the analysis of the German TAM use in football language should be extended to other structures. As it seems, there are further interesting deviations to be detected.

CORPUS

Online Databases

Cosmas II – Corpus Search, Management and Analysis System of the Leibniz-Institut für Deutsche Sprache (IDS). Available at: https://www.ids-mannheim.de/cosmas2/ (Accessed: June 16, 2021).

EMOLEX. Le lexique des émotions dans cinq langues européennes: sémantique, syntaxique et dimension discursive. ANR Franco-allemand édition 2009(ANR-09-FASH 017). Original website: http://emolex.u-grenoble3.fr/. Now available at: http://phraseotext.univ-grenoble-alpes.fr/emoBase (Accessed June 16, 2021).

Google. Available at: www.google.de (Accessed June 16, 2021).

Timestamped JSI web corpus 2014–2017 French. Created from Jozef Stefan Institute Newsfeed. Available at: https://www.sketchengine.co.uk/jozef-stefan-institute-newsfeed-corpus/ (Accessed June 16, 2021). (The corpus is part of the database Sketch Engine. Available at: https://www.sketchengine.co.uk (Accessed June 16, 2021)).

Further Data Cited in This Contribution

Aymé, M. (1968). Le passe-muraille, Les Sabines. Paris: Poche

FR 1 = Köln – Mainz. Lustgewinn im Hexenkessel [08.04.18]. Available at: https://www.fr.de/sport/fsv-mainz-05/lustgewinn-hexenkessel-11040097.html (Accessed June 16, 2021).

Spiegel 1 = Aussortierter Weltmeister Hummels zeigt jetzt Verständnis für Bundestrainer Löw [27.03.2019]. Available at: https://www.spiegel.de/sport/fussball/mats-hummels-zeigt-verstaendnis-fuer-bundestrainer-joachim-loew-a-1259864.html (Accessed June 16, 2021).

ZDF 1 = ZDF Sportstudio Live. UEFA EM 2020. Gruppe F: Deutschland – Ungarn – live. Third minute of additional time [23.06.2021]. Commentator: Oliver Schmidt. Available at: https://www.zdf.de/sport/fussball-em/deutschland-ungarn-spiel-36-livestream-100.html (Accessed June 23, 2021).

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

FUNDING

The research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) as part of the Collaborative Research Center SFB 1252 *Prominence in Language*, Project-ID 281511265, which is gratefully acknowledged.

ACKNOWLEDGMENTS

I thank Mary Chambers for proof-reading, and Barbara Zeyer, Laila López Armbruster and Vanessa Reibaldi for practical assistance. I thank Franziska Kretzschmar for an open ear and input concerning an experimental view on the problem which, however, I have not pursued at this point, and also for her support as my corresponding editor at the revision stage. I thank Olivier Kraif for explaining to me how Emolex can now be accessed. I thank the audience at a talk in Neuchâtel in June 2018 at the conference Chronos 13 for their feedback concerning football language. I thank Martin Becker, Mark Ellison, Klaus von Heusinger, and Petra Schumacher for their input following a talk on some of the data from this paper given in Cologne in June 2021. I thank my football-interested friends for discussing with me the "strange verbal inflections" used by German commentators on live football matches on national television. Finally, I thank three reviewers for their insightful remarks and constructive criticism. One of them suggested the term "immediacy of the experience" (see *Commonalities: Football Frame and Shifted Perspective Time*). The first draft of this paper was published on a preprint server: Egetenmeyer, J. (2021). Genre determining prediction: Non-standard TAM marking in football language. *arXiv* 2106.15872 [cs.CL]. Available at: https://arxiv.org/pdf/2106.15872.pdf (Accessed June 30, 2021).

REFERENCES

Altmann, G. T. M., and Kamide, Y. (1999). Incremental Interpretation at Verbs: Restricting the Domain of Subsequent Reference. *Cognition* 73 (3), 247–264. doi:10.1016/S0010-0277(99)00059-1

Altmann, G. T. M., and Kamide, Y. (2007). The Real-Time Mediation of Visual Attention by Language and World Knowledge: Linking Anticipatory (And Other) Eye Movements to Linguistic Processing. *J. Mem. Lang.* 57 (4), 502–518. doi:10.1016/j.jml.2006.12.004

Anderson, J. R. (1990). *The Adaptive Character of Thought*. Hillsdale: Lawrence Erlbaum Associates.

Arai, M., and Keller, F. (2013). The Use of Verb-specific Information for Prediction in Sentence Processing. *Lang. Cogn. Process.* 28 (4), 525–560. Available at: https://www.research.ed.ac.uk/en/publications/the-use-of-verb-specific-information-for-prediction-in-sentence-p (Accessed June 16, 2021).

Asher, N., and Lascarides, A. (2003). *Logics of Conversation*. Cambridge: Cambridge University Press.

Banfield, A. (1982). *Unspeakable Sentences: Narration and Representation in the Language of Fiction*. Boston, MA: Routledge and Paul.

Becker, M. (2014). *Welten in Sprache. Zur Entwicklung der Kategorie "Modus" in romanischen Sprachen*. Berlin; New York: De Gruyter.

Becker, M., and Egetenmeyer, J. (2018). A Prominence-Based Account of Temporal Discourse Structure. *Lingua* 214, 28–58. doi:10.1016/j.lingua.2018.08.002

Becker, M., Egetenmeyer, J., and Hinterwimmer, S. (2021). Perspective in German and French: Divergences in Formal Marking and Temporal Anchoring. *Z. Literaturwiss Linguistik* 51 (2), 321–356. doi:10.1007/s41244-021-00204-4

Berthonneau, A.-M., and Kleiber, G. (1999). Pour une réanalyse de l'imparfait de rupture dans le cadre de l'hypothèse anaphorique méronomique. *Cahiers de Praxématique* 31, 1–30. Available at: http://praxematique.revues.org/2853 (Accessed June 16, 2021).

Bres, J. (1999). L'imparfait dit narratif tel qu'en lui-même (le cotexte ne le change pas). *Cahiers de Praxematique* 32, 87–117. doi:10.4000/praxematique.2852

Bres, J. (2005). *L'imparfait dit narratif*. Paris: CNRS Éditions.

Burkhardt, A. (2006). Sprache und Fußball. Linguistische Annäherung an ein Massenphänomen. *Muttersprache* 116, 53–73.

Caudal, P. (Submitted). On So-Called 'Tense Uses' in French as Context-Sensitive Constructions.

de Saussure, L., and Sthioul, B. (1999). L'imparfait narratif: point de vue (et images du monde). *Cahiers de Praxématique* 32, 1–14. Available at: http://praxematique.revues.org/2855 (Accessed June 30, 2021).

De Swart, H. (2007). A Cross-Linguistic Discourse Analysis of the Perfect. *J. Pragmatics* 39 (12), 2273–2307. doi:10.1016/j.pragma.2006.11.006

Declerck, R. (2011). "The Definition of Modality," in *Cognitive Approaches to Tense, Aspect, and Epistemic Modality*. Editors A. Patard and F. Brisard (Amsterdam; PA: John Benjamins Publishing Company), 21–44. doi:10.1075/hcp.29.05dec

Dery, J. E., and Koenig, J.-P. (2015). A Narrative-Expectation-Based Approach to Temporal Update in Discourse Comprehension. *Discourse Processes.* 52 (7), 559–584. doi:10.1080/0163853X.2014.966293

Diwersy, S., Goossens, V., Grutschus, A., Kern, B., Kraif, O., Melnikova, E., et al. (2014). Traitement des lexies d'émotion dans les corpus et les applications d'EmoBase. *Corpus.* 13, 269–293. doi:10.4000/corpus.2537

Duden (2009). *Duden. Die Grammatik. Unentbehrlich für richtiges Deutsch. 8., überarbeitete Auflage, Bd. 4, herausgegeben von der Dudenredaktion*. Mannheim; Zürich: Dudenverlag.

Eckardt, R. (2014). *The Semantics of Free Indirect Discourse. How Texts Allow Us to Mind-Read and Eavesdrop*. Leiden; Boston: Brill.

Egetenmeyer, J. (2021). Genre Determining Prediction: Non-standard TAM Marking in Football Language. *arXiv*. Available at: https://arxiv.org/pdf/2106.15872.pdf (Accessed June 30, 2021).

Egetenmeyer, J. (In press). "Time Updating Uses of the French Imparfait Extending across Genres," in *Aspects of Tenses, Modality, and Evidentiality*. Editors L. Baranzini and L. de Saussure (Leiden; Boston: Brill Rodopi).

Ehrlich, S. (1990). *Point of View. A Linguistic Analysis of Literary Style*. London; New York: Routledge.

Engel, D. M., and Labeau, E. (2005). Il etait une fois un match de foot : l'événement sportif comme objet de narration. *Revue Romane* 40 (2), 199–218. doi:10.1111/j.1600-0811.2005.00031.x

Escandell-Vidal, M. V. (Submitted). Understanding the Spanish Imperfecto. Temporal Dimensions and Evidential Effects.

Fabricius-Hansen, C. (1999). "Moody time": Indikativ und Konjunktiv im deutschen Tempussystem. *Z. Literaturwiss Linguistik* 29 (1), 119–146. doi:10.1007/BF03379172

Facques, B. (2002). "Passé composé, imparfait et présent dans les récits journalistiques : des alternances aux ruptures temporelles," in *Les temps du passé français et leur enseignement*. Editors E. Labeau and P. Larrivée (Amsterdam; New York: Rodopi), 105–133.

Fillmore, C. J. (1977). "The Case for Case Reopened," in *Syntax And Semantics 8*. Editors P. Cole and J. M. Sadock (New York: Academic Press), 59–81. Available at: http://www.icsi.berkeley.edu/pubs/ai/casefor277.pdf (Accessed June 16, 2021).

Fillmore, C. J. (2006). "Chapter 10. Frame Semantics," in *Cognitive Linguistics: Basic Readings*. Editor D. Geeraerts (Berlin; New York: De Gruyter Mouton), 373–400. doi:10.1515/9783110199901.373

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

Gallmann, Peter. (2007). "Morphologische Probleme der deutschen Konjunktive," in *Sprachliche Motivation. Zur Interdependenz von Inhalt und Ausdruck*. Editors P. Gallmann, C. Lehmann, and R. Lühr (Tübingen: Narr), 45–80.

Gosselin, L. (1999). Le sinistre Fantômas et l'imparfait narratif. *Cahiers de praxematique* 32, 19–42. doi:10.4000/praxematique.2847

Graf, T., Philipp, M., Xu, X., Kretzschmar, F., and Primus, B. (2017). The Interaction between Telicity and Agentivity: Experimental Evidence from Intransitive Verbs in German and Chinese. *Lingua* 200, 84–106. doi:10.1016/j.lingua.2017.08.006

Grevisse, B. (1997). *Le temps des journalistes: Essai de narratologie médiatique*. Louvain-la-Neuve: CIACO.

Guéron, J. (2015). "Subjectivity and Free Indirect Discourse," in *Sentence and Discourse*. Editor J. Guéron (Oxford: Oxford University Press), 256–286. doi:10.1093/acprof:oso/9780198739418.003.0012

Hanulíková, A., van Alphen, P. M., van Goch, M. M., and Weber, A. (2012). When One Person's Mistake is Another's Standard Usage: The Effect of Foreign Accent on Syntactic Processing. *J. Cogn. Neurosci.* 24 (4), 878–887. doi:10.1162/jocn_a_00103

Hennig, M. (2000). *Tempus und Temporalität in geschriebenen und gesprochenen Texten*. Tübingen: Niemeyer.

Herzog, C. (1981). *Le passé simple dans les journaux du XXe siècle*. Berne: Francke.

Hinterwimmer, S. (2017). "Two Kinds of Perspective Taking in Narrative Texts," in Proceedings of the 27th Semantics and Linguistic Theory Conference (SALT 27). Editors D. Burgdorf, J. Collard, S. Maspong, and B. Stefánsdóttir (Washington: Linguistic Society of America), 282–301. doi:10.3765/salt.v27i0.4153

Jasinskaja, K., and Karagjosova, E. (2020). "Rhetorical Relations," in *The Wiley Blackwell Companion to Semantics*. Editors D. Gutzmann, L. Matthewson, C. Meier, H. Rullmann, and T. E. Zimmerman (Oxford: Wiley-Blackwell), 1–29. doi:10.1002/9781118788516.sem061

Kamp, H., and Rohrer, C. (1983). "Tense in Texts," in *Meaning, Use, and Interpretation of Language*. Editors R. Bäuerle, C. Schwarze, and A. von Stechow (Berlin; New York: de Gruyter), 250–269.

Kehler, A. (2002). *Coherence, Reference, and the Theory of Grammar*. Stanford: CSLI Publications.

Kehler, A. (2011). "74. Cohesion and Coherence," in *Semantics: An International Handbook Of Natural Language Meaning* 2. Editors C. Maienborn, K. von Heusinger, and P. Portner (Berlin; Boston: De Gruyter Mouton), 1963–1987. doi:10.1515/9783110255072.1963

Kern, B., and Grutschus, A. (2014). "Surprise vs étonnement: Comportement Discursif et Perspectives Contrastives," in *Les émotions dans le discours. Emotions in discourse.* Editors P. Blumenthal, I. Novakova, and D. Siepmann (Frankfurt a. M.: Peter Lang), 187–198.

Koch, P., and Oesterreicher, W. (2011). *Gesprochene Sprache in der Romania: Französisch, Italienisch, Spanisch*. Berlin; New York: de Gruyter. doi:10.1515/9783110252620

Koch, P., and Oesterreicher, W. (2012). ""Language of Immediacy – Language of Distance. Orality and Literacy from the Perspective of Language Theory and Linguistic History," in *Communicative Spaces. Variation, Contact, and Change. Papers in Honour of Ursula Schaefer*. Editors C. Lange, B. Weber, and G. Wolf (Frankfurt a. M.: Peter Lang), 441–473. doi:10.15496/publikation-20415

Kuperberg, G. R. (2013). "The Proactive Comprehender: What Event-Related Potentials Tell Us about the Dynamics of Reading Comprehension," in *Unraveling The Behavioral, Neurobiological, and Genetic Components Of Reading Comprehension*. Editors B. Miller, L. Cutting, and P. McCardle (Baltimore: Paul Brookes Publishing), 176–192. Available at: https://projects.iq.harvard.edu/kuperberglab/files/kuperberglab/files/kuperberg_predictivecomprehension_chapter.pdf (Accessed June 16, 2021).

Kuperberg, G. R., and Jaeger, T. F. (2015). What Do We Mean by Prediction in Language Comprehension? *Lang. Cogn. Neurosci.* 31, (1), 139–159. doi:10.1080/23273798.2015.1102299

Labeau, E. (2004). Le(s) temps du compte rendu sportif. *French Lang. Stud.* 14 (2), 129–148. doi:10.1017/S0959269504001619

Labeau, E. (2006). French Television Talk: What Tenses for Past Time?. *Int. J. Corpus Linguist.* 11 (1), 1–28. Available at: http://publications.aston.ac.uk/id/eprint/698/1/Int._journal_of_Corpus.pdf (Accessed June 16, 2021). doi:10.1075/ijcl.11.1.02lab

Labeau, E. (2007). “Et un, ou deux, ou trois ? Les temps-champions du reportage sportif depuis 1950,” in *Sémantique et diachronie du système verbal français*. Editors E. Labeau, C. Vetters, and P. Caudal (Amsterdam: Rodopi), 203–233. doi:10.1163/9789401203838_010

Lehecka, T. (2015). “Collocation and Colligation,” in *Handbook of Pragmatics 19, 2015 Installment*. Editors J. Östman and J. Verschueren (Amsterdam; Philadelphia: John Benjamins), 1–20. Available at: https://pure.mpg.de/rest/items/item_2304786_2/component/file_2304785/content (Accessed June 16, 2021). doi:10.1075/hop.19.col2

Lehmann, C. (1991). “Predicate Classes and Participation,” in *Partizipation. Das sprachliche Erfassen von Sachverhalten*. Editors H. Seiler and W. Premper (Tübingen: Narr), 183–239.

Leirbukt, O. (2008). *Untersuchungen zur temporalen Umfunktionierung des Konjunktivs II im heutigen Deutsch*. Tübingen: Max Niemeyer.

Levy, R. (2008). Expectation-based Syntactic Comprehension. *Cognition* 106 (3), 1126–1177. doi:10.1016/j.cognition.2007.05.006

Lewis, D. (1979). Counterfactual Dependence and Time’s Arrow Noûs. *Special Issue on Counterfactuals and Laws* 13, (4), 455–476. doi:10.2307/2215339

Molendijk, A., de Swart, H., Vetters, C., Borillo, A., Bras, M., Le Draoulec, A., et al. (2004). “Meaning and Use of Past Tenses in Discourse,” in *Handbook of French Semantics*. Editors F. Corblin and H. de Swart (Stanford: CLSI-Publications), 271–308.

Palmer, F. R. (2001). *Mood and Modality*. Cambridge: Cambridge University Press.

Philipp, M., Graf, T., Kretzschmar, F., and Primus, B. (2017). Beyond Verb Meaning: Experimental Evidence for Incremental Processing of Semantic Roles and Event Structure. *Front. Psychol.* 8, 1806. doi:10.3389/fpsyg.2017.01806

Portner, P. (2009). *Modality*. Oxford: Oxford University Press.

Quintero Ramírez, S., and Carvajal Carvajal, P. I. (2017). Funciones sintáctico-textuales de tiempos verbales en notas deportivas. *Rev. Invest. Lingüíst.* 20, 221–242. Available at: https://revistas.um.es/ril/article/view/316201 (Accessed June 16, 2021).

Ritz, M.-E. A. (2010). The Perfect Crime? Illicit Uses of the Present Perfect in Australian Police Media Releases. *J. Pragmat.* 42 (12), 3400–3417. doi:10.1016/j.pragma.2010.05.003

Salkie, R. (2009). “Degrees of Modality,” in *Modality in English: Theory and Description*. Editors R. Salkie, P. Busuttil, and J. van der Auwera (Berlin: Mouton de Gruyter), 79–104. doi:10.1515/9783110213331.79

Schank, R. C., and Abelson, R. P. (1977). *Scripts, Plans, Goals and Understanding. An Inquiry into Human Knowledge Structures*. Hillsdale, NJ: Erlbaum.

Schrott, A. (2012). Die Zeiten ändern sich. Zur Verwendung des *imparfait* in narrativen Kontexten. *Romanistisches Jahrbuch* 62 (1), 137–164.

Schumacher, P. B., and Avrutin, S. (2011). Register Affects Language Comprehension: ERP Evidence from Article Omission in Newspaper Headlines. *J. Neurolinguist.* 24 (3), 304–319. doi:10.1016/j.jneuroling.2010.10.002

Sinclair, J. (2005). “Corpus and Text – Basic Principles,” in *Developing Linguistic Corpora. A Guide to Good Practice*. Editor M. Wynne (Oxford: Oxbow Books), 1–16. Available at: http://users.ox.ac.uk/~martinw/dlc/chapter1.htm (Accessed June 16, 2021).

Squires, L. (2019). Genre and Linguistic Expectation Shift: Evidence from Pop Song Lyrics. *Lang. Soc.* 48 (1), 1–30. doi:10.1017/s0047404518001112

Sthioul, B. (2000). “Passé simple, imparfait et sujet de conscience,” in *Passé et parfait*. Editors A. Carlier, V. Lagae, and C. Benninger (Amsterdam; Atlanta: Rodopi), 79–93.

Tasmowski-De Ryck, L. (1985). L’imparfait sans et avec rupture. *Langue française* 67 (3), 59–77. doi:10.3406/lfr.1985.4651

Verkuyl, H., Vet, C., Borillo, A., Bras, M., Le Draoulec, A., Molendijk, A., et al. (2004). “Tense and Aspect in Sentences,” in *Handbook of French Semantics*. Editors F. Corblin and H. de Swart (Stanford: CLSI-Publications), 233–270.

Vetters, C. (1996). *Temps, aspect et narration*. Amsterdam; Atlanta: Rodopi.

Walker, J. (2008). “The Footballer’s Perfect – Are Footballers Leading the Way?,” in *The Linguistics of Football*. Editors E. Lavric, G. Pisek, A. Skinner, and W. Stadler (Tübingen: Gunter Narr), 295–303.

Weinrich, H. (1964). *Tempus. Besprochene und erzählte Welt*. München: C. H. Beck.

Wüest, J. (1993). Textsortenabhängiger Tempusgebrauch. *Z. für französische Sprache Literatur* 103, (3) 231–243.

Zifonun, B., Hoffmann, L., and Strecker, G. (1997). *Grammatik der deutschen Sprache, Band 1*. Berlin; New York: Walter De Gruyter. doi:10.1515/9783110872163

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

frontiers in Psychology

OPINION
published: 28 September 2021
doi: 10.3389/fpsyg.2021.734024

Prediction Impairment May Explain Communication Difficulties in Autism

Iris Scholten[1]*, *Catharina A. Hartman*[2] *and Petra Hendriks*[1]

[1] *Center for Language and Cognition Groningen, University of Groningen, Groningen, Netherlands,* [2] *Department of Psychiatry, Interdisciplinary Center Psychopathology and Emotion Regulation, University of Groningen, University Medical Center Groningen, Groningen, Netherlands*

Keywords: autism spectrum disorder, language processing, predictive processing, predictions, prediction impairment, language and communication difficulties

INTRODUCTION

A central research question on autism is how the communication difficulties of autistic individuals can be explained. In this opinion paper we put forward the hypothesis that autistic individuals have problems with language because of an underlying impairment in the ability to generate and update predictions about language. Our hypothesis combines well-established findings from the past decade indicating that linguistic predictions facilitate faster language processing with recent evidence suggesting that autistic individuals show abnormalities in predictions outside the field of language. Investigating linguistic predictions in autism can help clarify the mechanisms underlying the communication difficulties of individuals with ASD. Our hypothesis subsumes earlier mechanistic explanations involving theory of mind and executive functions.

OPEN ACCESS

Edited by:
Franziska Kretzschmar,
Leibniz Institute for the German Language (IDS), Germany

Reviewed by:
David H. V. Vogel,
Institute of Neuroscience and Medicine, Germany
Fernanda Dreux M. Fernandes,
University of São Paulo, Brazil

***Correspondence:**
Iris Scholten
l.i.scholten@rug.nl

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *30 June 2021*
Accepted: *02 September 2021*
Published: *28 September 2021*

Citation:
Scholten I, Hartman CA and Hendriks P (2021) Prediction Impairment May Explain Communication Difficulties in Autism. Front. Psychol. 12:734024. doi: 10.3389/fpsyg.2021.734024

COMMUNICATION IN AUTISM

Communication difficulties are a core component of autism. Research on communication typically distinguishes between structural language (i.e., the form and meaning of words and sentences) and pragmatic language (i.e., the use of language in social situations). In the past, many researchers assumed that autistic individuals mainly had problems with pragmatic language, such as irony (e.g., Happé, 1994; Leekam and Prior, 1994) and metaphors (e.g., Happé, 1994; Martin and McDonald, 2004). Some studies showed, however, that they have difficulties with structural language too (e.g., Brynskov et al., 2017; Wittke et al., 2017). For example, autistic individuals were found to have difficulties with which-questions (Prévost et al., 2017) and object relative clauses (Durrleman et al., 2015). Communication difficulties have been attributed to, among others, problems with theory of mind (ToM) or reduced executive functioning (EF), but these explanations do not speak to language problems beyond pragmatics. Additionally, our progress in gaining knowledge about ToM and EF in autism seems to have stagnated. We see this partly as a consequence of most studies using offline tasks, which only measure the ultimate response in a task but do not provide measures of the processes leading to this particular response (e.g., different cognitive processes can lead to the same outcome regarding ToM judgements or EF responses). Moreover, heterogeneity in outcomes (i.e., not every autistic individual shows inaccurate ToM judgements or EF responses; e.g., Baez et al., 2020; Deschrijver and Palmer, 2020) cannot be well-understood if the processes leading to these responses are not studied too. Thus, rethinking the theoretical foundations beyond ToM and EF is needed to accommodate structural language problems in autism. In addition, online measures (e.g., eye movements or brain activity) linked to the corresponding offline responses will foster insight into why some autistic individuals tend to have pragmatic and structural language problems, whereas others do not.

RECENT DEVELOPMENTS

Predictions in Language

Communication is fast and full of ambiguity. Thus, comprehenders must keep up with the speed of language and at the same time determine the intended meaning of a sentence (Crocker, 2005). Generating predictions are fundamental herein. Predictions about upcoming language speed up processing and thus help comprehenders to keep up with the speed of spoken language (e.g., Corps et al., 2018, 2019; Fitz and Chang, 2019; Kochari and Flecken, 2019; Shain et al., 2020).

It has been well-established that linguistic predictions are generated when particular linguistic information is activated in language users, even before the input that carries this information becomes available (Pickering and Gambi, 2018). For example, when hearing "John wants salt and pepper on his steak" a comprehender would be highly likely to predict lexical information, i.e., the word "pepper," after hearing "salt and" before actually hearing "pepper," because "salt and pepper" is a pair of words frequently used together in a fixed order. Preactivation of linguistic information is also empirically demonstrated by, for example, eye-tracking studies in which participants heard a sentence like "The boy eats the cake" while looking at pictures. Adults and children already look at the correct picture of a cake instead of competing pictures of other objects before they actually hear the word "cake" (e.g., Altmann and Kamide, 1999; Nation et al., 2003; Borovsky et al., 2012; Mani and Huettig, 2012). Prediction of the word "cake" comes from the verb "eats" which requires an object that is edible. This shows that comprehenders use lexical information to predict upcoming language.

Comprehenders also predict upcoming language on the basis of syntactic information. For example, Lukyanenko and Fisher (2016) showed that 3-year-old children are already faster and more likely to shift their gaze from an incorrect picture displaying only one thing to a correct picture displaying multiple things upon hearing a verb that requires a plural noun (e.g., "Where *are* the cookies?") compared to a verb that is uninformative about the number of the noun ("Do you *see* the cookies?"). Thus, toddlers use agreement between the number marking (singular or plural) of a verb and a noun to predict features of an upcoming noun, resulting in a faster identification of the correct picture. The studies described above show that preactivation of lexical as well as syntactical information helps to do some of the processing ahead of time so that comprehenders can process language fast, despite the speed with which sentences are produced and despite the pervasive ambiguity of language (Pickering and Gambi, 2018).

Given the ambiguity in language and the need for fast processing, predictions can sometimes also steer comprehenders in the wrong direction. This is the case if a comprehender's initial prediction turns out to be false. For example, in a sentence like "The horse raced past the barn fell" (Bever, 1970), comprehenders will predict that "the horse" is the subject of the sentence, because there is a tendency to interpret the first noun phrase in a sentence as the subject, and subjects as agents. Subsequently, they will predict that "raced" is the sentence's main verb (referring to what the horse did) and that "past the barn" is the direction in which the horse raced. Thus, after hearing "the horse raced past the barn," comprehenders will predict that the sentence is complete. However, upon hearing the final word of the sentence (the verb "fell") comprehenders will discover that the sentence was not yet complete, and that "raced past the barn" was a specification of "the horse" (i.e., the horse that was raced past the barn). That is, comprehenders will discover that they were led up the garden path. These so-called garden path sentences thus require comprehenders to update their initial incorrect prediction by replacing it with a new prediction, to arrive at the intended interpretation of the sentence.

Predictions Are Related to ToM and EF

Autistic individuals often have difficulties with language and communication. It is conceivable that these difficulties are a consequence of problems with predictions, and that the difficulties with ToM and EF often observed in autistic persons are linked to these prediction problems. ToM enables comprehenders to deduce why a person acted in a certain way or to anticipate how a person is likely to act. In this sense, ToM tasks are inherently prediction tasks, as comprehenders need to predict the actions of a person following a certain observation. Therefore, difficulties with ToM may be caused by a reduced ability to predict another person's behavior. EF is related to predictions as well, especially in situations where an initial prediction turns out to be incorrect. In such situations, comprehenders are required to inhibit their initial but false prediction and switch to an alternative interpretation while holding all information active in their working memory. Therefore, updating an incorrect predictions requires EF, in particular, cognitive inhibition (MacLeod, 2008), cognitive flexibility (Miyake and Friedman, 2012) and working memory (Miyake and Shah, 1999). This thus shows how ToM and EF are related to generating and updating predictions. However, as said earlier, the language difficulties seen in autism are broader than the hypotheses of ToM and EF can explain individually or together. In contrast, prediction impairments during language processing may explain the pragmatic and non-pragmatic language difficulties seen in autism and subsume the hypotheses of ToM and EF. Hence, ToM and EF are related to predictive abilities, but the process of generating and updating predictions is broader and has more explanatory power.

Predictions Beyond Language

Recent studies have illustrated abnormalities in predictive abilities in autistic individuals outside the domains of language (see Cannon et al., 2021, for a recent review of empirical evidence). Neuroimaging and eye-tracking studies using tasks in which participants are presented with predictable repetitive stimuli that are infrequently interrupted by an unpredictable deviant stimulus found that autistic individuals showed an altered response compared to neurotypical controls (Jeste et al., 2015; Balsters et al., 2017; Lawson et al., 2017; Goris et al., 2018). The results of these studies suggest that autistic individuals may be less surprised when their predictions are being violated, as indicated by reduced brain responses (e.g., Lawson et al., 2017).

This finding has been taken to mean that their predictions are less strong compared to those of neurotypical individuals (Pellicano and Burr, 2012), potentially because they struggle to generate precise internal prior beliefs about the world (both social and non-social; Pellicano and Burr, 2012; Friston et al., 2013; van de Cruys et al., 2014; Lawson et al., 2017), especially in temporally demanding environments where the environment changes fast in time (e.g., Pellicano and Burr, 2012; Hohwy et al., 2016; Vogel et al., 2019). Being less surprised after encountering a violation to a prediction could, in turn, mean that it is also harder to update a prediction, because the main function of surprise is to interrupt an ongoing action and reorient attention to the new, possibly significant stimulus (Kalat, 2015). This fits well with the finding in the language domain that predictions allow for faster processing. Indeed, the abnormalities in predictive abilities within autistic individuals found outside the language domain triggered our question if these abnormalities also occur within the language domain.

OUR PROPOSAL

To explain the language and communication problems of autistic individuals, we put forward the Linguistic Prediction Impairment Hypothesis. This hypothesis states that autistic individuals show abnormalities in generating and updating predictions about language and can be seen as the linguistic version of a more general hypothesis about predictive processing differences between autistic individual and neurotypicals (see Pellicano and Burr, 2012; Sinha et al., 2014; van de Cruys et al., 2014). We hypothesize that this explains why autistic individuals:

- process language slower than their neurotypical peers (e.g., Kamio et al., 2007; Henderson et al., 2011; Bavin et al., 2014; Arunachalam and Luyster, 2018), as predictions facilitate faster language comprehension (Corps et al., 2018);
- have more problems with pragmatic language than structural language (e.g., Ludlow et al., 2017), as pragmatic language is more influenced by contextual information, making it less predictable;
- have difficulties with structural language nonetheless (e.g., Brynskov et al., 2017; Prévost et al., 2017; Wittke et al., 2017), as predictions are generated at every level of language (sounds, words, sentences and their meanings);
- have particular problems interpreting language when the initial prediction turns out to be incorrect (as is evidenced by the findings of Durrleman et al., 2015; Martins et al., 2018; Sukenik and Friedmann, 2018; Was et al., 2018), as updating predictions requires EF which is often found to be impaired in autistic individuals;
- show individual variability in their linguistic performance (e.g., Pearson and Hodgetts, 2020), as predictive abilities may vary strongly in autistic individuals;
- tend to have difficulties with ToM tasks as well as linguistic tasks requiring speaker-hearer coordination, as predictions about other people's actions are needed to succeed in these tasks (Schuwerk et al., 2016).

CONCLUDING REMARKS

In this opinion paper, we have put forward the Linguistic Prediction Impairment Hypothesis, which states that autistic individuals have difficulties with language because they have difficulties with generating and updating predictions about language. This hypothesis could provide directions for further research and lead to new insights on language processing in autism, especially when focusing on the identification of subtle effects of predictions using methods that capture language processing online. While focused on the language and communication difficulties of autistic individuals, our proposal has broader implications. Prediction generation and prediction updating are less needed in restricted and repetitive situations. In such situations the future is more predictable and autistic individuals may therefore prefer such situations and behaviors (see also Pellicano and Burr, 2012; van de Cruys et al., 2014; Hohwy et al., 2016; Vogel et al., 2019). This allows for the integration of the DSM-5 diagnostic criteria of (A) impairments in social communication and interaction and (B) restricted, repetitive behaviors (American Psychiatric Association, 2013). Thus, our proposal would result in a more unified view of the core features of autism. Moreover, by emphasizing the importance of using online tasks measuring language processing instead of offline tasks merely registering the ultimate response, our proposal may lead to a better insight in the cognitive processes underlying linguistic behavior and may additionally increase our insights in the role of ToM and EF in generating and updating predictions about language.

AUTHOR CONTRIBUTIONS

IS wrote the first draft of the manuscript. CH and PH gave feedback on the first draft. All authors contributed to writing and revising the manuscript and read and approved the final version.

FUNDING

This research was supported by grants from the Netherlands Organization for Scientific Research NWO (PhD in the Humanities Grant no. 322-89-012, IS).

REFERENCES

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

American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders: DSM-5*. Washington, DC: American Psychiatric Association. doi: 10.26613/jca/3.2

Arunachalam, S., and Luyster, R. J. (2018). Lexical development in young children with autism spectrum disorder (ASD): How ASD may affect intake from the input. *J. Speech Lang. Hear. Res.* 61, 2659–2672. doi: 10.1044/2018_JSLHR-L-RSAUT-18-0024

Baez, A. C., Dajani, D. R., Voorhies, W., Parladé, M. V., Alessandri, M., Britton, J. C., et al. (2020). Parsing heterogeneity of executive function in typically and atypically developing children: a conceptual replication and exploration of social function. *J. Autism Dev. Disord.* 50, 707–718. doi: 10.1007/s10803-019-04290-9

Balsters, J. H., Apps, M. A. J., Bolis, D., Lehner, R., Gallagher, L., and Wenderoth, N. (2017). Disrupted prediction errors index social deficits in autism spectrum disorder. *Brain* 140, 235–246. doi: 10.1093/brain/aww287

Bavin, E. L., Kidd, E., Prendergast, L., Baker, E., Dissanayake, C., and Prior, M. (2014). Severity of autism is related to children's language processing. *Autism Res.* 7, 687–694. doi: 10.1002/aur.1410

Bever, T. G. (1970). The cognitive basis for linguistic structures. *Cogn. Dev. Lang.* 279, 1–61.

Borovsky, A., Elman, J. L., and 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. *J. Exp. Child Psychol.* 112, 417–436. doi: 10.1016/j.jecp.2012.01.005

Brynskov, C., Eigsti, I. M., Jørgensen, M., Lemcke, S., Bohn, O. S., and Krøjgaard, P. (2017). Syntax and morphology in danish-speaking children with autism spectrum disorder. *J. Autism Dev. Disord.* 47, 373–383. doi: 10.1007/s10803-016-2962-7

Cannon, J., Brien, A. M. O., Bungert, L., and Sinha, P. (2021). Prediction in autism spectrum disorder: A systematic review of empirical evidence. *Autism Res.* 13, 474–488. doi: 10.1002/aur.2482

Corps, R. E., Crossley, A., Gambi, C., and Pickering, M. J. (2018). Early preparation during turn-taking: Listeners use content predictions to determine what to say but not when to say it. *Cognition* 175, 77–95. doi: 10.1016/j.cognition.2018.01.015

Corps, R. E., Pickering, M. J., and Gambi, C. (2019). Predicting turn-ends in discourse context. *Lang. Cogn. Neurosci.* 34, 615–627. doi: 10.1080/23273798.2018.1552008

Crocker, M. W. (2005). "Retional models of comprehension: Addressing the performance paradox," in: *Twenty-First Century Psycholinguistics: Four Cornerstones* (Mahwah, NJ: Erlbaum). doi: 10.5465/ame.1998.1109048

Deschrijver, E., and Palmer, C. (2020). Reframing Social Cognition: Relational Versus Representational Mentalizing. *Psychol. Bull.* 146, 941–969. doi: 10.1037/bul0000302

Durrleman, S., Hippolyte, L., Zufferey, S., Iglesias, K., and Hadjikhani, N. (2015). Complex syntax in autism spectrum disorders: A study of relative clauses. *Int. J. Lang. Commun. Disord.* 50, 260–267. doi: 10.1111/1460-6984.12130

Fitz, H., and Chang, F. (2019). Language ERPs reflect learning through prediction error propagation. *Cogn. Psychol.* 111, 15–52. doi: 10.1016/j.cogpsych.2019.03.002

Friston, K. J., Lawson, R., and Frith, C. D. (2013). On hyperpriors and hypopriors: comment on pellicano and burr. *Trends Cogn. Sci.* 17:1. doi: 10.1016/j.tics.2012.11.003

Goris, J., Braem, S., Nijhof, A. D., Rigoni, D., Deschrijver, E., Van de Cruys, S., et al. (2018). Sensory prediction errors are less modulated by global context in autism spectrum disorder. *Biol. Psychiatry* 3, 667–674. doi: 10.1016/j.bpsc.2018.02.003

Happé, F. G. E. (1994). An advanced test of Theory of Mind: Understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *J. Autism Dev. Disord.* 24, 129–154. doi: 10.1007/BF02172093

Henderson, L. M., Clarke, P. J., and Snowling, M. J. (2011). Accessing and selecting word meaning in autism spectrum disorder. *J. Child Psychol. Psychiatry Allied Discipl.* 52, 964–973. doi: 10.1111/j.1469-7610.2011.02393.x

Hohwy, J., Paton, B., and Palmer, C. (2016). Distrusting the present. *Phenomenol. Cogn. Sci.* 15, 315–335. doi: 10.1007/s11097-015-9439-6

Jeste, S. S., Kirkham, N., Senturk, D., Hasenstab, K., Sugar, C., Kupelian, C., et al. (2015). Electrophysiological evidence of heterogeneity in visual statistical learning in young children with ASD. *Dev. Sci.* 18, 90–105. doi: 10.1111/desc.12188

Kalat, J. W. (2015). *Biological Psychology, 12th Edn*. Boston, MA: Cengage Learning.

Kamio, Y., Robins, D., Kelley, E., Swainson, B., and Fein, D. (2007). Atypical lexical/semantic processing in high-functioning autism spectrum disorders without early language delay. *J. Autism Dev. Disord.* 37, 1116–1122. doi: 10.1007/s10803-006-0254-3

Kochari, A. R., and Flecken, M. (2019). Lexical prediction in language comprehension: a replication study of grammatical gender effects in Dutch. *Lang. Cogn. Neurosci.* 34, 239–253. doi: 10.1080/23273798.2018.1524500

Lawson, R. P., Mathys, C., and Rees, G. (2017). Adults with autism overestimate the volatility of the sensory environment. *Nat. Neurosci.* 20, 1293–1299. doi: 10.1038/nn.4615

Leekam, S. R., and Prior, M. (1994). Can autistic children distinguish lies from jokes? A second look at second-order belief attribution. *J. Child Psychol. Psychiatry* 35, 901–915. doi: 10.1111/j.1469-7610.1994.tb02301.x

Ludlow, A. K., Chadwick, E., Morey, A., Edwards, R., and Gutierrez, R. (2017). An exploration of sarcasm detection in children with attention hyperactivity deficit disorder. *J. Commun. Disord.* 70, 25–34. doi: 10.1016/j.jcomdis.2017.10.003

Lukyanenko, C., and Fisher, C. (2016). Where are the cookies? Two- and three-year-olds use number-marked verbs to anticipate upcoming nouns. *Cognition* 146, 349–370. doi: 10.1016/j.cognition.2015.10.012

MacLeod, C. M. (2008). "The concept of inhibition in cognition," in: *Inhibition in Cognition*, eds D. S. Gorfein and C. M. MacLeod (Washington, DC: American Psychological Association). doi: 10.1037/11587-001

Mani, N., and Huettig, F. (2012). Prediction during language processing is a piece of cake-But only for skilled producers. *J. Exp. Psychol.* 38, 843–847. doi: 10.1037/a0029284

Martin, I., and McDonald, S. (2004). An exploration of causes of non-literal language problems in individuals with Asperger syndrome. *J. Autism Dev. Disord.* 34, 311–328. doi: 10.1023/B:JADD.0000029553.52889.15

Martins, A., Santos, A. L., and Duarte, I. (2018). "Comprehension of relative clauses vs. control structures in SLI and ASD," in: *Proceedings of the 42nd Annual Boston University Conference on Language Development* (Somerville, MA: Cascadilla Press).

Miyake, A., and Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Curr. Dir. Psychol. Sci.* 21, 8–14. doi: 10.1177/0963721411429458

Miyake, A., and Shah, P. (1999). *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control*. Cambirdge: Cambirdge University Press. doi: 10.1017/cbo9781139174909

Nation, K., Marshall, C. M., and Altmann, G. T. M. (2003). Investigating individual differences in children's real-time sentence comprehension using language-mediated eye movements. *J. Exp. Child Psychol.* 86, 314–329. doi: 10.1016/j.jecp.2003.09.001

Pearson, A., and Hodgetts, S. (2020). Can cerebral lateralisation explain heterogeneity in language and increased non-right handedness in autism? A literature review. *Res. Dev. Disabil.* 105:103738. doi: 10.1016/j.ridd.2020.103738

Pellicano, E., and Burr, D. (2012). When the world becomes "too real": A Bayesian explanation of autistic perception. *Trends Cogn. Sci.* 16, 504–510. doi: 10.1016/j.tics.2012.08.009

Pickering, M. J., and Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychol. Bull.* 144, 1002–1044. doi: 10.1037/bul0000158

Prévost, P., Tuller, L., Barthez, M. A., Malvy, J., and Bonnet-Brilhault, F. (2017). Production and comprehension of French wh-questions by children with autism spectrum disorder: A comparative study with specific language impairment. *Appl. Psycholinguist.* 38, 1095–1131. doi: 10.1017/S0142716417000169

Schuwerk, T., Sodian, B., and Paulus, M. (2016). Cognitive mechanisms underlying action prediction in children and adults with autism spectrum condition. *J. Autism Dev. Disord.* 46, 3623–3639. doi: 10.1007/s10803-016-2899-x

Shain, C., Blank, I. A., van Schijndel, M., Schuler, W., and Fedorenko, E. (2020). fMRI reveals language-specific predictive coding during naturalistic sentence comprehension. *Neuropsychologia* 138:107307. doi: 10.1016/j.neuropsychologia.2019.107307

Sinha, P., Kjelgaard, M. M., Gandhi, T. K., Tsourides, K., Cardinaux, A. L., Pantazis, D., et al. (2014). Autism as a disorder of prediction. *Proc. Natl. Acad. Sci. U.S.A.* 111, 15220–15225. doi: 10.1073/pnas.1416797111

Sukenik, N., and Friedmann, N. (2018). ASD is not DLI: Individuals with autism and individuals with syntactic DLI show similar performance level in syntactic tasks, but different error patterns. *Front. Psychol.* 9:279. doi: 10.3389/fpsyg.2018.00279 van de Cruys, S., Evers, K., van der Hallen, R., van Eylen, L., Boets, B., De-Wit, L., et al. (2014). Precise minds in uncertain worlds: Predictive coding in autism. *Psychol. Rev.* 121, 649–675. doi: 10.1037/a0037665

Vogel, D., Falter-Wagner, C. M., Schoofs, T., Krämer, K., Kupke, C., and Vogeley, K. (2019). Interrupted time experience in autism spectrum disorder: empirical evidence from content analysis. *J. Autism Dev. Disord.* 49, 22–33. doi: 10.1007/s10803-018-3771-y

Was, C. A., Sansosti, F. J., and Graham, E. (2018). Eye movements while processing syntactically ambiguous sentences in adolescents with autism spectrum disorders: preliminary outcomes and benchmarking data. *Glob. J. Intellectual Dev. Disabil.* 4, 76–81. doi: 10.19080/gjidd.2018.04.555645

Wittke, K., Mastergeorge, A. M., Ozonoff, S., Rogers, S. J., and Naigles, L. R. (2017). Grammatical language impairment in autism spectrum disorder: exploring language phenotypes beyond standardized testing. *Front. Psychol.* 8:532. doi: 10.3389/fpsyg.2017.00532

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

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

frontiers in Psychology

ORIGINAL RESEARCH
published: 11 November 2021
doi: 10.3389/fpsyg.2021.719447

Verb-Mediated Prediction in Bilingual Toddlers

Ane Theimann[1], Ekaterina Kuzmina[2] and Pernille Hansen[2,3]*

[1] Department of Linguistics and Scandinavian Studies, University of Oslo, Oslo, Norway, [2] Center for Multilingualism in Society Across the Lifespan, University of Oslo, Oslo, Norway, [3] Department of Humanities, Inland Norway University of Applied Sciences, Hamar, Norway

Prediction is an important mechanism for efficient language processing. It has been shown that as a part of sentence processing, both children and adults predict nouns based on semantically constraining verbs. Language proficiency is said to modulate prediction: the higher proficiency, the better the predictive skill. Children growing up acquiring two languages are often more proficient in one of them, and as such, investigation of the predictive ability in young bilingual children can shed light on the role of language proficiency. Furthermore, according to production-based models, the language production system drives the predictive ability. The present study investigates whether bilingual toddlers predict upcoming nouns based on verb meanings in both their languages, and whether this ability is associated with expressive vocabulary. Seventeen Norwegian-English bilingual toddlers (aged 2;5–3;3), dominant in Norwegian, participated in the study. Verb-mediated predictive ability was measured *via* a visual world paradigm (VWP) experiment, including sentences with semantically constraining and neutral verbs. Expressive vocabulary was measured by MacArthur-Bates CDI II. The results suggested that the toddler group predicted upcoming noun arguments in both their dominant and non-dominant languages, but were faster in their dominant language. This finding highlights the importance of language dominance for predictive processing. There was no significant relationship between predictive ability and expressive vocabulary in either language.

Keywords: semantic prediction, sentence processing, visual world paradigm (VWP), eye-tracking, bilinguals, toddlers, children

OPEN ACCESS

Edited by:
*Franziska Kretzschmar,
Leibniz Institute for the German Language (IDS), Germany*

Reviewed by:
*Anouschka Foltz,
University of Graz, Austria
Judith Schlenter,
University of Cologne, Germany*

Correspondence:
*Ane Theimann
ane.theimann@iln.uio.no*

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *02 June 2021*
Accepted: *18 October 2021*
Published: *11 November 2021*

Citation:
Theimann A, Kuzmina E and Hansen P (2021) Verb-Mediated Prediction in Bilingual Toddlers. Front. Psychol. 12:719447. doi: 10.3389/fpsyg.2021.719447

INTRODUCTION

One of the reasons why auditory language processing is so efficient is linguistic prediction, which implies pre-activation of linguistic input before it has been uttered (Huettig, 2015; Karaca et al., 2021). A growing body of research has shown that both children (Borovsky et al., 2012; Mani and Huettig, 2012; Mani et al., 2016) and adults predict upcoming linguistic input during auditory language comprehension (Altmann and Kamide, 1999; Hintz et al., 2017; Ito et al., 2017). Karaca et al. (2021) suggest that language proficiency facilitates prediction, while Mani and Huettig (2012) argue that the predictive ability is connected to language production and, as such, to expressive vocabulary.

Children growing up acquiring two languages can possibly have a substantial variance in proficiency between these languages. In addition, they can have highly different expressive vocabulary sizes in the two languages (Conboy and Thal, 2006). Hence, investigation of language prediction in children who acquire more than one language could potentially shed more light on the factors contributing to its development. Studies investigating prediction in language comprehension in bilingual children are few (Brouwer et al., 2017; Lemmerth and Hopp, 2019; Meir et al., 2020), and to the best of our knowledge, no studies have focused on this ability in bilingual toddlers.

In their seminal work, Altmann and Kamide (1999) discovered that monolingual English-speaking adults looked toward a specific object faster when they were given a verbal cue to which object would be mentioned. In their visual world paradigm (VWP) experiment, the adult participants listened to audio stimuli consisting of sentences such as *The boy will eat/move the cake*, while they looked at visual stimuli depicting different objects where one was the target. For example, with respect to the aforementioned sentence: a boy, a cake, a ball, a toy train, and a toy car. While all of the objects depicted were *movable*, only the cake was *edible*. The researchers found that the participants' gaze moved toward the cake faster upon hearing the semantically constraining verb *eat* than the more semantically neutral verb *move*. These findings were in favor of the hypothesis that adults predict nouns based on the semantic relationship between verbs and nouns; the adults in this study predicted upcoming nouns based on the semantic restrictions of verbs.

Previous studies also provide ample evidence for presence of predictive processing in monolingual children as young as 2 years old. Mani and Huettig (2012) employed the VWP to investigate whether monolingual German-speaking toddlers could use semantic cues represented by verb meanings to predict upcoming nouns. The toddlers listened to sentences such as *The boy eats/sees the big cake*, while looking at a screen with two pictures, where only one object was *edible*. The toddlers made predictive eye movements upon hearing semantically restrictive verbs (e.g., *eat*), but not when hearing non-restrictive verbs (e.g., *see*). Similarly, Borovsky et al. (2012) found that monolingual English-speaking children aged 3–10 years predict nouns based on verbs as well as sentential theme. The children were presented with four pictures (e.g., a treasure, a ship, a bone, and a cat), while hearing sentences such as *The pirate hides the treasure* or *The dog hides the bone*. In addition to semantic cues, there are other available cues to pre-activate upcoming linguistic input, such as prosodic, phonological, and morphosyntactic cues. An example of a morphosyntactic cue is grammatical gender, which can be used already by young children to predict upcoming nouns. Lew-Williams and Fernald (2007) showed that already by the age of 3, Spanish-speaking monolinguals identified target objects faster based on the gender-marked articles (el/la). The children heard sentences such as *Encuentra la pelota* 'Find the$_{FEM}$ ball', and saw two pictures of objects that were either both feminine or of differing grammatical gender. The children found the object faster in the different-gender trials, suggesting that they used the gender-marked article as a cue.

The use of gender-marked articles to predict upcoming nouns has also been studied in adult bilinguals and L2 learners. The results are conflicting. Lew-Williams and Fernald (2010) used the same VWP study as with the 3-year-old described above, to investigate if adult English-speaking L2 learners of Spanish had the ability to predict based on the articles in Spanish. The adult L2-learners in this study did not predict based on gender-marked articles. In a study with the same design as Lew-Williams and Fernald (2010), but with more experienced L2 speakers of Spanish, Grüter et al. (2012) found that the L2 speakers used the gender marked article to predict familiar nouns. However, the L2 speakers were less efficient in their use of the predictive cue compared to native speakers of Spanish. Surprisingly, the L2 learners were better at using the gender marked article to predict novel nouns than they were with familiar nouns. In another study on adult English-speaking L2 learners of Spanish, Dussias et al. (2013) found that more experienced L2 learners predicted based on grammatical gender, whereas those with less experience did not.

Investigations on adult bilingual's ability to predict have looked at not only morphosyntactic cues, but also semantic cues. For instance, Dijkgraaf et al. (2017) used the VWP to investigate late bilingual adults' ability to predict upcoming nouns based on the semantic relationship between verbs and nouns. Dutch-English bilinguals (with dominance in Dutch) and a control group of English monolinguals heard sentences such as *Mary knits/loses a scarf*, while looking at a screen depicting four objects where all could be *lost*, but only one was *knittable*. Dijkgraaf et al. (2017) argue that it is important to test bilinguals' predictive ability in both languages, due to individual differences connected to this ability, such as proficiency and vocabulary sizes. The researchers found that the bilinguals predicted based on semantic cues in both languages, but slower than the monolinguals (prediction effects reached significance 100 ms later in both languages). In a more recent study, Dijkgraaf et al. (2019) investigated Dutch-English bilinguals' ability to predict upcoming semantic information in both their languages. The bilinguals saw four pictures, where three were distractor pictures and the fourth was either the target picture or a semantically related competitor. For instance, when the bilinguals heard the sentence *Her baby doesn't like to drink from a bottle*, the target picture was a bottle, but in the semantically related trial it was a picture of a glass. The researchers found that the bilinguals predicted the semantics of target words in both conditions and in both languages. However, the prediction effect size was larger in the L1 than in the L2. Hopp (2015) investigated whether adult English L2 learners of German make predictions based on morphosyntactic cues (i.e., case marking) and verb semantics. The results showed that the bilinguals did not use morphosyntactic cues to predict, but they did use semantic cues.

Bilinguals are seldom completely balanced between their languages, neither in use nor in proficiency (Grosjean, 1989). Karaca et al. (2021) postulate that *language proficiency* modulates predictive processing: More proficient monolingual children and L2 learners are more likely to predict during sentence comprehension. Furthermore, based on previous studies on the predictive ability in monolingual children and adult L2 leaners, the researchers argue that language proficiency modulates the predictive ability in both L2 and L1. Similarly, Kaan (2014) argues that language users' *lexical representations*, specifically the quality of these representations, affects the ability to predict.

The researcher defines the quality of lexical representations as stability and accuracy of the language users' knowledge of it—its form, meaning, and use. Thus, lexical representations of higher quality have fewer lexical competitors and will be activated and chosen faster and/or more accurately during language processing. Further, Kaan (2014) argues that through *exposure* to a specific language, one learns to associate certain linguistic elements with each other, and one stores the frequency of how often specific linguistic information occurs in the same context. Bilinguals are intriguing in this respect: with exposure divided between two languages, both linguistic representations and associations between them might be weaker, potentially affecting the ability to predict upcoming linguistic elements (Kaan, 2014). She further argues that although predictive processing in an L2 is similar to that in L1, it might be affected by less language exposure. Similarly, the weaker links hypothesis (Gollan et al., 2008) postulates that since bilinguals divide their time between two languages, they have weaker links between semantics and phonology in both languages compared to monolingual peers. These weaker links could result in a reduced ability to predict. From these assumptions, one would expect individual variation between bilinguals, depending on their exposure, proficiency, and use of each of their languages. Thus, increased language exposure and use and a higher proficiency could lead to more efficient predictive processing.

The studies described above investigated the predictive ability in adults speaking more than one language. To date, there are few studies devoted to predictive processing in bilingual children, especially of children younger than 3 years old. Lemmerth and Hopp (2019) tested whether bilingual Russian-German 8- and 9-year-old could predict upcoming nouns based on gender-marked articles (der/die/das) in German, and compared them to monolingual German children (also aged 8–9 years). The study included both simultaneous bilinguals, who acquired two languages from birth, and sequential bilinguals, who acquired or one from birth and another later on, The children heard sentences such as *Wo ist der/die/das gelbe [N]?* 'Where is the$_{MASC/FEM/NEUT}$ yellow [N]?', while looking at pictures of four objects, of which only one had the grammatical gender mentioned in the sentence. The researchers found that the simultaneous bilingual children could use gender information to predict regardless of gender congruency, while the successive bilingual children would only predict when there was gender congruency between the two languages. Meir et al. (2020) investigated whether Russian-Hebrew bilingual children (4–8 years old) had the ability to predict upcoming nouns based on case-marking cues, and compared them to monolingual Russian children (aged 3–6 years) and Hebrew children (aged 4–8 years). The children looked at pictures, for example of a cabbage, a bunny and a fox. The researchers found that the bilingual children predicted based on case markers in Russian, as they looked at the agent (e.g., the fox) of the sentence upon hearing the accusative-marked NP (e.g., the bunny). However, they were slower than the monolingual Russian children. At the same time, the bilinguals used the case markers to predict also in Hebrew, whereas monolingual Hebrew children did not—as case-marking cues are assumed to be weighted lower than word order in Hebrew. Brouwer et al. (2017) tested the predictive ability based on semantic cues, employing the VWP, in Dutch monolingual and bilingual 4- and 5-year-old. The bilinguals spoke a variety of languages in addition to Dutch, but were only tested in Dutch. Of the bilinguals, 85% had learned Dutch before or around their first birthday, and their proficiency in Dutch was ranked as high. The children heard sentences such as *The boy eats/sees the big cake* while being presented with visual stimuli depicting two objects, where only one was edible. Brouwer et al. (2017) found that all the children (4- and 5-year-old monolinguals and bilinguals) predicted upcoming noun arguments based on the semantics of verbs. The researchers also found that the 4-year-old bilinguals predicted faster than their monolingual peers.

Although the number of studies on factors mediating predictive linguistic processing in bilingual children is relatively sparse, there are theories attempting to account for mediating factors of this ability for children and adults. According to production-based models, it is the production system that drives this ability (Pickering and Garrod, 2013; Pickering and Gambi, 2018). The foundational assumption of this theory is that the comprehension and the production systems are interwoven, allowing us to covertly imitate the speaker's production and predict their next word (Pickering and Garrod, 2013). Huettig (2015) sees production as an underlying mechanism for the predictive ability, and argues that in predictive processing we use "fully specified production representations" (p. 125).

Several studies have indeed shown a link between production (expressive vocabulary) and prediction. For instance, in a study by Ito et al. (2017), for half of the trials in a VWP experiment, the researchers asked the participants to just listen to audio stimuli, and for the other half they asked them to listen and shadow (i.e., repeat the sentence back as fast as they could). The researchers found that predictive eye movements happened earlier during the shadow tasks than during the listen tasks. They concluded that the study supports the hypothesis that production facilitates prediction. The link between production and predictive processing has also been found in studies with monolingual toddlers. For instance, Mani and Huettig (2012) showed that monolingual toddlers with larger expressive vocabularies (i.e., from 225 words) showed predictive eye movements suggesting that they were able to employ semantic cues for predictive processing. At the same time, the toddlers with smaller expressive vocabularies (i.e., fewer than 225 words) did not show predictive eye movements. Similarly, Mani et al. (2016) found that monolingual toddlers with larger expressive vocabularies had significantly more looks to the target picture during the predictive window, compared to children with smaller expressive vocabularies.

More studies investigating the predictive ability in bilingual children could help shed light on the role of proficiency and exposure to the languages in question. Compared to their monolingual peers, simultaneous bilinguals typically have larger *total vocabularies* (i.e., total sum of words known from all languages), comparable *conceptual vocabularies* (i.e., concepts that they have a word for in either one or both languages), and smaller language-specific vocabularies (Pearson et al., 1993; De Houwer et al., 2014). It is well-established that early grammatical development depends on lexical development (Bates and Goodman, 2001; Devescovi et al., 2005), a connection that holds *within each language* for bilinguals (Conboy and Thal, 2006). At the same time, studies of cross-linguistic influence point toward cognitive permeability between languages for simultaneous bilinguals (Döpke, 2001; Hulk, 2001), meaning that the processing of input in one language can indeed influence the acquisition of the other, as long as there is structural overlap between them. The Unified Model (MacWhinney, 2008, 2012) postulates that words that appear together, for example, verbs and nouns that often occur together, map on to each other. Words acquired in a non-dominant language may benefit from mappings made in the dominant language. Furthermore, according to the Unified Model, there is extensive transfer of knowledge from the dominant to the non-dominant language. Thus, an intriguing question is whether bilingual children's predictive ability relies on *within-language* vocabulary, as grammatical development in general, or the vocabulary in the strongest language, if predictive abilities in the non-dominant language comes as a result of cross-linguistic influence.

Simultaneously bilingual children are interesting for another reason: while we see great variability in the lexical development of monolingual children, there is reason to expect even more variability among bilinguals. They may be balanced between their languages or stronger in either, depending on a variety of factors, including the family language policy and the societal attitudes toward their languages. Hence, data from simultaneous bilinguals can potentially illuminate the relationship between prediction and expressive vocabularies. It is therefore important to look at both languages of the bilingual children. To date, there have been no studies investigating predictive ability based on semantic cues in both languages of bilingual children, and no studies at all investigating this ability in bilingual toddlers.

The Current Study

In the current study, we investigate verb-mediated prediction in a group of Norwegian-English bilingual toddlers dominant in Norwegian, and its relationship with their expressive vocabularies in both languages. Norwegian and English are structurally similar languages, with SVO (i.e., Subject-Verb-Object) order, which makes it possible to investigate verb-mediated prediction within the same sentence structure across languages. The study considers the following two research questions:

(1) Do Norwegian-English bilingual toddlers use verb meanings to predict upcoming noun arguments in Norwegian and/or English? If they do, is there a difference in speed of predictive processing between the dominant (i.e., Norwegian) and non-dominant (i.e., English) language?
(2) Is the linguistic predictive ability associated with linguistic production skills, specifically expressive vocabulary?

Based on previous studies, we hypothesized that the group of Norwegian-English bilingual toddlers: firstly, would predict noun arguments based on semantically constraining verbs in their dominant language (i.e., Norwegian). Secondly, we hypothesized that the toddlers would either not predict in their non-dominant language (i.e., English) or predict significantly slower compared to their dominant language.

With regard to the second research question, we hypothesized that we would find a significant positive association between the predictive ability in one language and expressive vocabulary size in the same language. This hypothesis was based on Mani et al. (2016) reporting the predictive ability to correlate with expressive vocabulary, and the findings of Conboy and Thal, 2006 indicating that lexical development in one language primarily leads to better grammatical abilities in the same language.

In addition to the two main research questions, we had one exploratory goal. Given that bilingual children acquire words from two languages at the same time, we explored relationships between the total vocabulary (i.e., the total sum of words in both languages) and predictive ability in either language.

MATERIALS AND METHODS

Participants

We recruited 18 simultaneous Norwegian-English bilingual toddlers for the study. We excluded one participant due to poor comprehension of the stimuli in the eye-tracking experiment, leaving us with 17 participants (6 females and 11 males) aged 2;5–3;3 (years;months) (M = 2;8, SD = 0.26). We recruited the toddlers *via* personal networks and posts on social media. To collect information on each toddler's language background, their parents filled out an electronic questionnaire based on the Parent of Bilingual Children Questionnaire (PaBiQ) (COST Action IS0804, 2011; Norwegian version; Hansen and Simonsen, 2016). We used their responses to calculate the balance in the children's language exposure, following Hansen et al. (2019). According to this calculation, all the toddlers were dominant in Norwegian. Corroborating this analysis, the parents of all the toddlers reported that their child felt most at home in Norwegian.

All the toddlers lived in or close to Oslo, and went to Norwegian-speaking day care. None of the toddlers used glasses, nor did any of them have hearing impairments or frequent ear infections. The toddlers' expressive vocabularies were assessed by the parents filling out the electronic version of the parental report tool MacArthur-Bates Communicative Development Inventories Words and Sentences (i.e., MB-CDI II),[1] developed for 16– to 36-month-old children in Norwegian (Kristoffersen and Simonsen, 2012), and for 16– to 30-month-old children in English (Fenson et al., 2007). In line with the reports on language dominance, all the toddlers had larger vocabularies in Norwegian (ranging from 281 to 664 words, M = 549.41, SD = 95.54) than in English (ranging from 20 to 565 words, M = *217.56*, SD = 156.97). The toddlers' vocabulary scores, including information about age and gender, is given in **Supplementary Table 1**. Parents gave written informed consent before participation in the study. The toddlers received a small toy after completing each session. Prior to

[1] Since only MB-CDI II was used in this study, for the remainder of this paper *CDI* will be used to refer to *MB-CDI II*.

commencing data collection, the Norwegian Center for Research Data (NSD) had approved the study.

To ensure the reliability of the eye-tracking experiment, we piloted it on 20 Norwegian-English bilingual adults (8 females and 12 males) aged 22–58 years (M = 28.4, SD = 7.4). These participants gave written consent to participate in the current study, and had the chance to win a gift certificate.

Materials

Eye-Tracking Experiment

The eye-tracking experiment employed the VWP and consisted of 56 stimulus sets, divided into four lists, two in each language, with14 stimuli sets in each list. Each stimuli set included one audio stimulus (i.e., either a semantically constraining or a neutral sentence—e.g., *The boy* ***eats/takes*** *the green apple*) and one visual stimulus, with three corresponding pictures: a picture of a boy/girl displayed in the top middle part of the screen, which functioned as a fixation picture, and the picture pairs (target and distractor) displayed lower on the right and left sides of the screen (see **Figure 1**). Including a fixation picture of the sentential subject made the participants look toward that picture at the start of the sentence. This minimized the possibility that the participants looked toward the target picture too early or by chance. We placed the fixation picture at the top middle at the screen because it is more natural to look at the top of the screen first. In the fixation pictures, both the boy and the girl looked straight down, so that the participants would not be biased toward any of the pictures. The picture pairs were bigger (9.7 cm°×°6.5 cm) than the fixation picture (5.5 cm°×°5.5 cm). We edited the pictures with Gimp, version 2.8 (The GIMP Team, 2018), so that all the pictures had a white background and no shadows. We found most of the pictures in the picture database Colourbox.com (2018), and the rest via searches on the internet.

In every picture pair, the pictures had similar size, color, and shape, and both were either photos or vectors. If the target picture was animate, the distractor picture was also animate. The object in the distractor picture was semantically and associatively unrelated to the object in the target picture and to the constraining verb. The reason behind this was that previous VWP studies have shown that participants tend to look more toward visual objects that are semantically related to the spoken word, than to visual objects that are unrelated (Huettig et al., 2006). The reason for only including two pictures (target and distractor) in addition to the fixation picture was to prevent toddlers from becoming overwhelmed by the visual stimuli. The position of the target picture appearing on the right or on the left side of the screen (see **Figure 1**) was counterbalanced. The position of the target picture was also counterbalanced across conditions.

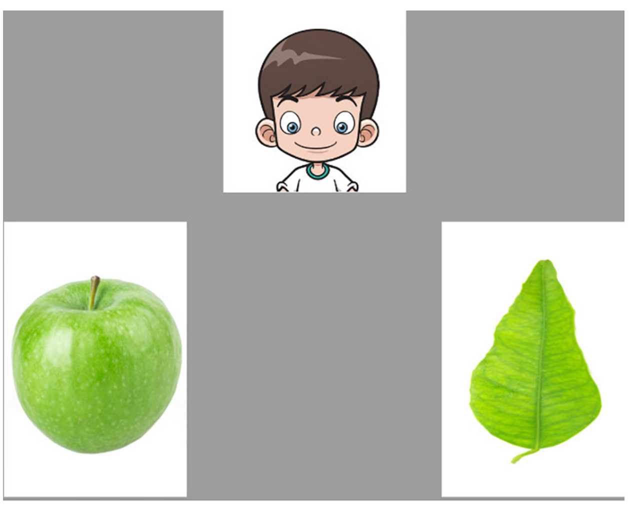


FIGURE 1 | Example of the visual stimuli for the auditory stimulus *This is a boy. The boy eats the green apple*.

For each language, 14 sentence pairs were created (all sentences can be found in **Supplementary Table 2**). A sentence pair consisted of one sentence with a semantically constraining verb (e.g., *eats*) and one sentence with a neutral verb (e.g., *takes*). Within one language, each verb was used only once in the experiment. The same semantically constraining verbs were used in both languages. However, the noun arguments following the verbs differed between the languages. For example, the corresponding sentence pair for *The boy* ***eats/takes*** *the orange carrot* in English was *Gutten* ***spiser/tar*** *det grønne eplet* 'The boy **eats/takes** the green apple' in Norwegian. Using different noun arguments ensured that participants saw each picture pair only once during the entire experiment consisting of two sessions. For example, for the sentence pairs described above, the target pictures were a green apple or an orange carrot in the Norwegian and English tasks respectively. We designed the sentences with neutral verbs so that the verbs would plausibly fit with both pictures; for instance, if the verb was *pick up,* it was possible to pick up both objects in the pictures. In some cases, we had to include a preposition after the verb for the stimulus sentences to ensure grammatical correctness (e.g., *The girl sits* ***on*** *the cold bench*), or to make verbs semantically constraining (e.g., *The girl draws* ***with*** *the blue crayon*).

A short context sentence preceded each sentence, namely *Her er det en jente/gutt* 'Here there is a girl/boy' in Norwegian and *This is a girl/boy* in English. The context was neutral, so the toddlers were not primed to look at the target or distractor pictures. From experience, toddlers are easily distracted when they hear names of family members or friends, we therefore decided that all the sentences should start with *The boy* or *The girl.*

The experimental material consisted of four lists: two in Norwegian and two in English, created to balance the visual and auditory stimuli across the participants, while also avoiding repetition. The two lists from each language were evenly distributed among the toddlers. In other words, one list from each language was used for half of the toddlers. Equally many sentences in each list began with *The boy* and *The girl*, half of which had semantically constraining and neutral verbs, respectively. Each child saw each picture pair only once. Trials were fully randomized on a by-participant basis by the experimental software.

A Norwegian-English bilingual female speaker recorded the auditory stimuli in a quiet environment with a Zoom Q2n, with 48,000 Hz. We asked her to record the auditory stimuli because of her clear pronunciation of Norwegian and English, and because she had partly been growing up in an English-speaking society and partly in a Norwegian-speaking society. To edit the auditory stimuli we used Audacity, version 2.2.2 (Audacity Team, 2021). We edited the length of a short pause (M = 812.47 ms; SD = 210.74) after the context sentence, to make sure the verb onset was at exactly 3,500 ms into the trial. Similarly, we edited the length of another short pause (M = 564.23 ms; SD = 143.23) after the verb, to make sure the noun onset was at 5,300 ms. This gave the toddlers a time window of 1,800 ms to predict the upcoming noun (e.g., This is a boy. [pause] The boy eats [pause] the green apple). No linguistic cues appeared during the predictive window that could bias the toddlers to look toward either of the pictures.

All the adult pilot participants completed the Norwegian version first, and the English version about 2 weeks later. After each session, we asked the participants how sentences sounded to them and whether the visual stimuli fitted the sentences. Six adults noted that the audio stimuli sounded child-directed. All agreed that the sentences sounded natural, that the pictures matched the task, and that the two pictures shown at the same time were equally salient.

Procedure

The toddlers performed the eye-tracking task either in the Socio-Cognitive laboratory at the University of Oslo, or at the toddler's daycare center. The toddlers performed the eye-tracking task twice—first in Norwegian, and 1–2 weeks later in English. At the beginning of each session, we introduced the toddlers to a stuffed animal, and told them that the stuffed animal could only speak the language of that session. Toddlers sat on their parent's lap, facing a monitor. The eye-tracker was located underneath this monitor. Eye movements were recorded from the right eye with an SMI RED25mobile Eye Tracker, with a sample rate of 250 Hz. Auditory stimuli were presented through a speaker, connected to the monitor. We controlled and monitored the experiment by a laptop computer. The experimenter instructed the parents to sit still and not say anything during the experiment, so as to not guide the toddlers to look at any of the pictures.

We used an image of a bee, shrinking in on itself, to calibrate the toddlers' eye movements. To make the calibration more playful, the experimenter told the toddlers that the bee would fly for them if they looked at it. After a successful calibration, the experimenter told the toddlers that they would hear some stories about a boy and a girl, and that they should look at the pictures on the monitor in front of them. After the calibration, two practice trials followed. After each trial, we asked the toddlers to point to the pictures that matched the sentence they had just heard. Previous studies have shown that young children pay more attention performing eye-tracking tasks when they are asked to point to the correct picture after each trial (Trueswell, 2008). During the practice trials, if the toddlers did not point to the target picture, we repeated the practice trials until they did. In each trial, the three pictures appeared and stayed on the screen for 1,000 ms before the auditory stimuli started. Once the participant had pointed to a picture, the next trial was started by the experimenter. If a toddler pointed to the distractor picture, we interpreted it as incorrect sentence understanding. After half of the trials, a new calibration started, and the toddlers that needed it had a break before the second calibration. Each session lasted about 15 min. Within a week of each test session, the parents filled out the CDI form for the language tested that day.

Data Analysis

For the analysis, we only used data from trials where the toddlers understood the sentences correctly. As previously mentioned, if a toddler pointed to the distractor picture at the end of a sentence, we interpreted it as incorrect sentence understanding. Therefore, 9% of the trials were excluded (43 out of 476). All of the adults understood all of the sentences, so we kept all trials in the pilot study. We used both fixations and saccades in the analysis because young children have less stable patterns of eye-movements and tend to saccade within areas of cognitive fixation (Aring et al., 2007). Conventionally, fixations are defined as time periods when eyes fixate on a specific area and stay relatively stationary—from tens of milliseconds to several seconds, while saccades are defined as rapid eye movements between any two fixations (Holmqvist et al., 2011). In the remainder of this paper, *fixations* will refer to both saccades and fixations used in the analysis.

We did not perform statistical analysis on the data from the adult group and relied on visual investigation of the fixation curves only, since the pilot study only served as a proof of concept for the chosen experimental design. For the analysis of the data from the toddler group, we used the divergence point analysis reviewed in great detail and with remarkable clarity by Stone et al. (2020). We encourage our readers to acquaint themselves with the before mentioned paper. The analysis script in the current study is adapted from the tutorial provided in **Supplementary Material** by Stone et al. (2020).

The main goal of the divergence point analysis is to allow researchers studying online unfolding of language comprehension to estimate specific timepoints of effect onsets. Once estimated, the effect onset timepoints can be directly compared between experimental groups and/or conditions in order to conclude in which experimental group and/or condition the effect onset manifests earlier. Several methods can be used to perform the divergence point analysis and these methods have their specific advantages and disadvantages (Stone et al., 2020). In the current study, we used Generalized Logistic Mixed-effect Models (GLMM; Barr, 2008) and our effect of interest was the onset of verb-mediated predictive processing. We operationalized its onset as the timepoint that (1) is located between the verb and noun onsets in the constraining condition and (2) corresponds to a significant increase in target-fixations compared to distractor-fixations.

To analyze the data, we first defined the critical predictive window as the time period between the verb onset + 300 ms and the noun argument onset + 300 ms. Given that, firstly, adults use at least 200 ms to launch a saccade (Altmann and Kamide, 1999;

Salverda et al., 2014) and, secondly, that children are generally slower than adults at launching saccades (Bucci and Seassau, 2012; Lemoine-Lardennois et al., 2016), 300 ms were added to the verb onset as well as to the noun onset. We also added a buffer of 1,700 ms after the noun onset + 300 ms to detect a divergence point in the neutral condition where we did not expect any predictive processing.

Secondly, we used the GLMM. We fitted a generalized logistic mixed-effect model to data from each 20 ms time slot of this critical predictive window to compare binomial distributions of target and distractor fixations in each time slot. The time slot where a significant difference between the number of target- and distractor-fixations was observed for the first time was defined as the divergence point, specifically the onset of the verb-mediated processing. This method requires to run multiple statistical tests, which is associated with an increased risk of making a Type I error, that is, detecting an effect when it is in fact absent in reality. Therefore, to adjust for multiple comparisons, we used the false discovery rate (FDR) control correction (Benjamini and Hochberg, 1995) when fitting the model. We preferred the FDR correction to the Bonferroni correction, because the latter is associated with reduced statistical power and consequently increased probability to miss an effect that in fact exists in reality (i.e., a Type II error). The FDR correction makes *p*-values from each significance tests larger based on a specific algorithm. This results in a lower number of false positives passing the initially chosen alpha-level.

Following Mani and Huettig (2012), the ability to predict upcoming linguistic information (i.e., predictive ability) was operationalized as the difference in the number of target-fixations between the semantically constraining and neutral conditions within the critical predictive window. To measure the strength of the relationship between CDI vocabulary scores and predictive ability in each language separately, Spearman correlation coefficients were used. The same method was used to explore whether there is a relationship between the overall productive vocabulary and predictive ability. R (R Core Team, 2021) and RStudio (Rstudio Team, 2021) were used for the data analysis. The R script, detailed report of the analysis, and data can be found in the **Supplementary Material**.

RESULTS

Results of the Pilot Study With the Adult Group

The main goal of the pilot study was to ensure that the proposed experimental design captures verb-mediated prediction. As can be seen from **Figure 2**, the adult group showed predictive processing in both Norwegian and English: Percentages of target-fixations clearly increased (1) within the critical predictive window in the constraining condition and (2) after the noun onset in the neutral condition. Based on these results from the visual inspection of the fixation curves, we concluded that the proposed experimental design is suitable for studying verb-mediated predictive processing.

Results of the Main Study With the Toddler Group

Figure 3 summarizes performance of the toddler group and displays a noticeable increase in target fixations happening within the critical predictive window in the constraining conditions well as after the noun onset in the neutral condition in both Norwegian and English. Thus, this descriptive plot already suggests that the toddler group predicted upcoming nouns based on verb meanings in both their dominant and non-dominant language. To estimate exactly when this verb-mediated predictive processing started, a divergence point analysis was performed.

Figure 4 displays performance of the toddler group within the critical predictive window. The triangle point depicts the divergence point estimate calculated with the FDR correction. **Table 1** summarizes the divergence point estimates for different conditions and languages in the toddler group.

For Norwegian, the divergence point estimate was before the noun onset (i.e., 1,800 ms after the verb onset) in the constraining condition: 1,020 ms after the verb onset. As expected, in the neutral condition the divergence point estimate was after the noun onset pointing to the absence of predictive processing. These results suggest that the toddler group used verb meanings to predict upcoming noun arguments in their dominant language, Norwegian.

For English, the divergence point estimate was within the critical predictive window in the constraining condition: 1,620 ms after the verb onset. Similar to the Norwegian results, the FDR divergence point estimate in the neutral condition was outside of the critical predictive window pointing to the absence of predictive processing. These results provide evidence for verb-mediated predictive processing in the toddler group in their non-dominant language, English.

The toddler group was generally faster in both constraining and neutral conditions in their dominant language, Norwegian, compared to English. Specifically, in the neutral condition capturing verb-based integration of noun arguments, the toddlers were 240 ms faster in Norwegian. For the constraining condition this difference was 600 ms.

The predictive ability was calculated as the difference in the number of target-fixations between the semantically constraining and neutral conditions within the critical predictive window. There were no significant correlations between predictive ability and productive vocabulary in either Norwegian, $r = -0.14$, $p = 0.59$, or English, $r = 0.19$, $p = 0.47$. **Figure 5** shows relationships between productive vocabulary and predictive ability in both languages.

There was also no significant relationship between the total productive vocabulary and predictive ability in either Norwegian, $r = -0.26$, $p = 0.30$, or English, $r = 0.002$, $p = 0.996$. **Figure 6** shows relationships between total productive vocabulary and predictive ability in both languages.

DISCUSSION

The present study is the first to investigate the semantic predictive ability in bilingual toddlers in both their languages. We set

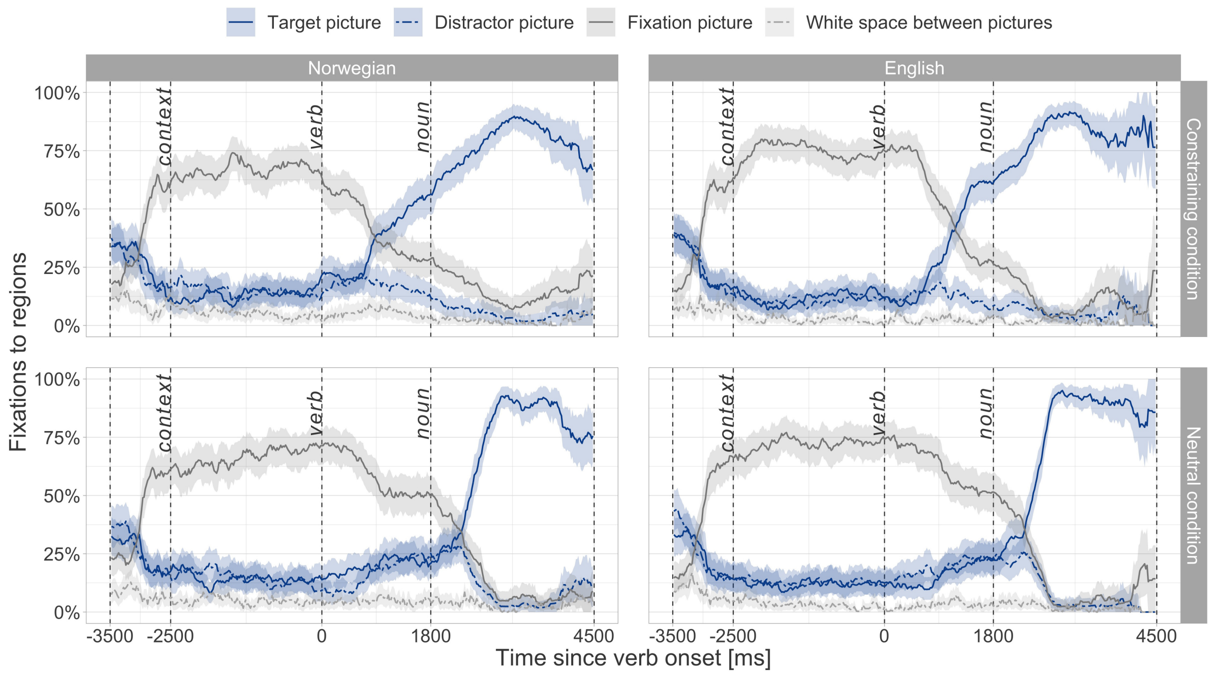


FIGURE 2 | Fixation curves made of mean fixation percentages with 95% confidence intervals calculated per 20 ms time bins for the four areas of interest in the adult group. "*Context*" refers to the time point when a context sentence starts, "*verb*"—when a verb starts, and "*noun*"—when a noun argument starts.

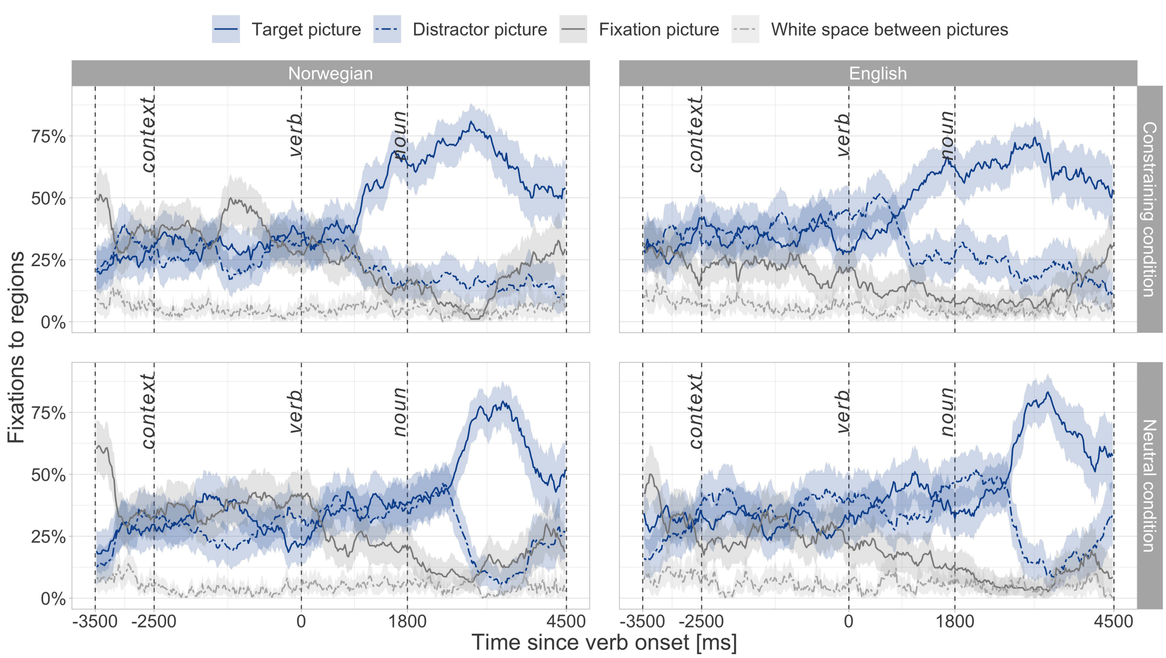


FIGURE 3 | Fixation curves representing mean fixation percentages with 95% confidence intervals calculated per 20-ms time bin for the four areas of interest in the toddler group. "*Context*" refers to the time point when a context sentence starts, "*verb*"—when a verb starts, and "*noun*"—when a noun argument starts.

out to answer two research questions. The first question was whether bilingual toddlers predict upcoming nouns based on verb meanings in both their dominant and non-dominant languages—namely Norwegian and English, respectively—and, if they do so, whether there is a difference in speed of predictive processing between the two languages. The second question was whether toddlers' production skills, specifically expressive vocabulary, mediates this predictive ability. Linguistic predictive

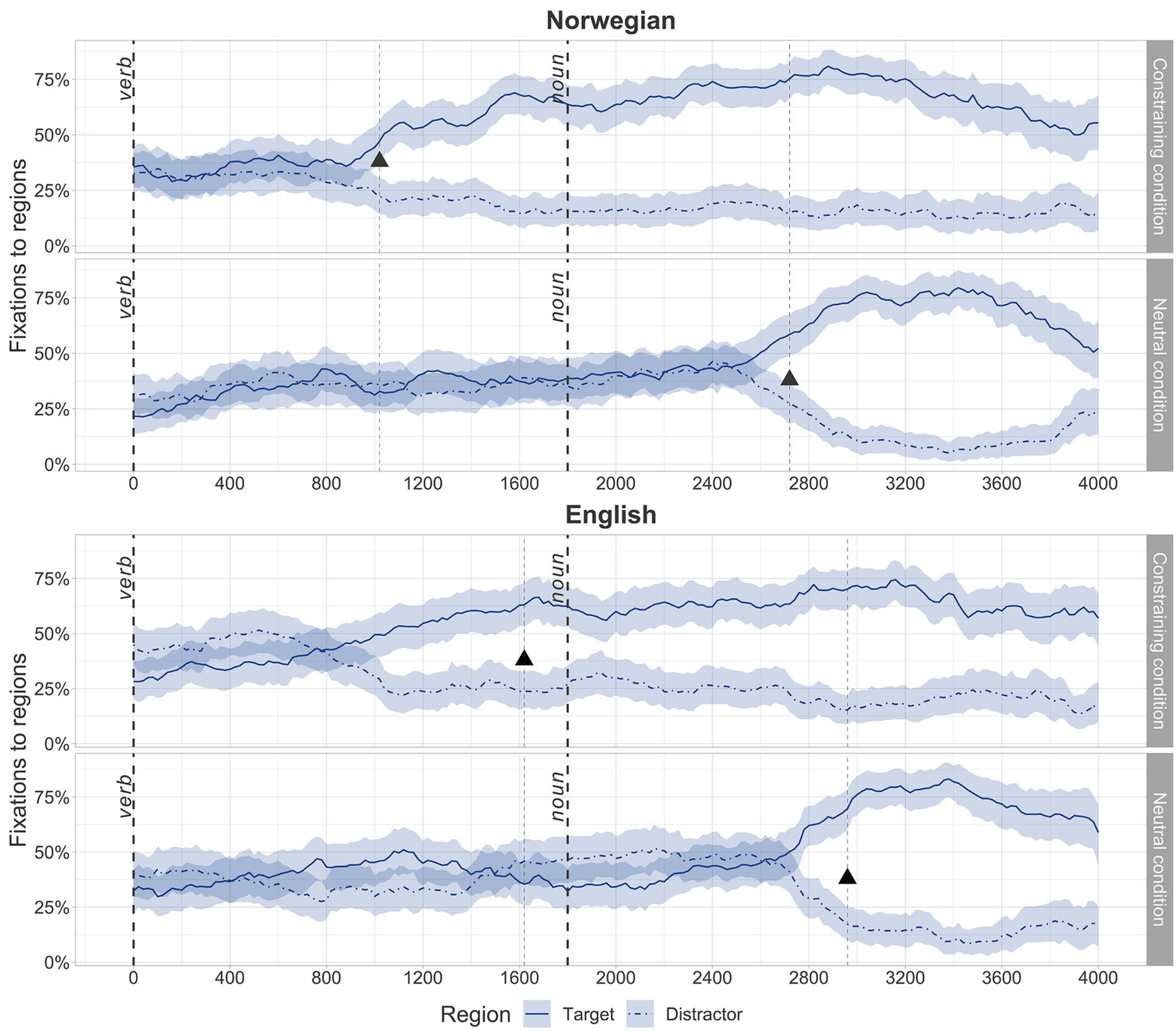


FIGURE 4 | Fixation curves for the critical predictive window. Triangles depict the divergence point estimate calculated with the generalized logistic mixed-effect models (GLMM) using the false discovery rate (FDR) correction.

TABLE 1 | Summary of the divergence point estimates.

Condition	Norwegian			English			Difference between the divergence point estimates, ms
	Time after verb onset, ms	*z*-score	*p*-value	Time after verb onset, ms	*z*-score	*p*-value	
Constraining	1,020	3.00	<0.05	1,620	2.95	<0.05	600
Neutral	2,720	3.28	<0.05	2,960	3.49	<0.05	240

processing was investigated by means of an eye-tracking experiment employing the VWP. The expressive vocabulary sizes were assessed with the Norwegian and English MacArthur-Bates Communicative Development Inventories Words and Sentences (i.e., MB-CDI II). Below, we will discuss our findings in relation to these questions.

Concerning our first research question, we hypothesized, firstly, that the toddler group would predict noun arguments based on semantically constraining verbs in their dominant language (i.e., Norwegian). Secondly, we hypothesized that they would either not predict in their non-dominant language (i.e., English) or predict significantly slower compared to their dominant language.

The results from the current study support the first part of this hypothesis: As expected, in Norwegian the divergence point estimate was within the critical predictive window (i.e., after the verb- and before the noun-onset) in the constraining condition and outside of it in the neutral condition. As previously

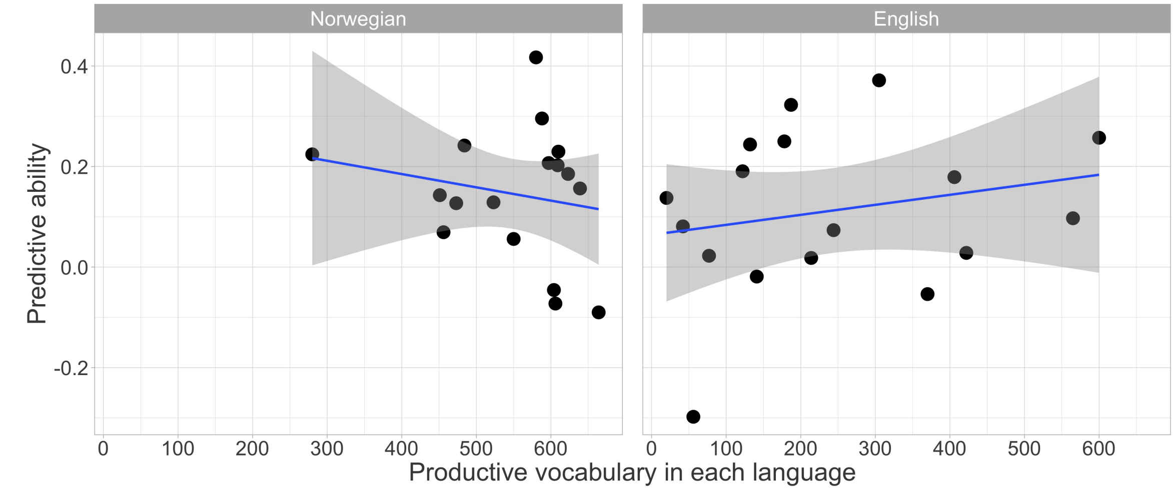


FIGURE 5 | Correlations between the predictive ability and expressive vocabulary.

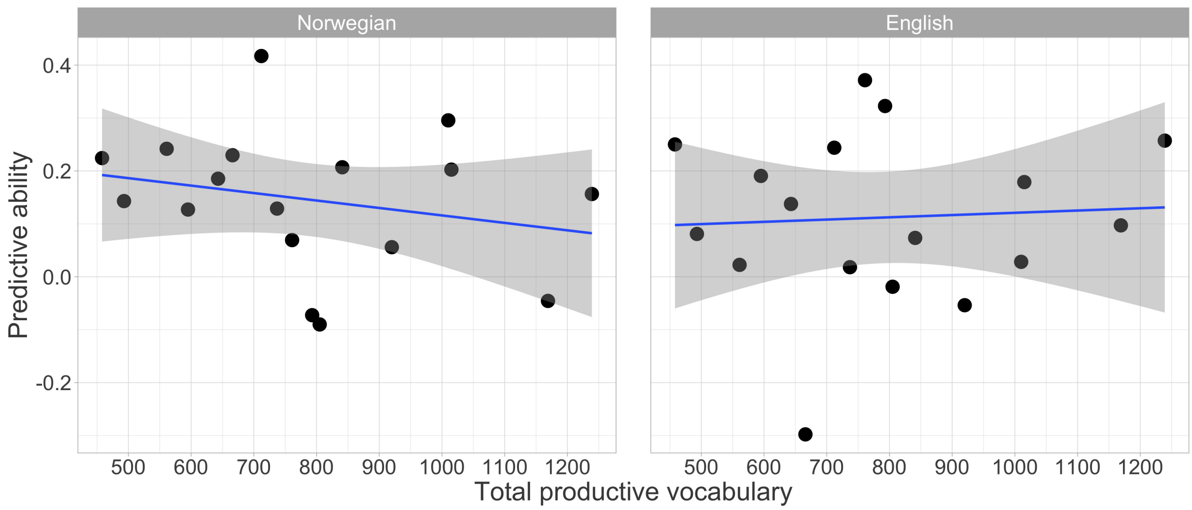


FIGURE 6 | Correlations between the predictive ability and total expressive vocabulary.

mentioned, the divergence point was defined as the first time slot with a significant difference between the number of target- and distractor-fixations. These results provide clear evidence for the presence of verb-mediated predictive processing of noun arguments in toddlers in their dominant language. These findings are in line with findings from previous research. In particular, previous studies have shown that monolingual toddlers at the age of 2 years (Mani and Huettig, 2012; Mani et al., 2016) as well as monolingual children aged 3–10 years (Borovsky et al., 2012) predict upcoming lexical information during sentence comprehension. Additionally, Brouwer et al. (2017) found that bilingual children aged 4 years old were able to predict in their majority language using semantic cues similarly to their monolingual peers.

The second part of our hypothesis connected to the first research question was also supported by the current findings: in English, the divergence point estimate was within the critical predictive window in the constraining condition and outside of it in the neutral condition. Thus, we concluded that the current study provides evidence for verb-mediated predictive processing in toddlers in a non-dominant language.

With regards to the speed of processing, the toddler group was generally faster in their dominant language, Norwegian, compared to their non-dominant language, English. In the constraining condition, which taps into predictive processing, the processing advantage for the domain language was 600 ms.

In English, the divergence point estimate were later in both neutral and constraining conditions in English compared to Norwegian (see **Table 1**). To the best of our knowledge, there are no existing studies on semantic prediction in bilingual toddlers in both their languages. However, the current results are in line with experimental evidence from studies with adults, finding that adult bilinguals predict slower in their L2 compared to monolinguals. For instance, Dijkgraaf et al. (2019) compared Dutch-English bilinguals' L1 to their L2 and found that they predicted based on semantic cues slower in L2 than in L1. Chun and Kaan (2019) showed that adult Chinese L2 learners of English predicted 180 ms slower than native English speakers when listening to English. Dijkgraaf et al. (2017) investigated semantic verb-mediated prediction and identified that adult Dutch-English bilinguals predicted 100 ms slower in both of their languages compared to native English speakers.

The current study is the first to test semantic prediction skills in bilingual 2;5- to 3;3-year-old toddlers in both languages. By testing both languages with almost identical stimuli in the same individuals, it is possible to investigate the differences between the prediction abilities in two languages within individuals. This eliminates between-group differences (e.g., speed differences in eye movements, cultural and linguistic backgrounds, as well as individual cognitive differences), which is difficult to avoid when comparing bilinguals' non-dominant language to native speakers (Dijkgraaf et al., 2019).

In the present study, the toddlers had less exposure to English than to Norwegian, potentially resulting in weaker mental representations in this language, and in turn slower prediction. These findings are in line with the theoretical accounts of Gollan et al. (2008), Kaan (2014), and Karaca et al. (2021), and indicate that semantic prediction is less efficient in the non-dominant language. Following the weaker links hypothesis (Gollan et al., 2008), the less efficient prediction could be due to weaker links between verb meanings and their arguments in the less practiced language. The current findings are also in line with the Unified Model (MacWhinney, 2008, 2012). Findings from studies of simultaneous bilingual language acquisition indicate that knowledge from one language can indeed influence the acquisition of another, as long as there is structural overlap. Since Norwegian and English are structurally similar languages, the toddlers who participated in our study could be transferring their ability to predict from the dominant to the non-dominant language. In fact, cross-linguistic influence in semantic (or conceptual) knowledge could be what allows them to predict in their non-dominant language at all. When the toddlers have understood that certain verbs are followed by certain noun arguments (e.g., eat and something edible) in their strongest language, they transfer this knowledge to the weakest language, and thus predict in this language as well. Thus, if a toddler knew the meaning of the verb *eat* in English, they could use their mental representations from Norwegian, their dominant language, to predict edible objects, even if they did not know the English names for the edible objects depicted on the screen.

Other studies of prediction with bilingual children also report findings that suggest transfer of knowledge from the dominant to the non-dominant language. Lemmerth and Hopp (2019) showed that Russian-German successive bilingual 8- and 9-year-old would only predict nouns based on grammatical gender in cases where there was gender congruency between Russian and German. The researchers argue that all nouns in both languages with the gender they heard get activated, so the nouns with gender congruency across languages benefited from the activation and eased prediction. At the same time, the nouns with incongruent gender suffered from competition effects. In a study with Russian-Hebrew bilinguals aged 4–8 years, the bilinguals were slower to predict based on case markers compared to their monolingual Russian-speaking peers (Meir et al., 2020). However, in contrast to Hebrew monolinguals, only the bilinguals used the case markers to predict in Hebrew. This indicates that the bilinguals transferred their case marker knowledge from Russian, where case markers are commonly used to predict, to Hebrew, where case markers are used less as a cue to predict.

To summarize, the results we obtained to answer our first research question, suggest that bilingual toddlers at the age of 2;6 are able to use verb meanings to predict upcoming noun arguments in both of their languages. However, they are faster at predicting in the language reported as their dominant one, where they have the largest vocabulary, according to parental reports.

The second research question was whether there was an association between predictive ability and production skills represented by expressive vocabulary. We hypothesized that there would be a positive correlation between the predictive ability and expressive vocabulary sizes; the toddlers with higher productive vocabulary sizes would predict faster. Previous studies on grammatical and lexical development in bilingual children indicate that lexical development in one language leads to better grammar in the same language only. Therefore, we expected a positive correlation between the predictive ability in one language and the expressive vocabulary in the same language.

The results did not support this hypothesis. There were no significant correlations between predictive ability and expressive vocabulary size within either of the two languages. This is in line with findings from Dijkgraaf et al.'s (2017) study, where the researchers did not find a significant relationship between predictive ability and production in the respective languages in a group of adult bilinguals. The researchers assumed that the method they used to measure bilinguals' vocabularies (i.e., LexTALE) was not sensitive enough to reflect variation in production skills in the group of adults, and that, in turn, this could have explained the non-significant results they obtained. Previous studies have provided evidence for the production-based models, suggesting that one of the possible mechanisms for language prediction during sentence comprehension is language production skills (e.g., Mani and Huettig, 2012; Mani et al., 2016; Martin et al., 2018). For instance, Mani and Huettig (2012) found a positive correlation between monolingual 2-year-olds' predictive ability and their productive skills. The toddlers with expressive vocabulary sizes from 225 words according to the MB-CDI had better predictive ability compared to the toddlers with smaller expressive vocabularies. Mani and Huettig argue that production skill is one of the underlying mechanisms, and that toddlers with larger expressive vocabularies will have stronger links between constraining verbs and their arguments, which in turn facilitates predictive processing. However, while the toddlers in the present study predicted, there was no link between the predictive ability and the same language vocabulary. We used the same method to measure expressive vocabulary as Mani and Huettig (2012). The obvious difference between the two studies is that their study concerned monolingual toddlers while and the present study concerned bilingual toddlers. As mentioned in the introduction, there are different ways to look at bilinguals' vocabularies, which we will discuss below. The findings in the present study do not provide evidence in support of the theoretical account where productive vocabulary is seen as an underlying mechanism of the predictive ability.

In addition to the two research questions, we had an exploratory goal: to investigate whether there was a correlation between the predictive ability in each of the languages and the total expressive vocabulary. There was no correlation between the predictive ability in either of the languages and the total expressive vocabulary. It is possible that this result is an artifact of the measure used. As mentioned in the introduction, bilingual children's lexical development can be assessed by first assessing their vocabulary in each language separately, and then use those data to calculate either a conceptual vocabulary or a total vocabulary (Pearson et al., 1993; De Houwer et al., 2014). Both of these measures have their advantages. This study calculated the total vocabulary, that is, the sum of words produced in Norwegian and English. The measure requires no qualitative judgment from the researcher, but may give inflated numbers, particularly for children who know many cognates, that is, words from different languages overlapping in form and meaning. Hence, a child scores 2 points for producing both the Norwegian word *banan* and its English equivalent *banana*, and 1 point for producing the Norwegian word only. The measure of conceptual vocabulary, on the other hand, is meant to reflect the number of concepts the child has a word for, regardless of which language the concept is in. Here, a child would score only 1 point whether they produced both the Norwegian *banan* and the English *banana* or only one of these words. This measure is more conservative, not inflated by the existence of cognates, and it may yield a more valid picture of bilingual children's lexical knowledge. However, the calculation requires the researcher to map conceptual equivalents among the words that the children produce in their two languages. This task is not trivial, as complete conceptual equivalents across languages are rare (Pavlenko, 2009; de Groot, 2013). Some words may seem to have an absolute conceptual equivalent, but a complete overlap for all the uses in a range of situations and contexts of the word is hard to find. Thus, establishing a child's conceptual vocabulary size is not straightforward. However, it is nevertheless possible that it is a more relevant measure for studies of verb-mediated prediction.

The current study is not without limitations. The first limitation we wish to address regards participant groups. Future studies could include a group of English-Norwegian simultaneous bilingual toddlers, to see if they would have the opposite results from the toddlers in this study: if they would show faster predictive abilities in English. Including a group of simultaneous bilingual toddlers with larger vocabularies in English than in Norwegian could help answering the question if proficiency is key to prediction. Future studies with older and more proficient bilingual children could shed light on whether the predictive ability in their non-dominant language would increase with increased proficiency in this language. In addition, more research on prediction in bilingual children at different ages, and with different proficiency levels between their languages would increase our knowledge on the role of proficiency for predictive ability. Another limitation of the current study is that we did not measure the toddlers' receptive vocabulary or other receptive language processing skills. The current study focused on the relationship between the predictive ability and production, however there are also models that link prediction to comprehension (Chang et al., 2006; Gambi and Pickering, 2013; Dell and Chang, 2014; Ness and Meltzer-Asscher, 2021). Future studies should investigate a relationship between the predictive ability and both production and comprehension. For both modalities, the measure of conceptual vocabulary may be more suitable than the total vocabulary measure used here. Concerning our statistical analyses, the main drawback of GLMM is its inability to measure variability of the divergence points. As such it is not suited for statistical comparison of these points between conditions and/or groups. In a recent paper, Stone et al. (2020) suggest using bootstrapping to enable estimation of uncertainty around the divergence points. While it is uncertain how well-suited a limited dataset with high variability such as ours is for bootstrapping, we see this as a promising direction for the field.

CONCLUSION

To conclude, findings from the current study suggest that bilingual toddlers predict upcoming nouns based on the semantic restrictions of verbs in both their dominant and non-dominant languages. However, they are faster in their dominant language. These findings are in line with the weaker links hypothesis: less exposure and lower proficiency in the non-dominant language lead to weaker mental representations and associations, which, in turn, result in slower linguistic predictive processing. Due to young age and limited experience with the non-dominant language, a toddler's ability to predict may still be developing in this language. Despite their lower proficiency in the non-dominant language the bilingual toddlers in our study still predicted in this language. Following the Unified Model, the prediction in the non-dominant language could be explained by a transfer of this ability from the dominant to the non-dominant language. Findings from previous studies concerning the possible association between the predictive ability and production are conflicting. The results from the current study do not lend support to the theoretical account where productive vocabulary is seen as an underlying mechanism of the predictive ability: there was no relationship between the predictive ability and the expressive vocabularies in either language or with total vocabulary. Based on the conflicting findings on the expressive vocabulary's role on predictive ability, more research is needed to investigate this relationship further.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

The study was reviewed and approved by Norsk senter for forskningsdata (NSD). Written informed consent to participate in this study was provided by the participants, and in case of the toddlers, by their legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

AT initiated and led the present project, from design *via* data collection to the writing process, provided the idea to investigate prediction in bilingual children, designed and revised the experiment, led the recruitment of participants, collected the data, and drafted the introduction, methods, and discussion sections. EK provided the idea to study verb-mediated prediction, performed the statistical analysis, and wrote and revised the parts on analysis and results. EK and PH contributed with suggestions on methodology. PH wrote parts of the introduction and discussion. All authors provided critical feedback and helped shape the study and the manuscript.

FUNDING

This work was partly supported by the Research Council of Norway through its Centres of Excellence funding scheme, project numbers 223265 and 250093.

ACKNOWLEDGMENTS

We would like to thank the participants, children, and their parents, as well as the pilot participants for providing the data for this research. We are grateful to Løvetann Ripoll for recording the auditory stimuli. We also gratefully acknowledge MultiLing's Socio-Cognitive Laboratory for access to their facilities. The two reviewers have provided insightful comments and suggestions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.719447/full#supplementary-material

REFERENCES

Altmann, G. T., and Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition* 73, 247–264. doi: 10.1016/s0010-0277(99)00059-1

Aring, E., Grönlund, M. A., Hellström, A., and Ygge, J. (2007). Visual fixation development in children. *Graefe's Arch. Clin. Exper. Ophthalmol.* 245, 1659–1665.

Audacity Team (2021). Audacity(R): Free Audio Editor and Recorder [Computer Software]. Version 2.2.2. Available online at: https://audacityteam.org/ (accessed August 31, 2018).

Barr, D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *J. Mem. Lang.* 59, 457–474. doi: 10.1016/j.jml.2007.09.002

Bates, E., and Goodman, J. C. (2001). "On the inseparability of grammar and the lexicon: evidence from acquisition," in *Language Development: The Essential Readings*, eds M. Tomasello and E. Bates (Hoboken, NJ: Blackwell Publishing), 134–162.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Statist. Soc. Ser. B* 57, 289–300.

Borovsky, A., Elman, J. L., and 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. *J. Exper. Child Psychol.* 112, 417–436. doi: 10.1016/j.jecp.2012.01.005

Brouwer, S., Özkan, D., and Küntay, A. C. (2017). "Semantic prediction in monolingual and bilingual children," in *Cross-Linguistic Influence in Bilingualism*, eds E. Blom, L. Cornips, and J. Schaeffer (Amsterdam: John Benjamins), 49–74.

Bucci, M. P., and Seassau, M. (2012). Saccadic eye movements in children: a developmental study. *Exper. Brain Res.* 222, 21–30. doi: 10.1007/s00221-012-3192-7

Chang, F., Dell, G. S., and Bock, K. (2006). Becoming syntactic. *Psychol. Revi.* 113:234. doi: 10.1037/0033-295x.113.2.234

Chun, E., and Kaan, E. (2019). L2 Prediction during complex sentence processing. *J. Cult. Cogn. Sci.* 3, 203–216. doi: 10.1007/s41809-019-00038-0

Colourbox.com (2018). Available online at: https://www.colourbox.com (accessed August 15, 2018).

Conboy, B. T., and Thal, D. J. (2006). Ties between the lexicon and grammar: Cross-sectional and longitudinal studies of bilingual toddlers. *Child Dev.* 77, 712–735. doi: 10.1111/j.1467-8624.2006.00899.x

COST Action IS0804 (2011). *Parents of Bilingual Children Questionnaire(PaBiQ). A Part of the LITMUS Battery (COST IS0804)*.

de Groot, A. M. B. (2013). "Bilingual memory: a short introduction," in *The Psycholinguistics of Bilingualism*, eds F. Grosjean and P. Li (Chichester: Wiley Blackwell), 171–191. doi: 10.1002/brb3.2121

De Houwer, A., Bornstein, M. H., and Putnick, D. L. (2014). A bilingual–monolingual comparison of young children's vocabulary size: evidence from comprehension and production. *Appl. Psycholinguist.* 35:1189. doi: 10.1017/S0142716412000744

Dell, G. S., and Chang, F. (2014). The P-chain: relating sentence production and its disorders to comprehension and acquisition. *Philos. Trans. R. Soc. B Biol. Sci.* 369:20120394. doi: 10.1098/rstb.2012.0394

Devescovi, A., Caselli, M. C., Marchione, D., Pasqualetti, P., Reilly, J., and Bates, E. (2005). A crosslinguistic study of the relationship between grammar and lexical development. *J. Child Lang.* 32:759.

Dijkgraaf, A., Hartsuiker, R. J., and Duyck, W. (2017). Predicting upcoming information in native-language and non-native-language auditory word recognition. *Biling. Lang. Cogn.* 20, 917–930.

Dijkgraaf, A., Hartsuiker, R. J., and Duyck, W. (2019). Prediction and integration of semantics during L2 and L1 listening. *Lang. Cogn. Neurosci.* 34, 881–900.

Döpke, S. (2001). "The interplay between language-specific development and crosslinguistic influence," in *Cross-Linguistic Structures in Simultaneous Bilingualism*, Vol. 21, ed. S. Döpke (Amsterdam: John Benjamins Publishing).

Dussias, P. E., Kroff, J. R. V., Tamargo, R. E. G., and Gerfen, C. (2013). When gender and looking go hand in hand: grammatical gender processing in L2 Spanish. *Stud. Second Lang. Acquisit.* 35, 353–387. doi: 10.1017/s0272263112000915

Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., and Reznick, J. S. (2007). *MacArthur-Bates Communicative Development Inventories: User's Guide and Technical Manual*. Baltimore, MD: Brookes.

Gambi, C., and Pickering, M. J. (2013). Prediction and imitation in speech. *Front. Psychol.* 4:340. doi: 10.3389/fpsyg.2013.00340

Gollan, T. H., Montoya, R. I., Cera, C., and Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: aging, bilingualism, and the weaker links hypothesis. *J. Mem. Lang.* 58, 787–814. doi: 10.1016/j.jml.2007.07.001

Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain Lang.* 36, 3–15. doi: 10.1016/0093-934x(89)90048-5

Grüter, T., Lew-Williams, C., and Fernald, A. (2012). Grammatical gender in L2: a production or a real-time processing problem? *Sec. Lang. Res.* 28, 191–215. doi: 10.1177/0267658312437990

Hansen, P., Łuniewska, M., Simonsen, H. G., Haman, E., Mieszkowska, K., Kołak, J., et al. (2019). Picture-based vocabulary assessment versus parental questionnaires: a cross-linguistic study of bilingual assessment methods. *Intern. J. Biling.* 23, 437–456.

Hansen, P., and Simonsen, H. G. (2016). *PaBiQ, Norsk Versjon: Spørreskjema Om Flerspråklige Barn*. Oslo, NO: Universitet i Oslo.

Hintz, F., Meyer, A. S., and Huettig, F. (2017). Predictors of verb-mediated anticipatory eye movements in the visual world. *J. Exper. Psychol. Learn. Mem. Cogn.* 43:1352. doi: 10.1037/xlm0000388

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., and Van de Weijer, J. (2011). *Eye Tracking: A Comprehensive Guide to Methods and Measures*. Oxford: Oxford University Press.

Hopp, H. (2015). Semantics and morphosyntax in predictive L2 sentence processing. *Intern. Rev. Appl. Linguist. Lang. Teach.* 53, 277–306.

Huettig, F. (2015). Four central questions about prediction in language processing. *Brain Res.* 1626, 118–135. doi: 10.1016/j.brainres.2015.02.014

Huettig, F., Quinlan, P. T., McDonald, S. A., and Altmann, G. T. M. (2006). Models of high-dimensional semantic space predict language-mediated eye movements in the visual world. *Acta Psychol.* 121, 65–80. doi: 10.1016/j.actpsy.2005.06.002

Hulk, A. (2001). "Non-selective access and activation in child bilingualism: the syntax," in *Cross-Linguistic Structures in Simultaneous Bilingualism*, Vol. 21, ed. S. Döpke (Amsterdam: John Benjamins Publishing).

Ito, A., Dunn, M., and Pickering, M. (2017). Effects of language production on prediction: word vs. picture visual world study. *Poster Presented at Architectures and Mechanisms of Language Processing*, Tokyo.

Kaan, E. (2014). Predictive sentence processing in L2 and L1: what is different? *Linguist. Approach. Biling.* 4, 257–282.

Karaca, F., Brouwer, S., Unsworth, S., and Huettig, F. (2021). "Prediction in bilingual children: the missing piece of the puzzle," in *Prediction in Second Language Processing and Learning*, eds E. Kaan and T. Grüter (Amsterdam: Benjamins).

Kristoffersen, K., and Simonsen, H. (2012). *Tidlig Spårkutvikling hos Norske Barn: MacArthur-Bates Foreldrerapport for Kommunikativ Utvikling*. Oslo: NOVUS Forlag.

Lemmerth, N., and Hopp, H. (2019). Gender processing in simultaneous and successive bilingual children: cross-linguistic lexical and syntactic influences. *Lang. Acquisit.* 26, 21–45.

Lemoine-Lardennois, C., Alahyane, N., Tailhefer, C., Collins, T., Fagard, J., and Doré-Mazars, K. (2016). Saccadic adaptation in 10–41 month-old children. *Front. Hum. Neurosci.* 10:241. doi: 10.3389/fnhum.2016.00241

Lew-Williams, C., and Fernald, A. (2007). Young children learning Spanish make rapid use of grammatical gender in spoken word recognition. *Psychol. Sci.* 18, 193–198. doi: 10.1111/j.1467-9280.2007.01871.x

Lew-Williams, C., and Fernald, A. (2010). Real-time processing of gender-marked articles by native and non-native Spanish speakers. *J. Mem. Lang.* 63, 447–464. doi: 10.1016/j.jml.2010.07.003

MacWhinney, B. (2008). "A unified model," in *Handbook of Cognitive Linguistics and Second Language Acquisition*, eds P. Robinson, C. Nick, and Ellis (Hillsdale, NJ: Lawrence Erlbaum Press), 343–371.

MacWhinney, B. (2012). "The logic of the unified model," in *Handbook of Second Language Acquisition*, eds M. Susan, Gass, and Alison Mackey (New York, NY: Routledge), 211–227.

Mani, N., Daum, M. M., and Huettig, F. (2016). "Proactive" in many ways: Developmental evidence for a dynamic pluralistic approach to prediction. *Q. J. Exper. Psychol.* 69, 2189–2201. doi: 10.1080/17470218.2015.1111395

Mani, N., and Huettig, F. (2012). Prediction during language processing is a piece of cake—But only for skilled producers. *J. Exper. Psychol. Hum. Percept. Perform.* 38:843. doi: 10.1037/a0029284

Martin, C. D., Branzi, F. M., and Bar, M. (2018). Prediction is production: the missing link between language production and comprehension. *Sci. Rep.* 8, 1–9. doi: 10.1038/s41598-018-19499-4

Meir, N., Parshina, O., and Sekerina, I. A. (2020). "The interaction of morphological cues in bilingual sentence processing: an eye-tracking study," in *Proceedings of the 44th Annual Boston University Conference on Language Development*, eds M. Brown and A. Kohut (Sommerville, MA: Cascadilla Press), 376–389.

Ness, T., and Meltzer-Asscher, A. (2021). Rational adaptation in lexical prediction: the influence of prediction strength. *Front. Psychol.* 12:735849. doi: 10.3389/fpsyg.2021.735849

Pavlenko, A. (2009). "Conceptual representation in the bilingual lexicon and second language vocabulary learning," in *The Bilingual Mental Lexicon: Interdisciplinary Approaches*, ed. A. Pavlenko (Bristol: Multilingual Matters), 125–160.

Pearson, B. Z., Fernández, S. C., and Oller, D. K. (1993). Lexical development in bilingual infants and toddlers: comparison to monolingual norms. *Lang. Learn.* 43, 93–120.

Pickering, M. J., and Gambi, C. (2018). Predicting while comprehending language: a theory and review. *Psychol. Bull.* 144:1002. doi: 10.1037/bul0000158

Pickering, M. J., and Garrod, S. (2013). An integrated theory of language production and comprehension. *Behav. Brain Sci.* 36, 329–347.

R Core Team (2021). *R: A Language and Environment for Stastistical Computing [Computer Software]*. Vienna: R Fundation for Statistical Computing.

Rstudio Team (2021). *Rstudio: Integrated Development for R. [Computer Software]*. Boston, MA: Rstudio Team.

Salverda, A. P., Kleinschmidt, D., and Tanenhaus, M. K. (2014). Immediate effects of anticipatory coarticulation in spoken-word recognition. *J. Mem. Lang.* 71, 145–163. doi: 10.1016/j.jml.2013.11.002

Stone, K., Lago, S., and Schad, D. J. (2020). Divergence point analyses of visual world data: applications to bilingual research. *Biling. Lang. Cogn.* 1–9. doi: 10.1017/S1366728920000607

The GIMP Team (2018). Gimp 2.8. [Computer Software]. Available online at: https://www.gimp.org

Trueswell, J. C. (2008). Using eye movements as a developmental measure within psycholinguistics. *Lang. Acquis. Lang. Disord.* 44:73.

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

frontiers in Communication

ORIGINAL RESEARCH
published: 09 December 2021
doi: 10.3389/fcomm.2021.703351

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

Kyla McConnell and Alice Blumenthal-Dramé*

English Department, University of Freiburg, Freiburg, Germany

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

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

Edited by:
Phillip M. Alday,
Max Planck Institute for Psycholinguistics, Netherlands

Reviewed by:
Jeroen van Paridon,
University of Wisconsin-Madison, United States
Chiara Celata,
University of Urbino Carlo Bo, Italy

****Correspondence:***
Kyla McConnell
kyla.mcconnell@anglistik.uni-freiburg.de

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Communication

Received: *30 April 2021*
Accepted: *03 November 2021*
Published: *09 December 2021*

Citation:
McConnell K and Blumenthal-Dramé A (2021) Usage-Based Individual Differences in the Probabilistic Processing of Multi-Word Sequences. Front. Commun. 6:703351. doi: 10.3389/fcomm.2021.703351

INTRODUCTION

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

A myriad of studies confirm that comprehenders are sensitive to word-level distributional patterns such as word frequency (see, for example, Gries and Divjak, 2012). Experimental evidence shows that these statistical constraints can be utilized in the service of prediction on diverse levels of language comprehension, including the first sound of an upcoming word (DeLong et al., 2005; but see ; Nieuwland et al., 2018; Yan et al., 2017), the grammatical gender of an upcoming word (Wicha et al., 2003), the semantic features of expected lexical items (Federmeier and Kutas, 1999), and specific words that fit sentence-level context (Otten and Van Berkum, 2008). Strong views of prediction in language processing, such as the view marshalled by the "predictive coding" framework, posit that the brain's central function is to predict across all types of sensory input, from completing a friend's sentence to driving down a familiar road. According to this view, the mind predicts constantly and sensory input is largely ignored unless it disconfirms these predictions (Clark, 2013; Hohwy, 2013; Lupyan and Clark, 2015).

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

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

Experience-Based Individual Differences

Adult native speakers were long thought to achieve a certain ideal attainment, in which they had a complete and uniform understanding of their native language. However, usage-based approaches posit that human skills are highly plastic and shaped by experience (Evans and Levinson, 2009; Dąbrowska, 2015; Dąbrowska, 2018). Recent research has highlighted that language attainment within adult native speakers is modulated by both endogenous constraints (e.g., executive functions, statistical learning abilities, personality traits) and exogenous, experience-related variables (notably, the quality and quantity of the input) (Andringa and Dąbrowska, 2019; Dąbrowska, 2019; Frost et al., 2019; Medimorec et al., 2019; Divjak and Milin, 2020; Kidd and Donnelly, 2020; Ryskin et al., 2020; Kerz and Wiechmann, 2021).

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

Reading Experience

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

Although more proficient readers show reduced word frequency effects, more reading experience seems to pattern with heightened prediction effects (Kukona et al., 2016; Favier et al., 2021). Populations with smaller vocabularies, such as non-native speakers and low literates, rely less on predictions (Mishra et al., 2012; Huettig and Mani, 2015; Pickering and Gambi, 2018). Children's predictive ability in comprehension covaries with their language production ability, which may be linked to their experience with language overall (Mani and Huettig, 2012). Domain-specific text experience can also encourage prediction; job-seekers have lower voice onset times in predictively completing job-related multi-word sequences ("good communication…") (Verhagen et al., 2018). Neurolinguistic studies also support the idea that increased reading experience leads to greater sensitivity to predictability, though the effect seems to be sensitive to task demands, especially time pressure (Ng et al., 2017; Tabullo et al., 2020).

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

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

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

Age

Age has widely been assumed to have a modulatory effect on predictive processing; however, the direction of this effect is still under debate (Gordon et al., 2016; Dave et al., 2018; Payne and Silcox, 2019). On the one hand, older adults generally have greater *crystallized intelligence* (i.e., knowledge accumulated throughout the lifespan), which is comprised of non-linguistic world knowledge, vocabulary (Brysbaert et al., 2016; Sánchez-Izquierdo and Fernández-Ballesteros, 2021), schematic or generalized representations of common occurrences (Ghosh and Gilboa, 2014), and more entrenched distributional knowledge (such as which units of language typically co-occur in language use) (Ramscar et al., 2014; Whitford and Titone, 2019). It has been hypothesized that older adults may rely on their superior crystallized knowledge to engage more readily in linguistic prediction, possibly as a strategy to compensate for perceptual (auditory, visual) decline or for age-related slowing (for review, see Gordon et al., 2016; Payne and Silcox, 2019).

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

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

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

In summary, we expected healthy older people to show heightened effects of lexical-level association patterns than their younger counterparts, independent of their reading experience. This effect could be driven by two different sources: probabilistic processing may be relied on more heavily in cases where the processor has sensory processing deficits or additional experience may simply lead to more efficient language processing in general.

Transition Probabilities in Language Learning and Processing

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

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

Moreover, empirical research has demonstrated that bigrams with lower TPs are more effortful to process, as measured in terms of processing speed, accuracy and brain activation (McDonald and Shillcock, 2003a; McDonald and Shillcock, 2003b; Boston et al., 2008; Demberg and Keller, 2008; Levy, 2008; Frank and Bod, 2011; Frank, 2013; Smith and Levy, 2013; Frank et al., 2015; Hale, 2016; Willems et al., 2016; Lopopolo et al., 2017; Nelson et al., 2017; Lowder et al., 2018). High TPs have also been shown to support language comprehension under challenging speech or reading circumstances, arguably because they allow for the top-down activation of missing or degraded lower-level components (Lorenz and Tizón-Couto, 2019). Likewise, it has been demonstrated that partial input primes high- but not low-TP completions (e.g., *govern* primes *government* more than it primes *governor*, because of higher TPs between *govern*- and -*ment* than between *govern*- and -*or*) (Blumenthal-Dramé, 2017; Blumenthal-Dramé et al., 2017). These different strands of empirical research, along with the inherent forward-looking directionality of FTP, have been taken to suggest that FTP taps into predictive online processing (for discussion, see Aurnhammer and Frank, 2019). BTP, by contrast, does not tap into prediction processes, which by definition must occur before the sensory signal has been encountered. Rather, BTP is likely to correlate with the ease of retrodiction or integration processes, which connect a physically present sensory signal to the previous input to generate a unified and coherent mental representation (Ferreira and Chantavarin, 2018; van Paridon and Alday, 2020; Onnis et al.).

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

MATERIALS AND METHODS

Stimuli

Bigram Extraction

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

Bigrams were rejected if they were emotionally arousing (violent, sexual, or religious) or specialist terms from restricted domains like sports or medicine, so pairs like "primal screams" and "canonical coefficients" were not eligible. Additionally, only bigrams containing one lexical (base) morpheme and up to two derivational and inflectional morphemes were permitted, because words made up of more than one lexical morpheme could themselves be seen as collocations at the morphological level ("widespread," "lifestyles"). Similarly, words beginning with removable prefixes were not included ("inexperienced," "unusual"). Both words in the bigram also had to have consistent spelling in both British and American English and be less than 12 letters long.

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

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

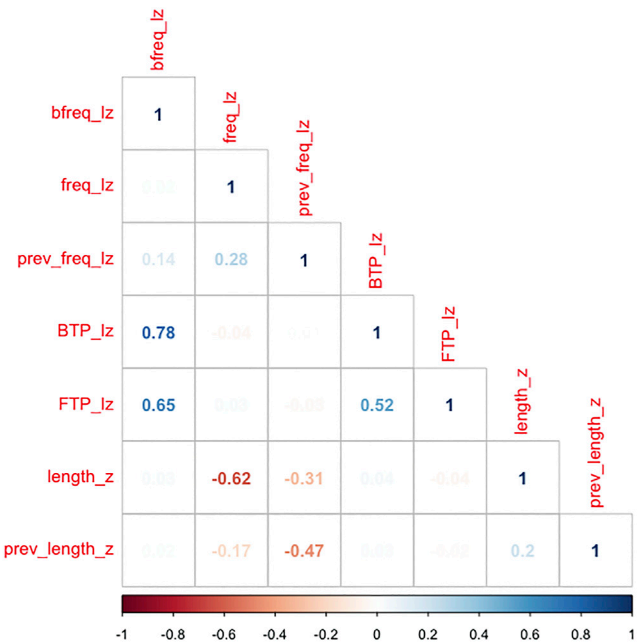


FIGURE 1 | Correlation between stimuli characteristics (Spearman's ρ): log-transformed bigram frequency (bfreq_lz), log-transformed current word frequency (freq_lz), log-transformed previous word frequency (prev_freq_lz), log-transformed backward transition probability (BTP_lz), log-transformed forward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

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

Bigram Pairing

Of course, word associations and other lexical-level factors are not the only force affecting lexical processing; n-gram (or chunk) frequency is also a major factor (Lorenz and Tizón-Couto, 2019; Supasiraprapa, 2019). Usage-based approaches posit that the comprehender does not access, concatenate, or integrate the component words of high-frequency n-grams but rather retrieves chunks of varying sizes holistically (Blumenthal-Dramé, 2017; Ambridge, 2020; Havron and Arnon, 2021). Further, higher n-gram frequency correlates with greater speed and accuracy in comprehension, production, and acquisition regardless of other lexical-level factors, and this effect is consistently found in self-paced reading paradigms (McConnell and Blumenthal-Dramé, 2019).

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

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

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

Sentences

Critical bigrams were embedded in sentence onsets that were designed to be equally plausible with any of the four matched bigrams, so that onsets could be counterbalanced across participants. Critical words were followed by three-word spillover regions that were held as consistent as possible over the two pairs to maximize comparability. These spillover regions were followed by at least one further word to catch sentence-end wrap-up effects. See **Table 1** for example stimuli. Each participant saw one list (A or B), thus participants did not

[3]FTP = W1 frequency/bigram frequency; BTP = W2 frequency/bigram frequency

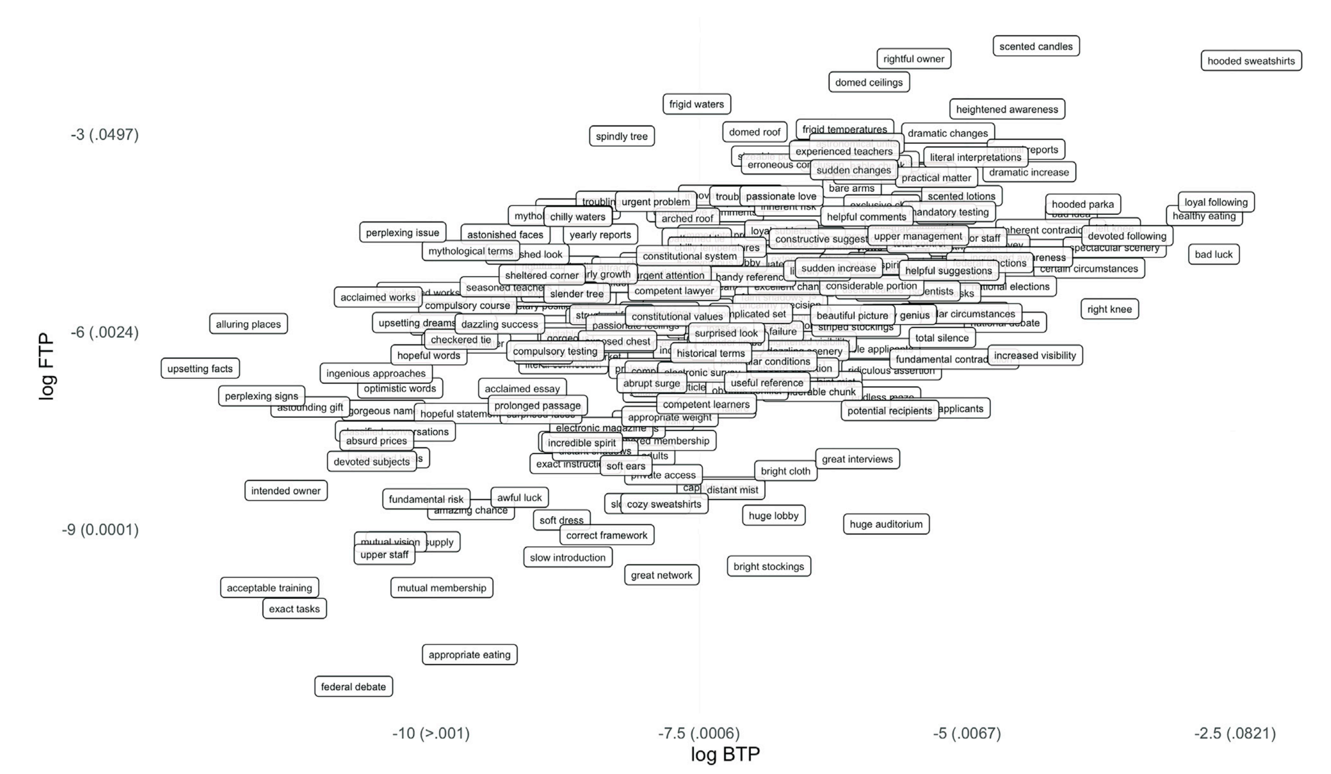


FIGURE 2 | Bigrams by log-transformed forward transition probability and log-transformed backward transition probability (approximate raw values in parentheses).

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

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

read highly similar sentences. Similar sentences (i.e., both labeled A1) were spaced maximally within a set, and shuffled in order across participants.

Sentence onsets were created to be maximally semantically neutral to avoid priming or prediction before the reader encountered the first word in the bigram (the modifier). To prevent uncontrolled semantic relations or priming, role nouns such as what? were excluded from sentence heads—instead, proper names, group nouns, or occasionally, inanimate subjects were used. Names were not repeated and the repetition of other lexical items was also restricted except when it came to highly frequent items (e.g., "little"). Sentences were further designed to have the critical bigram in different semantic and syntactic roles, to reduce predictability of stimuli. This means that in some items, the bigram took the role of the subject, in others it was the direct or indirect object, and in others still, it functioned as part of a fronted subordinate clause. Sentences varied in length from 8 to 21 words (median: 13) and did not contain words with regional spellings.

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

Self-Paced Reading

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

Participants were instructed to read as normally as possible and were informed that the sentences were not related to each other. They were told to answer questions with the best answer,

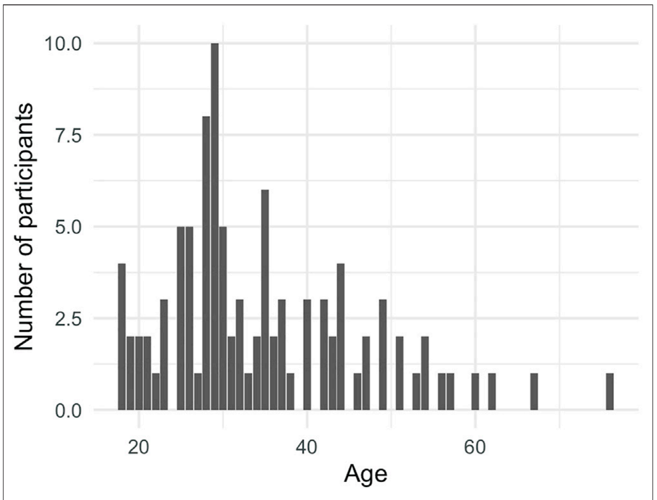


FIGURE 3 | Age of participants (self-reported).

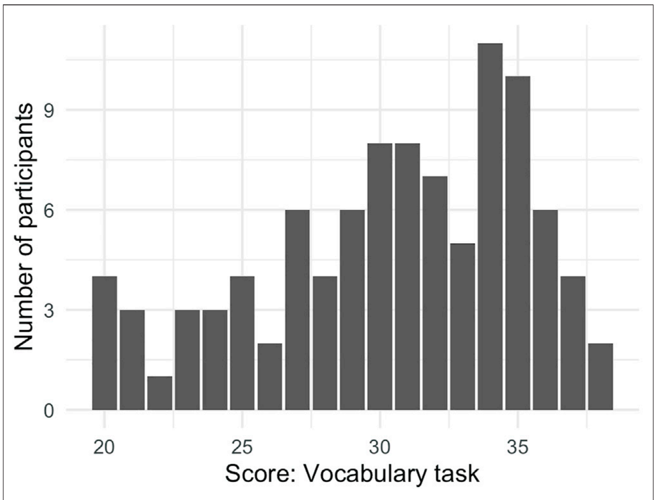


FIGURE 4 | Distribution of scores on the vocabulary task.

and to keep short breaks (such as drinking water) to the question screens, not the reading sections. The reading section was estimated to take about 45 min.

Participants

Participants were recruited online *via* Prolific (Damer and Bradley, 2018). Submissions for which less than 80% of SPR comprehension questions were answered correctly were not considered in the analysis. Recruitment stopped when 100 participants with useable submissions were reached; however, only 97 participants were used in the current study because two did not complete the vocabulary task and one received a score of only 28%. 56 participants were female. Education was measured on a 4-point scale: 19 participants had at least some high school education, 21 had attended technical or trade school, 43 had a bachelor's degree and 14 had an advanced degree (Master's or Ph.D.). All participants were required to be monolingual native speakers of English; 60 participants were British nationals and 37 were United States nationals[4].

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

Individual Difference Assessments

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

Vocabulary Task

To assess vocabulary, participants completed a timed version of the Shipley Institute of Living scale, in which they had to select the best synonym for a given word (abbrv. "vocab" below) (Shipley, 1940)[5]. For example, participants were asked to select a word that meant the same as "fortify" from the set: "strengthen," "submerge," "vent", "deaden". Participants were told not to use any other browser tabs and were presented the questions in a randomized order with 8 s to select the best option. Participant scores were calculated as the number of sets for which the correct synonym was selected; they ranged from 20 to 38 correct answers out of 40 questions (median: 31), see **Figure 4**.

Author Recognition Task

Similarly, participants completed an author recognition task (abbrv. ART), a commonly used assessment of reading amount that circumvents inflated scores that tend to arise from a direct question. By asking participants to identify authors, but not directly asking how much they read, those who see reading as a favorable pastime, but who do not read often themselves, are less able to inflate their own scores.

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

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

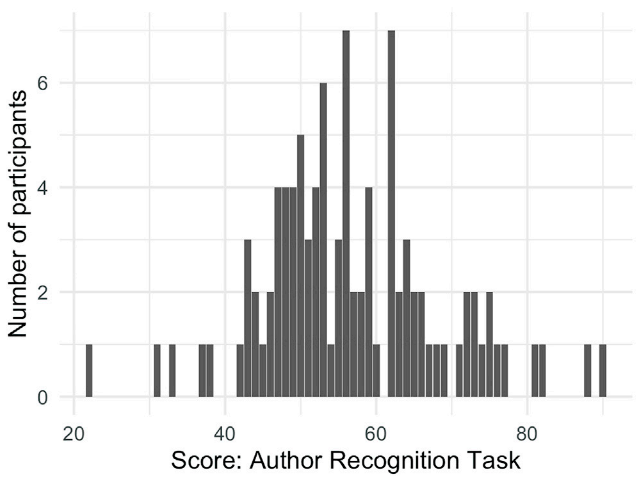


FIGURE 5 | Distribution of scores on the Author Recognition task.

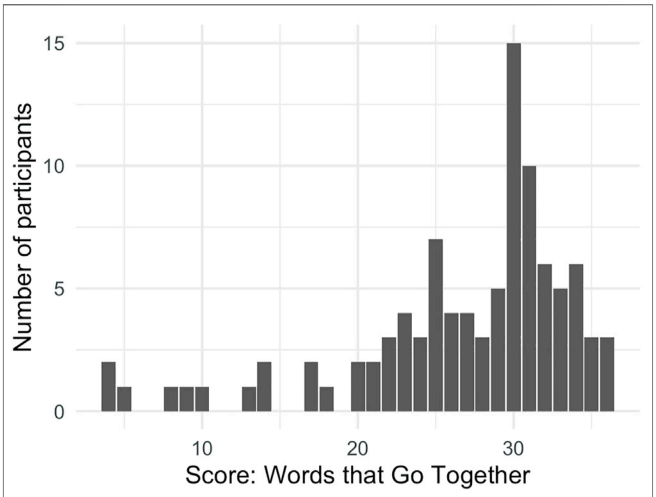


FIGURE 6 | Distribution of scores on the Words that Go Together task.

However, it should be kept in mind that participants from social groups that are well-read or educated may also know more authors names without having read any of the texts from those authors. In the current experiment, the ART task contained 65 authors names from Acheson et al., 2008, an update on Stanovich and West, 1989. Though Acheson and colleagues tested the list on a United States undergraduate audience, the median score across all participants in the United Kingdom was 7.5 points higher in our sample (57.5 vs. 50.0). Because many of the real author names were recognized by less than 10% of the undergraduates tested, just 40 additional foils were taken from Mar et al. (2006), which had a similarly structured task and openly published these names. Participants scores were calculated as the total number of correct answers minus a point for each incorrect guess and ranged from 22 to 90 (median: 56), see **Figure 5**.

Words That Go Together Task

Finally, participants were also presented the Words that Go Together task introduced in Dąbrowska, 2015 (abbrv. WGT). The task had 40 multiple-choice questions, each of which contained five collocated bigrams, where the participant was asked to select the pair that "sounds the most natural or familiar." Participants were further instructed to guess if they didn't know which pair sounded best. Sets varied in frequency; one example is: "hazard a guess," "chance a guess," "dare a guess," "gamble a guess," "risk a guess". Participant's scores were the number of pairs for which they chose the most highly collocated pair and ranged from 4 to 36 correct answers (median: 29), see **Figure 6**. Across median scores by country, participants from the United Kingdom scored five points higher than United States participants in our sample (30 vs. 25). However, since the participants from the United Kingdom scored higher on the US-based ART, this does not seem concerning in terms of a bias of the current task towards United Kingdom Englishes.

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

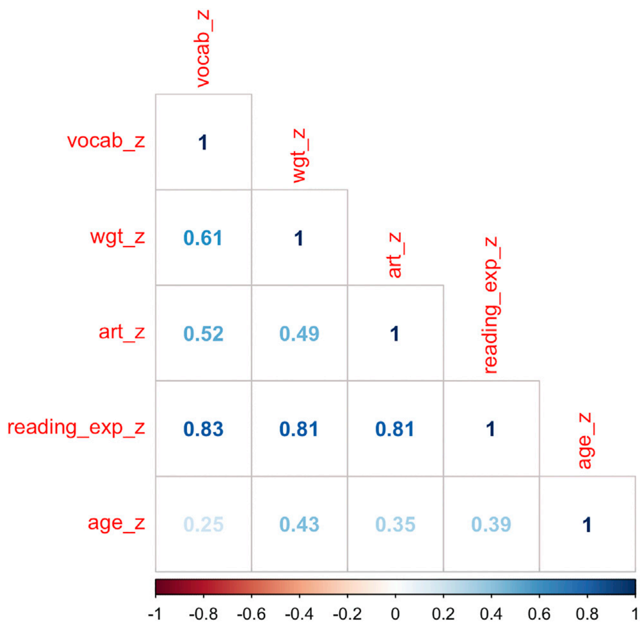


FIGURE 7 | Correlation (Spearman's ρ) of scores on the vocabulary task (vocab_z), the Words that Go Together task (wgt_z), the author recognition task (art_z), the overall reading experience metric (reading_exp_z), and age (age_z).

Modeling

Data was first prepared in R: in this stage, reading times were joined to frequency data extracted from COCA and word and bigram frequencies were log-transformed. RTs below 100 m or (subsequently) outside of 2.5 standard deviations of each participant's individual mean were removed[6] (total: 2.2%). RTs were log-transformed and then centered and scaled. All other numeric variables were also centered and scaled. The prepared data was then loaded into Julia and analyzed with linear mixed effects models using MixedModels.jl (Bates et al., 2021) and plotted in R with the ggeffects package via JellyMe4.jl.

The critical word was defined as the noun in the bigram, because at this point, participants were able to recognize the completed bigram. Our previous research showed that an association effect develops over the spillover region (the critical word and the two following words), so two words following the noun were also included and the variable "position" was effects coded with three levels (McConnell and Blumenthal-Dramé, 2019).

Random slopes were calculated on three grouping variables: participant, the first word in the bigram (W1) and the second word in the bigram (W2). Grouping on W1 pairs bigrams paired items based on the automatic frequency-matched bigram extraction. Grouping on W2 pairs bigrams allowed for bigrams to vary based on semantic similarity (see **Table 1**). To control for commonly observed self-paced reading effects, trial number, word length[7] and (effects-coded) position in sentence were included as covariates. We additionally control for level of education, which was Helmert coded as a factor with four levels.

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

RESULTS

For each model, we fit a four-way interaction between age, reading experience, position, and FTP/BTP, as we hypothesized that the effect of experience may change over the critical region as a factor of these two types of experience. However, no three- or four-way interactions were statistically significant.

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

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

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

DISCUSSION

In this self-paced reading experiment, we investigated probabilistic prediction and integration as parallel mechanisms in language processing, operationalized in terms of forward (FTP) and

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

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

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

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

Linear mixed model fit by maximum likelihood

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

logLik	−2 logLik	AIC	AICc	BIC		
−2001.35	4,002.70	4,098.70	4,098.77	4,540.67		
Variance components						
		Variance	**Std.Dev.**	**Corr.**		
id	(Intercept)	0.0723	0.2689			
	trial_number_z	0.0037	0.0608	−0.25		
	length_z	0.0005	0.0232	0.65	−0.02	
	prev_length_z	0.0008	0.0278	0.5	0.1	0.53
w2	(Intercept)	0.0002	0.0138			
	trial_number_z	0.0001	0.0093	−0.42		
w1	(Intercept)	0.0004	0.0200			
	trial_number_z	0.0001	0.0082	0.08		
Residual		0.0603	0.2456			
Number of obs: 73,690; levels of grouping factors: 97, 143, 135						
Fixed-effects parameters						
	Coef.	**Std. Error**	**z**	**Pr(>|z|)**		
(Intercept)	5.7113	0.0293	195.0300	< 0.0001		
FTP_lz	−0.0027	0.0018	−1.4600	0.1437		
reading_exp_z	−0.0305	0.0235	−1.3000	0.1947		
age_z	0.0977	0.0235	4.1500	< 0.0001		
position: noun	−0.0080	0.0021	−3.8000	0.0001		
position: spillover_1	0.0158	0.0015	10.4900	< 0.0001		
trial_number_z	−0.1125	0.0063	−17.7800	< 0.0001		
word_number_z	−0.0060	0.0014	−4.2800	< 0.0001		
length_z	0.0123	0.0027	4.6200	< 0.0001		
prev_length_z	0.0266	0.0031	8.4600	< 0.0001		
education: Trade school	−0.0980	0.0312	−3.1400	0.0017		
education: Undergraduate	−0.0008	0.0145	−0.0500	0.9581		
education: Grad school	0.0111	0.0143	0.7800	0.4366		
FTP_lz & reading_exp_z	0.0017	0.0012	1.4400	0.1494		
FTP_lz & age_z	−0.0026	0.0012	−2.2600	0.0240		
reading_exp_z & age_z	0.0019	0.0198	0.1000	0.9229		
FTP_lz & position: noun	0.0059	0.0014	4.3300	< 0.0001		
FTP_lz & position: spillover_1	−0.0008	0.0014	−0.6200	0.5351		
reading_exp_z & position: noun	−0.0102	0.0019	−5.4500	< 0.0001		
reading_exp_z & position: spillover_1	0.0038	0.0016	2.4100	0.0160		
age_z & position: noun	0.0048	0.0019	2.5800	0.0100		
age_z & position: spillover_1	-0.0011	0.0016	−0.6800	0.4980		
FTP_lz & reading_exp_z & age_z	0.0012	0.0010	1.2500	0.2125		
FTP_lz & reading_exp_z & position: noun	0.0008	0.0015	0.5200	0.6029		
FTP_lz & reading_exp_z & position: spillover_1	0.0002	0.0015	0.1600	0.8716		
FTP_lz & age_z & position: noun	0.0007	0.0015	0.4700	0.6389		
FTP_lz & age_z & position: spillover_1	−0.0008	0.0015	-0.5500	0.5828		
reading_exp_z & age_z & position: noun	−0.0008	0.0016	−0.5000	0.6169		
reading_exp_z & age_z & position: spillover_1	−0.0013	0.0013	−0.9800	0.3255		
FTP_lz & reading_exp_z & age_z & position: noun	0.0003	0.0012	0.2300	0.8192		
FTP_lz & reading_exp_z & age_z & position: spillover_1	0.0004	0.0012	0.3100	0.7580		

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

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

We initially considered two explanations for why older readers may process more probabilistically: because they have greater

TABLE 3 | Model output linear mixed effects model with backward transition probability.

Linear mixed model fit by maximum likelihood

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

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

Variance components

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

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

Fixed-effects parameters

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

experience with language and world knowledge in general, or because they are compensating for perceptual decline and cognitive slowing. In isolation, the results for age do not answer this question; however, the effect of reading experience provides insight into the question of whether an increased reliance on distributional knowledge is a luxury, in which case the most experienced readers should be better able to exploit distributional information to facilitate processing, or if it is rather a compensatory mechanism, in which case participants with *less* experience should show greater effects of transition probability.

Although in our sample, greater age and increased reading experience were positively correlated ($\rho = 0.39$, see **Figure 7**), they show effects in different directions. Thus, the two individual difference factors can be disentangled and even have competing effects on language comprehension processes. Interestingly, participants with less reading experience show increased

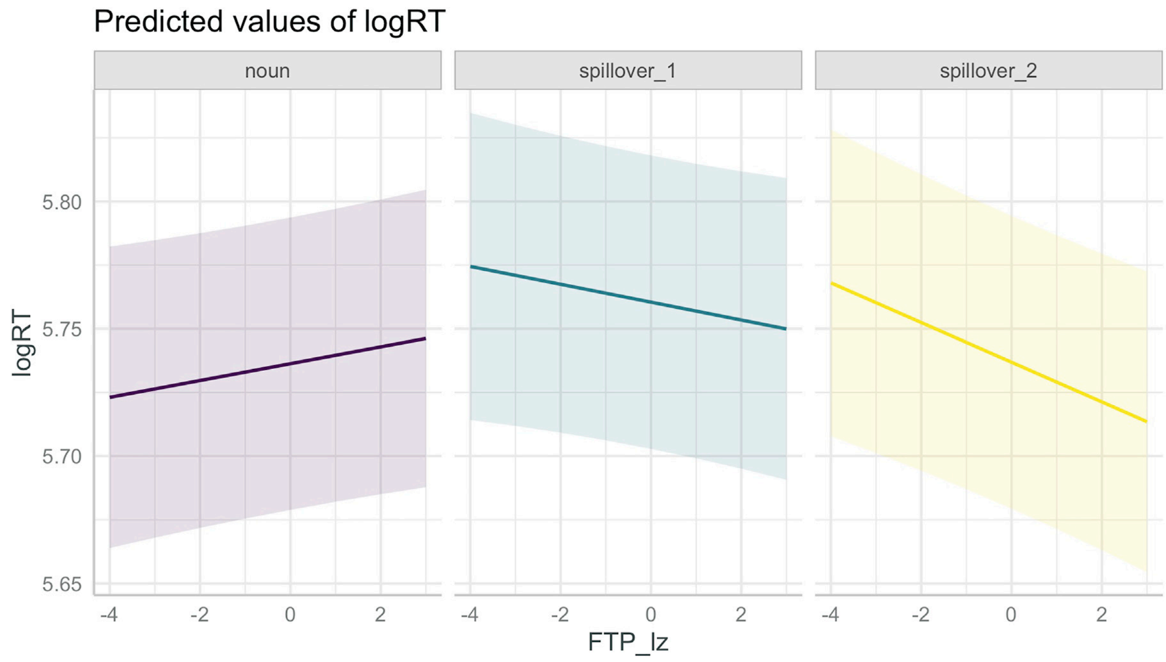


FIGURE 8 | Interaction between (log-transformed, z-scored) forward transition probability and position in the sentence (noun in associated modifier-noun bigram, first and second spillover words).

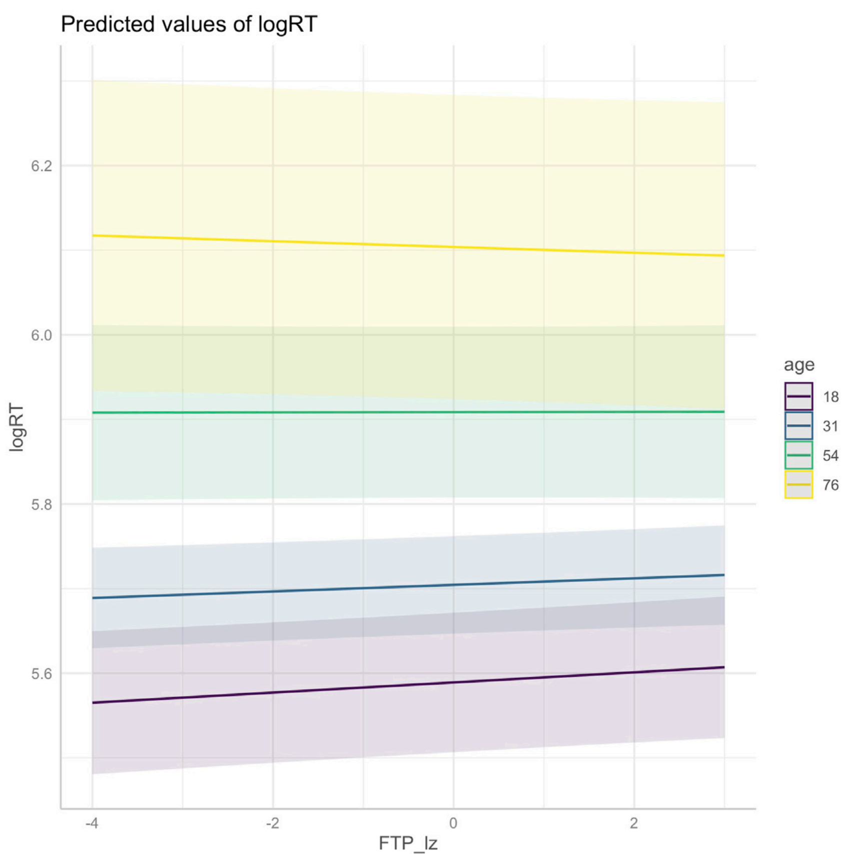


FIGURE 9 | Interaction between (log-transformed, z-scored) forward transition probability and age at sample ages (min: 18, median: 31, max: 76, and midway between median and max: 54).

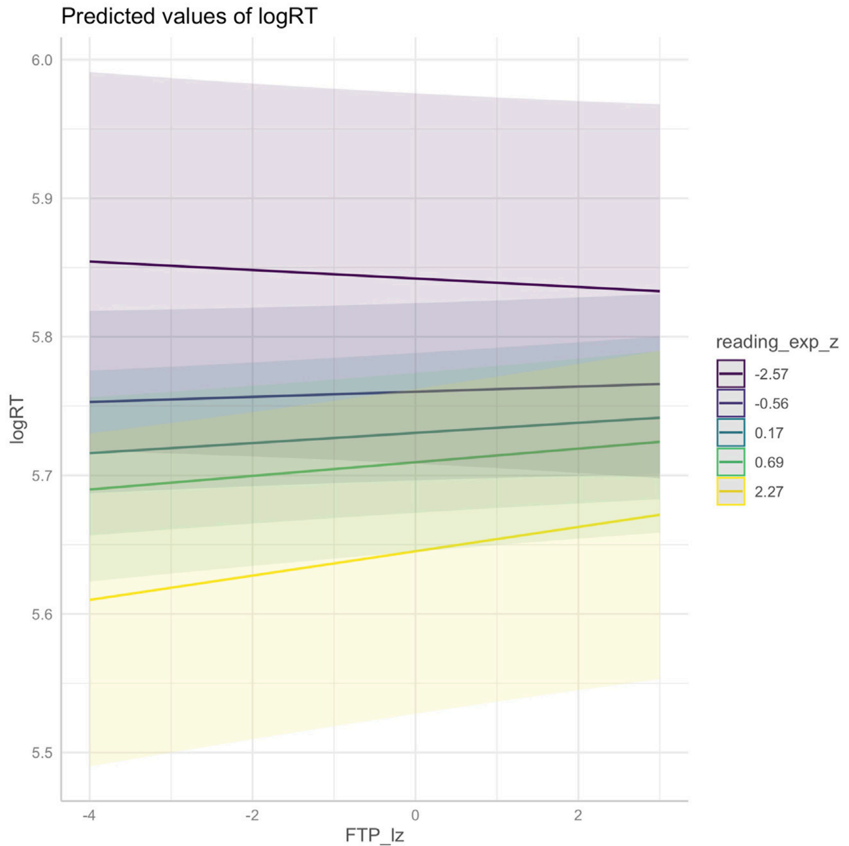


FIGURE 10 | Interaction between (log-transformed, z-scored) forward transition probability and reading experience (at the minimum, first quartile, median, third quartile, and maximum values).

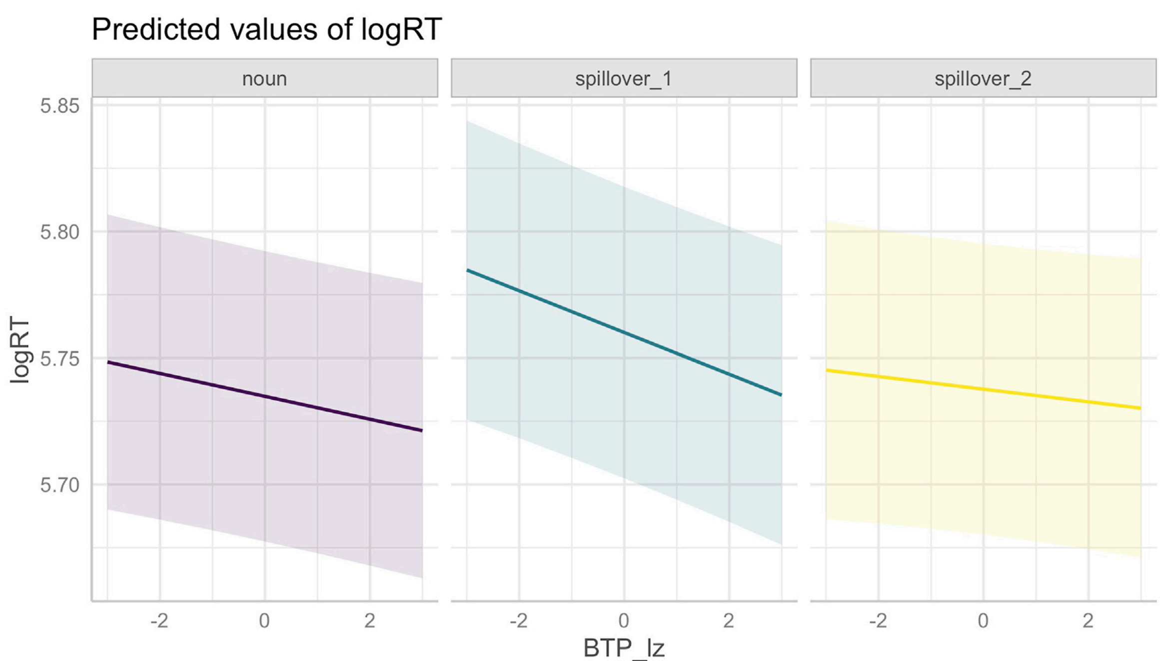


FIGURE 11 | Interaction between (log-transformed, z-scored) backward transition probability and position in the sentence (noun in associated modifier-noun bigram, first and second spillover words).

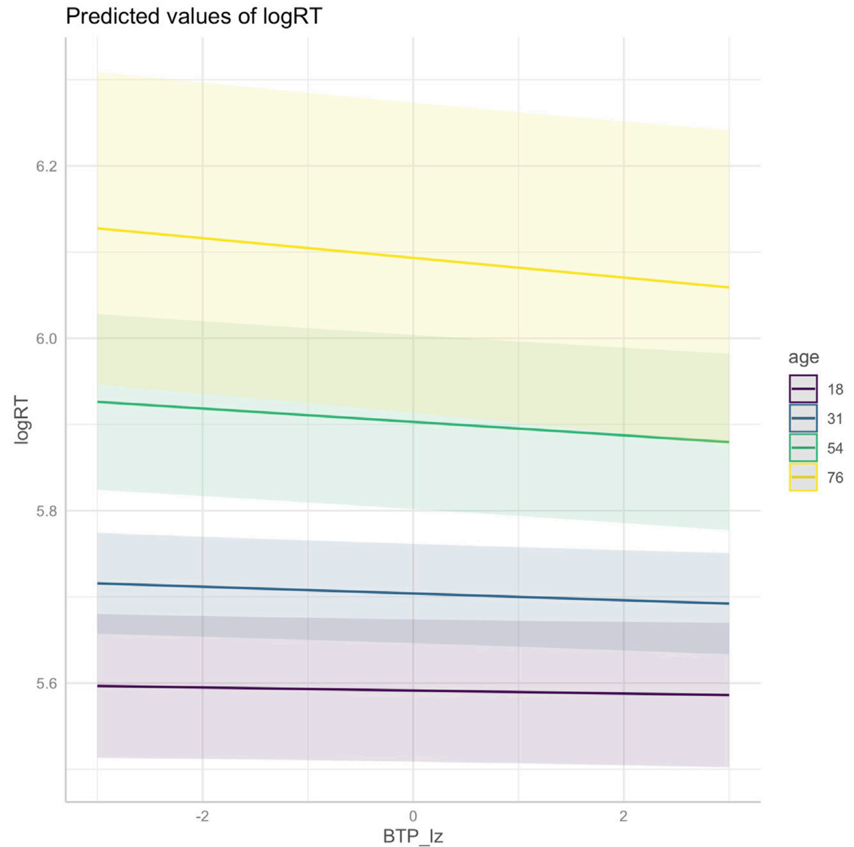


FIGURE 12 | Interaction between (log-transformed, z-scored) backward transition probability and age at sample ages (min: 18, median: 31, max: 76, and midway between median and max: 54).

speeding up effects of BTP and FTP, which seems to contradict the previous research that has found that more reading experience makes a comprehender more predictive (see Section 1.1.1). However, the majority of these studies operationalized prediction via cloze scores, which may rely more heavily on general world knowledge, ability to effectively use context, or other mechanisms (McDonald and Shillcock, 2003a; McDonald and Shillcock, 2003b; Boston et al., 2008, but see Frisson et al., 2005).

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

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

There are of course limitations to the current design. The self-paced reading paradigm, for example, may have affected some groups differentially; more experienced readers,

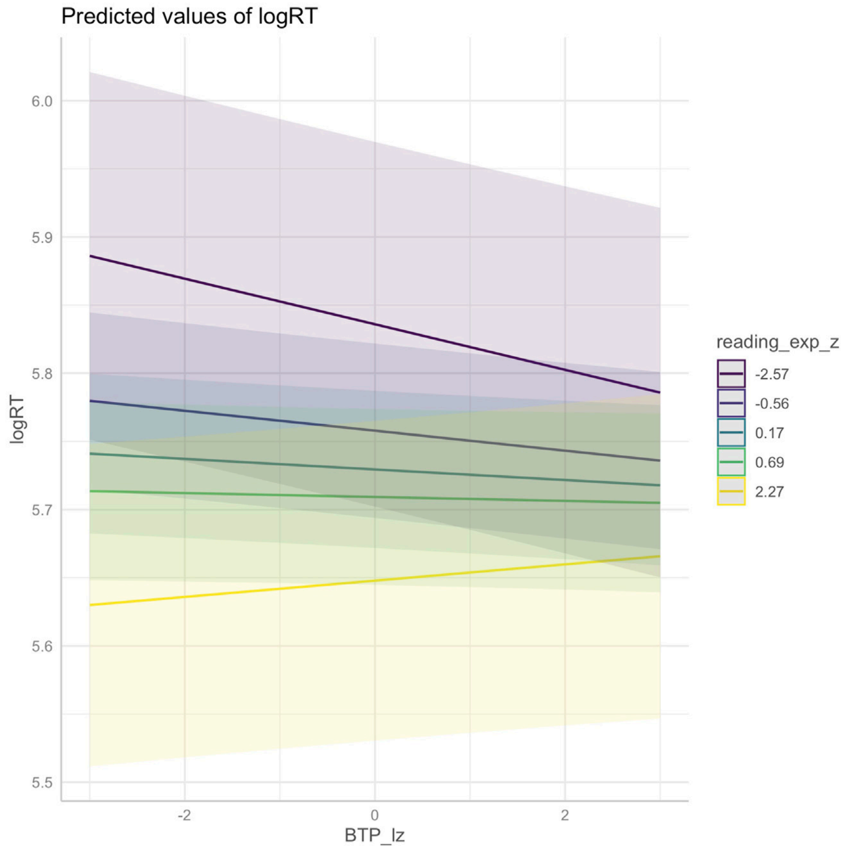


FIGURE 13 | Interaction between (log-transformed, z-scored) backward transition probability and reading experience (at the minimum, first quartile, median, third quartile, and maximum values).

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

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

Follow-up research is encouraged to consider TPs at more fine-grained levels of analysis (letters, phonemes, syllables, etc.), TPs between longer stretches of language [e.g. the predictability of the unit (*helpful discussion*), given the preceding context], and TPs at abstract levels of analysis [e.g., between the syntactic categories (Adj) and (N)] (Celata, 2020). Ultimately, it would also be desirable take into account TPs that cut across levels [e.g. between the morpheme *-ful* and the lexeme *discussion*, or between the lexeme *helpful* and the syntactic category (N)] and even TPs between context and language strings (e.g., how predictable is *helpful discussion* in an academic context?). Of course, this should be conducted from the perspective of

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

individual differences, among the groups posited here as well as those not focused on in this study.

CONCLUSION

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

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

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

AUTHOR CONTRIBUTIONS

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

FUNDING

Participant recruitment was funded by the Wissenschaftliche Gesellschaft in Freiburg im Breisgau. The article processing charge was funded by the Baden-Württemberg Ministry of Science, Research and Art and the University of Freiburg via the funding program Open Access Publishing. This research was supported by a grant from the German National Academic Foundation (Studienstiftung des deutschen Volkes) to KM.

ACKNOWLEDGMENTS

The authors thank Jiří Zámečník for his contributions to the programming of the online experiment, as well as the two reviewers and the editor for their constructive feedback on earlier drafts of the manuscript.

REFERENCES

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

frontiers in Communication

ORIGINAL RESEARCH
published: 16 December 2021
doi: 10.3389/fcomm.2021.779814

Co-Speech Movement in Conversational Turn-Taking

Samantha Gordon Danner[1], Jelena Krivokapić[2] and Dani Byrd[1]*

[1]University of Southern California, Los Angeles, CA, United States, [2]Department of Linguistics, University of Michigan, Ann Arbor, MI, United States

This study investigates co-speech movements as a function of the conversational turn exchange type, the type of speech material at a turn exchange, and the interlocutor's role as speaker or listener. A novel interactive protocol that mixes conversation and (non-read) nursery rhymes works to elicit many speech turns and co-speech movements within dyadic speech interaction. To evaluate a large amount of data, we use the density of co-speech movement as a quantitative measure. Results indicate that both turn exchange type and participant role are associated with variation in movement density for head and brow co-speech movement. Brow and head movement becomes denser as speakers approach overlapping speech exchanges, indicating that speakers increase their movement density as an interruptive exchange is approached. Similarly, head movement generally increases after such overlapping exchanges. Lastly, listeners display a higher rate of co-speech movement than speakers, both at speech turns and remote from them. Brow and head movements generally behave similarly across speech material types, conversational roles, and turn exchange types. On the whole, the study demonstrates that the quantitative co-speech movement density measure advanced here is useful in the study of co-speech movement and turn-taking.

Keywords: turn-taking, multimodal speech, head movement, brow movement, conversational interaction

OPEN ACCESS

Edited by:
Martine Grice,
University of Cologne, Germany

Reviewed by:
James P. Trujillo,
Radboud University Nijmegen, Netherlands
Mathias Barthel,
Humboldt University of Berlin, Germany

****Correspondence:***
Samantha Gordon Danner
sgdanner@usc.edu

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Communication

Received: *19 September 2021*
Accepted: *17 November 2021*
Published: *16 December 2021*

Citation:
Danner SG, Krivokapić J and Byrd D (2021) Co-Speech Movement in Conversational Turn-Taking. Front. Commun. 6:779814. doi: 10.3389/fcomm.2021.779814

INTRODUCTION

The goal of this study is to examine whether and how interacting speakers deploy co-speech movements of the brows and head at speech turn exchanges in a dyadic spoken language interaction. We focus on these movements because they are not directly associated with semantic meaning, and thus might lend themselves to interactional use. We use an interactive, non-read speaking task and a quantitative measure of movement density (velocity peaks per second—first used for co-speech movement in Danner et al. (2018)—to evaluate a large amount of speech turn and kinematic data (Gordon Danner et al., 2021) in addressing our questions.

Experimental linguistics has increasingly attended to an embodied perspective on spoken language interaction. Phonetic research has examined co-speech movements of the hands, head, eyes and facial features (McClave, 2000; Krahmer and Swerts, 2007; Cummins, 2012; Kim et al., 2014; Fuchs and Reichel, 2016) to illuminate prosodic structure, primarily elicited with read speech. A large body of research has examined the informational role of co-speech movement of an individual (Kita and Özyürek, 2003; Özyürek et al., 2005; Gullberg, 2010), but less is known about co-speech movement behaviors in *interactional* contexts (though see e.g., Nota et al., 2021; Trujillo et al., 2021; Duncan Jr, 1972; Latif et al., 2014; Mondada, 2007).

Research on the human capacity for turn-taking in conversation has observed that turn-taking is remarkably fast and flexible (Duncan Jr, 1972; Stivers et al., 2009). The average gap between speakers in spontaneous speech is ~200 ms (Levinson and Holler, 2014; Magyari et al., 2014; Garrod and Pickering, 2015; Roberts et al., 2015) and this gap is stable across a variety of languages and cultures (Stivers et al., 2009). As such, researchers have long been interested in what exactly happens at turns that enables the smooth flow of conversation, and research has focused on the ways that turn-end prediction and next-turn preparation are aided by attention to lexical, syntactic, semantic and prosodic content (De Ruiter et al., 2006; Magyari et al., 2014; Bögels and Torreira, 2015; Garrod and Pickering, 2015; Barthel et al., 2017). However, there are reasons to think that co-speech movement might be relevant for conversational turn-taking, in that listeners are sensitive to gestures and that these movements seem to facilitate comprehension (e.g., Kelly et al., 2010; Holler et al., 2014) and facilitate speech production, specifically by reducing cognitive load and facilitating lexical access (Krauss, 1998; Alibali et al., 2000; Goldin-Meadow et al., 2001; Melinger and Kita, 2007; Gillespie et al., 2014). Studies specifically examining co-speech movements at turn-ends or in an interactional context suggest that co-speech movements likely contribute to effective turn-taking; these include movements of the hands, head, face, eye blinking (Duncan Jr, 1972; Hadar et al., 1985; Mondada, 2007; Barkhuysen et al., 2008; Sikveland and Ogden, 2012; Levitan et al., 2015; Holler et al., 2017; Hömke et al., 2018; Zellers et al., 2019; Trujillo et al., 2021) and gaze (Barkhuysen et al., 2008; Bavelas et al., 2002; Stivers et al., 2009). The most examined of these are manual gestures, and it has been suggested that these contribute to the selection of the next speaker, to indicating the end of the turn of the current speaker, and to soliciting help in the interaction, and there is evidence that listeners respond to these gestures by taking up the offered turn (Bavelas et al., 1995; Duncan Jr, 1972). Two studies have specifically investigated how the timing of turn exchanges is affected by the presence of manual co-speech gestures. Holler et al. (2017) find that, at least in question-response pairs, turn exchanges are faster when the question is accompanied by a co-speech gesture. Trujillo et al. (2021) also examine question-response pairs but separate turn exchanges into overlapping exchanges and non-overlapping exchanges and find that both gaps and overlaps between speakers are shorter when questions are accompanied by gestures.

Our study focuses on head and brow movement. Among co-speech body movements, head movement has received significant attention in studies of communicative interaction (Hadar et al., 1985; Munhall et al., 2004; Krahmer and Swerts, 2007; Ishi et al., 2014). Ishi et al. (2014) find that head-motion type differs according to type of dialogue (for example, more head nods in questions than in turn-giving, see also Kendon, 1972) and that the frequency of some head movements is affected by the relationship between conversation partners. Head movement has been found to be more frequent during speaking than during listening (Hadar et al., 1983), but head nods are also known to be part of a listener's repertoire, e.g., as backchannelling (Duncan Jr, 1972). Listener nods in turn seems to be coupled with speakers' head nods (McClave, 2000) and can indicate a turn-taking request (Hadar et al., 1985). Nods can also be the first indicator of a response, preceding a verbal response (Stivers et al., 2009).

Beyond (whole) head movement, dynamic aspects of facial features are likely to be relevant to turn-taking and communicative interaction. Eyebrow movement has been studied in non-interactive spoken language (Krahmer and Swerts, 2007; Cvejic et al., 2010; Kim et al., 2014). Goujon (2015) finds that brow movements occur more frequently at the beginning of an utterance than elsewhere in the utterance, while Flecha-Garcia (2010) finds some evidence that such movement occurs at the start of hierarchically high discourse units. The frequency of eyebrow movement is also dependent on the speech material, with expression of personal opinions, for example, being related to more eyebrow movements (Goujon et al., 2015), and eyebrow movements also being more frequent in giving instructions than in asking questions (Flecha-García, 2010). Only a few studies have investigated the role of eyebrow movement in interactions. Guaitella et al. find that brow movements are significantly more likely to occur in the immediate vicinity of the initiation of a new speaking turn than elsewhere in conversations, and the authors link this to the speaker's intention to communicate (Guaïtella et al., 2009). Borràs-Comes et al. (2014) find that speakers of Catalan and Dutch use more eyebrow raises in questions than in responses. Similarly, Nota et al. (2021) find more eyebrow movements in questions compared to responses for Dutch and that they occur typically early in the utterance (before the onset of speech). Nota et al. (2021) suggest that this might be in order to allow the interlocuter more time to plan the response.

Many of the previous studies examine manual gestures that have an obviously interpretable communicative function, for example, gestures that are iconic, metaphoric, deictic, or pragmatic, and then relate these to the meaning or function in the utterance. It has also been suggested that the cessation of co-speech gestures functions for the listener as a signal for turn completion (Duncan Jr, 1972; Levinson and Torreira, 2015). Thus, together with information from the acoustic speech signal, co-speech gestures could help in predicting the end of the turn and concomitantly help in timing the onset of the next turn (e.g., Barkhuysen et al., 2008; Stivers et al., 2009; Holler et al., 2017; Nota et al., 2021). The general approach of studying the occurrence and placement of individual gestures, largely with meaningful interpretations, contrasts with our approach in that we examine the broad patterning of general movement density in the neighborhood of an interactional event of interest, namely a floor exchange. While our study does not directly test or model predictability of a floor exchange, by evaluating the patterning of co-speech movement at exchanges we lay the ground for future studies of how co-speech movement contributes to the management of interactions and we offer a new empirical strategy for assessing these complex multimodal signals.

The present study examines *the rate* or *density* of co-speech head nods and eyebrow raises at speech turn exchanges. Specifically, using non-read conversational interactions with robust opportunities for turn-taking, we examine whether the rate of co-speech movement varies as a function of proximity to a turn exchange, the type of speech turn, or the conversational role. Importantly, we examine *any* type of movement, regardless of its function. As will be explained in more detail in the next section, our study uses a measure common in articulatory speech production research, namely the rate/density of movement expressed in velocity peaks per second. This measure was first used by Danner et al. (2018), where it has been shown that distinct varieties of movement are used in different kinds of speech tasks. The benefit of this measure is that it can be extracted almost automatically, allowing for a large database of turn exchanges to be examined. Relatedly, Bavelas et al. (2008, 1992) find that the rate of movement of co-speech gestures differs depending on co-speaker visibility and on whether the speaker was alone or part of a dialogue. These findings suggest that the frequency of co-speech movements may aid the human capacity for efficient turn-taking and conversation and thus is a potentially useful measure for our study.

A variety of patterns around floor exchanges are possible from what little we know from the prior literature. Our hypothesis is that the density of co-speech movements will differ depending on the type of floor exchange—whether having overlapping speech or non-overlapping speech—as compared to speech that is non-exchange-adjacent. This is based in particular on findings that turn-end prediction is facilitated by utterance final (prosodic, syntactic, and lexical) information and on the evidence that co-speech movement differs utterance-finally compared to utterance-medially (Duncan Jr, 1972; Barkhuysen et al., 2008; Bögels and Torreira, 2015). We further examine whether the density of co-speech movement will differ as a function of the interlocuter's immediate role at the exchange, i.e., as "listener" or "speaker". Previous studies have focused mostly on one participant in an interaction, but a number of findings point to different functions of co-speech movement for the listener and for the speaker. Since our study is among the first to examine the co-speech movements of *both* participants in a robust sampling of conversational floor exchanges, specific predictions cannot yet be made as to whether listener and speaker will differ in the density of head and brow movement. For example, a speaker may increase the rate of their co-speech movements to indicate an upcoming turn end or to focus phrase edge material, and this may facilitate a listener's prediction of the end of the turn, thereby facilitating the turn exchange. Furthermore, a listener may increase their movement density as a precursor to interrupting, starting a turn, or as an act of affiliation while listening. These possibilities have motivated the current study of dyads by examining both the speech interval *approaching* a floor exchange—which we call Turn Approach—and the speech interval after the conversational baton changes hands to the other speaker—which we call Turn Receipt.

Our instrumental setup allows us to examine kinematic data for *both* listener and speaker simultaneously. And while most prior research on co-speech movement in interactions has focused on manual gestures (and to a lesser extent on head movement), our study quantifies and examines both whole-head and eyebrow movement. In comparison to the more studied manual gestures, these articulators are less directly associated with semantic meaning (other than the agreement and disagreement of head nods) and may lend themselves to interactional use. The experimental approach of the current study of dyads thus advances a more complete understanding of co-speech movement patterning with the goal of broadening our empirical knowledge of the interactional process taking place between conversing partners.

Strategies for Experimental Design

While motion-capture technology and other tools for detecting movement from video have existed for some years in speech and linguistic research (Levelt et al., 1985; Munhall et al., 1985; Yehia et al., 1998; Barbosa et al., 2008), recent advances have enabled researchers to examine conversational interaction in a variety of novel ways: from empirically quantifying movement in video recordings (Barbosa et al., 2012; Voigt et al., 2016), to directly tracking the kinematics of speech-accompanying movements (Ishi et al., 2014; Kim et al., 2014; Danner et al., 2018), to considering the speech articulator kinematics of two interacting speakers (Scobbie et al., 2013; Lee et al., 2018). The present study combines several of these tools with the goal of examining co-speech movement density at conversational turn exchanges for pairs of speakers in naturalistic, face-to-face conversation. By capitalizing on a dual-magnetometer setup described in the Methods section below (see also previous work from our laboratory e.g. Lee et al., 2018), we are able to collect time-aligned audio and kinematic signals from pairs of conversing speakers. The conversing speakers were seated facing each other, able to see one another's heads, arms/hands and torsos. This experimental setting offers the rare opportunity to collect kinematic data for two interacting speakers in a relatively natural setting, enabling the study of participants in their roles as *both* speakers and listeners during an interaction. While the exchange of these conversational roles has of course been extensively observed (Rochet-Capellan and Fuchs, 2014), the consideration of empirical data for the *co-speech movements* of interacting pairs in conversation, with annotation of conversational role (speaker and listener), has rarely been undertaken with *kinematic data for both participants in a dyadic interaction*.

A further advantage of our experimental protocol is that the rich kinematic data can be analyzed quantitatively, as described in detail in the methods section below. Specifically, using a method developed in Danner et al. (2018), a measure of *movement rate* (*density*) is algorithmically derived from brow and head movement kinematic velocity profiles (see for other uses of velocity profiles: Leonard & Cummins, 2011; Munhall et al., 1985; Ostry et al., 1987). This differs from manually annotating multiple gestural landmarks from video recordings and classifying them according to their communicative function, a technique used in many foundational studies of co-speech movement behavior (see overviews in Danner et al., 2018; Wagner et al., 2014). This method has given invaluable results but it has always, by necessity, been limited to a very small set of data. For example, Bavelas et al. (1995) was based on 88 gestures (selected from a larger corpus of 464 gestures), Holler et al. (2017) examined 281 question response sequences, Loehr (2004) is based on 164s of data and 147 gestures, and seminal work Kendon's (1972) was based on 90 s of data. While our method does not explicitly assess the communicative function of co-speech movement, it automatically detects movement occurrence,

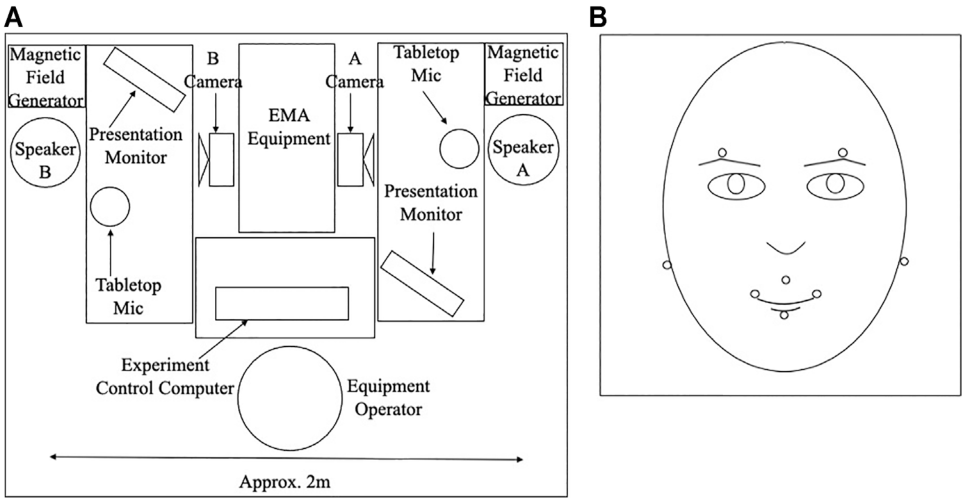


FIGURE 1 | Panel **(A)** shows the experiment room setup; Panel **(B)** shows EMA sensor placement schematic: right/left/center reference sensors, right/left brow sensors, right/left lip corner sensors, and lower lip sensor.

thereby enabling the examination of a much larger set of data (>85 min of speech, 3,110 exchanges and thousands of individual movements) than the earlier gesture annotation method. In order to elicit structured turn-taking that is not read speech, we developed a speech elicitation paradigm in which dyads cooperatively undertake a spoken language task that promotes significant interaction between the interlocutors (see also Danner, 2017; Geluykens & Swerts, 1992; Lee et al., 2018). Crucially, by not relying on reading, our study allows participants the opportunity to interact with one another in a visually engaged way that promotes naturalistic speech *and* co-speech behavior with ample opportunities for floor exchanges. This protocol was achieved by leveraging familiar nursery rhymes in a collaborative task, as described in detail below. While many studies use conversational interactions and non-read speech to examine co-speech movement, the short, easily predicted phrases of the present task provide *many* opportunities for participants to exchange speaking turns both related and unrelated to the nursery rhyme at hand. The prosodic and rhythmic structure of the nursery rhymes, along with the engaging collaborative nature of the task, promote speech-accompanying movements of the brow, head, and hands, as well as non-speech communication like smiling and laughing, as participants cooperate to complete a rhyme; for these reasons this data collection protocol is particularly suitable to examining our question of co-speech movement patterning in the approach and receipt of dyadic floor exchanges.

Taken together, the instrumental set-up utilized within the collaborative speech task allows for a robust quantified view of the speech and co-speech movements that are integral to human conversational interaction. This empirical data alongside the innovative experimental data elicitation paradigm offers a new window for advancing the understanding of the cognitive and linguistic processes that underlie the elegant human ability for conversation.

METHODS

Experiment Design and Stimuli

In order to study how fluctuating conversational conditions might affect co-speech movement behavior, we created a cooperative turn-taking task in which participants were asked to work together to recite fairly well-known English nursery rhymes. Nursery rhymes were selected for use in this task in order to elicit naturalistic interactions between two participants that were conducive to numerous speech turn exchanges. Nursery rhymes were suited to this goal due to their prosodic structure and the fact that, as somewhat familiar speech material, they could (after prompting) be recalled and produced with relative ease without being read. Participants were asked to complete the rhymes by taking turns and helping one another to complete the rhyme if either speaker were to forget the next portion of the rhyme. If the participants working together got irrevocably stuck trying to complete a rhyme, they could decide to give up and move on to the next rhyme, but this occurred quite rarely. These nursery rhymes are suited to the elicitation goals of this project because they are many in number and commonly known, they tend to be short (typically around 15–30 s for a solo production of one well-known verse), and they have a simple rhythmic and phrasal structure and accessible rhyme patterns (Fuchs and Reichel, 2016). It was useful for our elicitation purposes that most native speakers have been exposed to nursery rhymes as children but that the rhymes are not regularly encountered by older children and adults without children. We expected that our participants would have a baseline level of familiarity with nursery rhymes but may not remember a given nursery rhyme in exact detail.

This then provided an excellent opportunity for interaction between the participants, given that they were likely to need each other's help to remember and complete the rhyme or negotiate with each other as to when to give up and move on. It's important

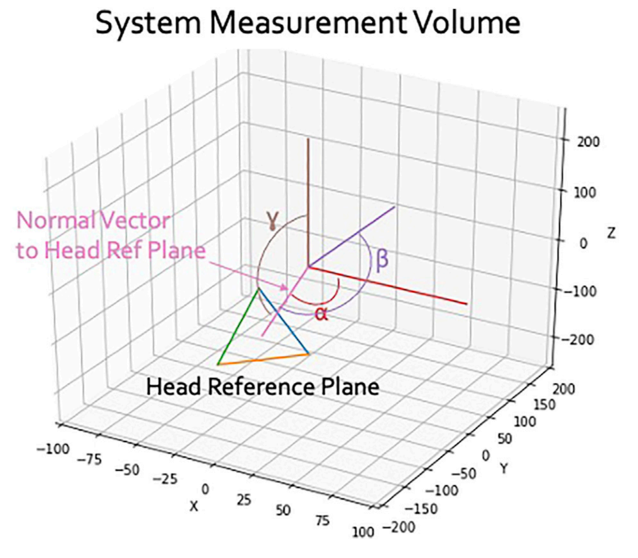


FIGURE 2 | Illustration of the head movement measure.

to note that the recorded conversational interaction for each dyad included a great deal of speech well beyond the production of nursery rhymes themselves, as the participants navigated the task they had been given to cooperate in. This meant that free conversational material—chatting between the speakers—was intermingled with the nursery rhyme production material. In fact a coding of the immediately turn final material (See *Exchange Types, Conversational Roles and Speech Content Types* section below) indicates that roughly 40% of the exchanges were conversational and not strictly nursery-rhyme production. All the speech material in the entire session was included in the analysis below, which is to say that the analyzed material included both free conversation and nursery rhyme production. Critically, the task elicited many floor exchanges with a variety of speech material, as was intended.[1]

To construct the nursery rhyme stimuli set for this experiment, we selected 24 common nursery rhymes found on the website *nurseryrhymes.org* (Granum, 2017). From this database of 202 unique nursery rhymes, we excluded rhymes that are not primarily in English (e.g., *Frère Jacques*), rhymes requiring stylized melodies or "dances" (e.g., *I'm a Little Teapot*), rhymes introduced within the last century (e.g., *Miss Suzy/Hello Operator*), rhymes of a religious nature (e.g., *Now I Lay Me Down to Sleep*), and rhymes longer than three stanzas (e.g., *Little Bunny* Foo). The titles of the 24 remaining rhymes were submitted to the Corpus of Contemporary English (Davies, 2008) and a Google search to norm for frequency of appearance (COCA count frequency ranged between 0 and 134 appearances, mean = 18.46; Google frequency ranged between 99,800–843 M hits; mean = 134.9 M). The two least frequent rhymes, Peter *Pumpkin Eater* and *There Was an Old Woman Who Lived in a Shoe*, were chosen for use as practice trials.

Subjects

Six pairs of previously unacquainted[2] speakers, henceforth *dyads*, participated in the experiment. The first two dyads recorded were used as pilot data to refine our data collection procedure and were excluded from further analysis, leaving four analyzed dyads, henceforth referred to as Dyads 1–4. Dyads 1 and 3 are composed of two female participants, while Dyads 2 and 4 are composed of one male and one female participant. Participants range in age between 19 and 40 (mean age: 27.75) and are native speakers of American English. All participants voluntarily completed the entire experiment, which lasted approximately 1.5–2.5 h, and all participants were naïve to the purpose of the study.

Data Acquisition

Participants were seated at facing desks approximately 2 m apart in a sound-insulated room in the University of Southern California Phonetics Laboratory. Each participant had a Wave (Northern Digital, Inc.) electromagnetic articulography (EMA) system positioned beside their head, a tabletop microphone and a computer monitor on their desk, and a tripod and video camera positioned in front of the desk, angled down toward the speaker. The monitor and microphone were placed to allow each participant an unobstructed view of the other participant's head and upper body. **Figure 1A** shows a schematic of the experiment room setup. Prior to the beginning of the experiment, participants were given the stimulus set of 24 nursery rhymes, each one printed on an individual sheet of paper, and were instructed to have a quick read-through of each rhyme only once, before putting that sheet of paper face-down on their desk. After both participants finished reading the set of nursery rhymes once through, the study personnel removed the papers from the participants' desks and commenced with EMA sensor placement.

Following standard EMA protocol, head reference EMA sensors were adhered externally at participants' left and right mastoid processes and internally on the gum above the upper incisor, using a temporary adhesive. An occlusal plane measure was then taken for each participant, after which study personnel placed the remaining EMA sensors on the lower lip (mid-sagitally on the vermillion border), the right and left brows (placed above the most mobile part of the brow), and on the right and left upper lip corners as close to where the upper and lower lips meet as possible. **Figure 1B** shows the sensor placement schematic. An XML-based Matlab tool for stimulus presentation and

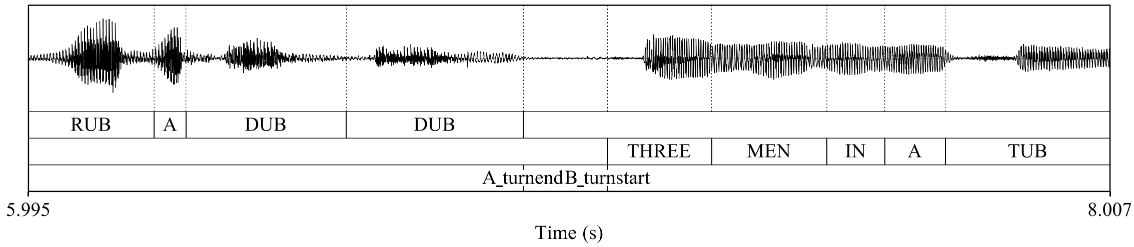
[1]While our task elicits both nursery rhyme material and spontaneous speech, it can be assumed that other types of interaction would give different results, as different tasks and contexts elicit different conversational strategies and different co-speech gesture behavior (see for example Danner et al., 2018; Dideriksen et al., 2019).

[2]Lack of familiarity between participants in a dyad was the only constraint we placed on their pairing. Although speaker age, gender, race or other perceived or real demographic information may affect some aspects of interaction, we have no reason to believe that these details impact the fine-grained movement behavior that is the object of this study, nor was this study designed to probe such myriad socio-linguistic variables.

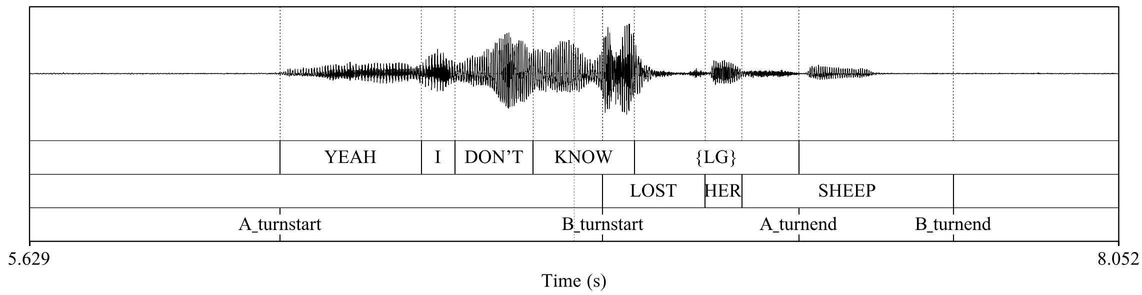
TABLE 1 | Selected examples of floor exchange analysis regions.

Exchange type examples

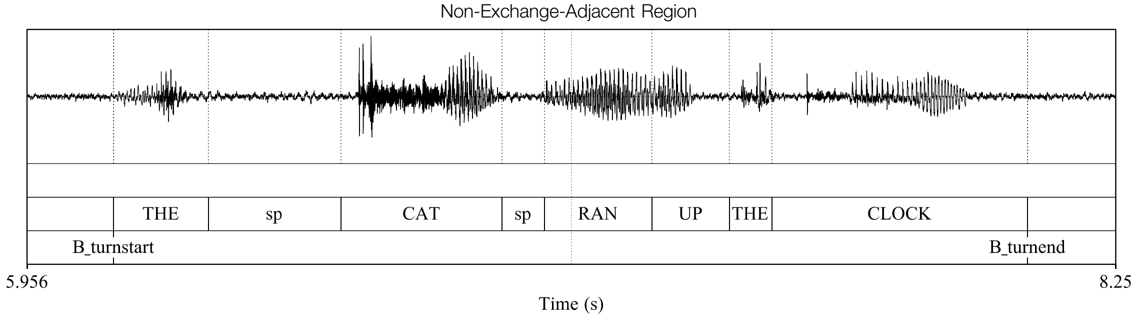
Non-Overlapping Exchange



Overlapping Exchange



Non-Exchange-Adjacent Region



experiment management called Marta (custom software written by Mark Tiede at Haskins Laboratories, New Haven, CT) was used to present stimuli to participants via separate monitors on their desks, and to save and organize recorded data by participant and trial for later analysis. The EMA sensor movement was sampled at 400 Hz, and speech audio was sampled at 44.1 kHz. GoPro cameras were used for video capture, using 1080 PPI resolution and a medium field of view in H.264 encoding; secondary GoPro audio was recorded in stereo at 48 kHz. (GoPro video and audio are not analyzed for the present study).

After sensor placement was complete, participants were jointly briefed on their tasks and the experiment began. The first task in the experiment was a brief mutual introduction between the two participants lasting 2 min, which served to familiarize the dyad members with one another and help them adapt to speaking with the EMA sensors. The next task was the primary experimental task of collaboratively completing nursery rhymes. The specific instructions that subjects received are described in **Supplementary Table S1**; the experimenters also verbally instructed the subjects that they will work as a team taking turns to say a nursery rhyme, going back and forth and helping one another finish the rhyme if someone gets stuck. On their screen, participants saw the first phrase of each rhyme, and for each rhyme, a graphic "star" was displayed on one participant's screen serving to denote who would start that particular rhyme; this alternated between speakers. The rhyme presentation order was randomized for each dyad, and before each trial, a beep sound was played as a go-signal (and to facilitate future alignment of video with EMA/audio). The dyads completed practice trials of two nursery rhymes that were not repeated in the main experiment, after which, experiment personnel answered any participant questions and provided feedback on whether the practice trial was performed in accordance with the instructions. The participants then proceeded to complete one block of the nursery rhyme task (24 nursery rhymes). After the conclusion of the entire first block, the speakers completed another 2 min conversation period in which they were asked to find out what they had in common with one another; this provided a rest from the semi-structured task (and possibly helped sustain a friendly affiliative atmosphere between the participants). Following this, the

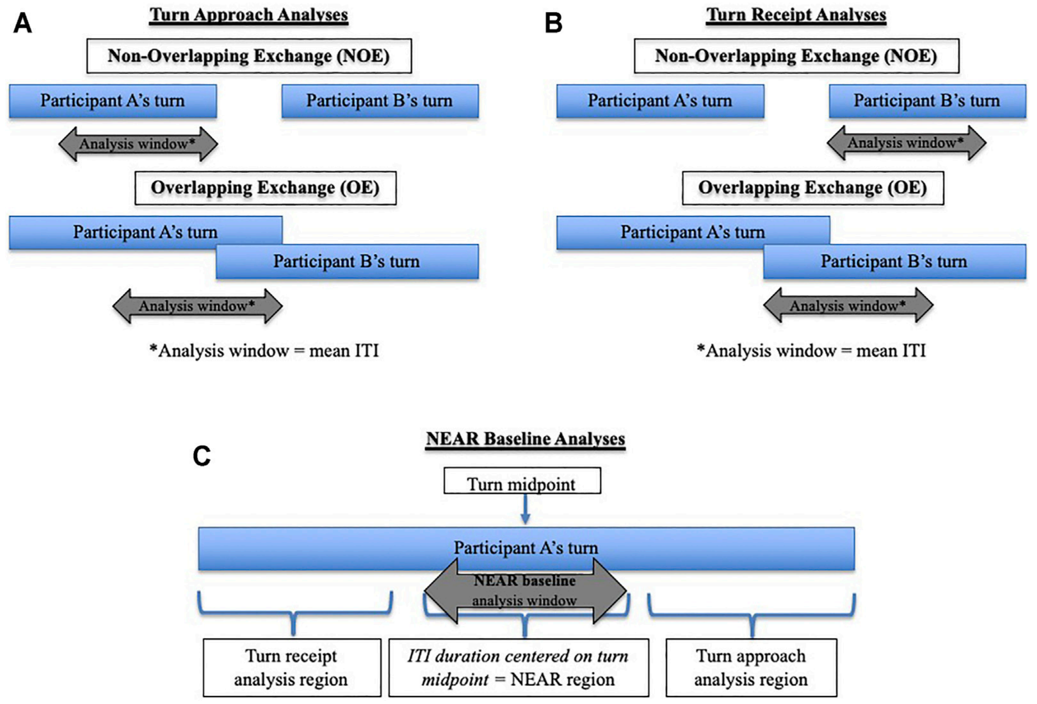


FIGURE 3 | Panel **(A)** shows the Turn Approach analysis schematic; Panel **(B)** shows the Turn Receipt analysis schematic; Panel **(C)** shows the NEAR region schematic.

participants were again given the chance to briefly read through the printed pages of nursery rhymes. The second repetition of the nursery rhyme task subsequently commenced with stimuli presented in the same randomized presentation order as in the first block for that dyad. The completion of the second round of the nursery rhyme task concluded the experiment, after which sensors were removed. The conversational and commonalities tasks were not included in the data analysis portion of the study; only blocks 1 and 2 of the nursery rhyme task were analyzed. That said, in addition to the nursery rhyme material itself, the design elicited a substantial amount of conversational material not related to the rhymes within those blocks as the dyads conversed and collaborated on the task.

This protocol yielded a large database of over 85 min of actual speech audio (Dyad S5S6: 33.67 min; Dyad S7S8: 19.59 min; Dyad S9S10: 12.61 min; Dyad S11S12: 19.43 min).

Data Processing

The kinematic trajectories of the EMA sensors were used to calculate gestural density of head and brow gestures as follows. EMA sensor trajectory data was prepared for use with the MATLAB-based analysis program Mview (custom software written by Mark Tiede at Haskins Laboratories, New Haven, CT) by interpolating missing data and extracting three-dimensional sensor trajectories from raw data. As is standard, EMA sensor data was rotated to a coordinate system aligned with the speaker's occlusal plane, and brow and lip-corner sensors were corrected for headmovement. Custom Python scripts were created to extract head and eyebrow movement data from the EMA sensor trajectory data.

Head movement data was derived as follows: the three-dimensional movement of the plane formed by the three head reference sensors (left and right external mastoid processes and just above the upper incisor [UI]) rotating around the projected EMA system origin was calculated at each sample. This head movement data was subsequently detrended and low-pass filtered at 5 Hz (Tiede et al., 2010). (Instantaneous) angular velocity derived from all three available dimensions of movement was then computed. Angular velocity peaks (in three dimensions) were extracted from all data for a given participant (as is common in EMA-derived signal analysis, the minimum velocity threshold for a given speaker was computed using 5% of the maximum observed value across all of that speaker's trials; below-threshold head velocity peaks were not considered). **Figure 2** shows an illustration of the head movement measure.

Brow movement was derived as follows: the y-dimensional Euclidean distance from the right brow sensor to the (fixed) upper incisor [UI] sensor was calculated. Brow movements were detrended and low-pass filtered at 12 Hz, and their instantaneous velocity was computed from the change in y-dimensional distance from the brow sensor to the fixed mandibular UI sensor. Positive instantaneous velocities, associated with upward-going brow movements, were used for all subsequent data analysis. Negative instantaneous velocities, associated with downward-going brow movements were not analyzed, as brow raising but not brow lowering has been observed to co-occur with discourse-relevant and with prosodically relevant acoustic events in speech (Flecha-García, 2010; Prieto et al., 2015). In the same manner as the head movement data, a minimum velocity peak threshold was computed

TABLE 2 | Summary of floor exchange type counts by Dyad.

Floor exchange type counts by Dyad	Near	NOE	OE	Totals
Dyad 1 (S5/S6)	130	570	394	1,094
Dyad 2 (S7/S8)	78	450	176	704
Dyad 3 (S9/S10)	106	282	162	550
Dyad 4 (S11/S12)	76	484	202	762
Totals	390	1786	934	3,110

for each participant's brow data using 5% of the maximum value across all of a given participant's trials, and only velocity peaks above this threshold were used.

For both head and brow movement data, the primary measure of interest was co-speech gesture density, measured as velocity *peaks per second* (PPS). Prior research has suggested that co-speech gestural density depends on speech and interlocutor context (Ishi et al., 2014; Danner et al., 2018). Gestural PPS is a time-normalized rate measure calculated for a variety of conversationally relevant regions, as described in detail below in *Exchange Types, Conversational Roles and Speech Content Types*.

The first author along with two trained research assistants produced a word-level transcription of the recorded speech. These transcriptions and the associated audio files were then submitted to the Penn Forced Aligner (Yuan and Liberman, 2008) for automatic text alignment, resulting in the production of Praat TextGrids (Boersma and Weenink, 2016) for each file. This implementation of forced alignment cannot attribute parts of a transcription to multiple speakers, so a subsequent annotation step was performed to check/correct the automatic alignment and to attribute speech to each of the two recorded speakers in a trial. After a transcription was produced using a two-channel audio file (one channel per speaker), speaker attribution was performed by separating the audio files into two mono channels, each of which was associated with only one speaker's microphone. The final TextGrids contain the automatic force-aligned transcriptions at phone and word levels, a tier for each of the two speakers in a given file containing only the speech attributed to a given speaker, and a point tier where the acoustic onset and offset of each participant's speech was annotated; the last three of these tiers are shown in the examples in **Table 1**. The annotations were used to demarcate participants' speech *turns*, with the acoustic onset and offset of each participant's speech corresponding to turn start and end points, respectively. Speech turn exchange events are described further below. All transcriptions, annotations, and turn start/end points were cross-checked by the first author and assistants for accuracy.

Exchange Types, Conversational Roles and Speech Content Types

The TextGrids described above were coded for different types of floor exchanges, the conversational role held by each speaker at each exchange, and whether the content of speech at the end of turns was rhyme-related or not. These coding decisions were made in the context of *analysis windows*. To determine the duration of the analysis window (which was dyad-dependent), the average inter-turn interval (ITI) for each dyad was computed as the average interval duration from the acoustic offset of a speech turn to the acoustic onset of the next speech turn across every trial of that dyad. The ITI was determined separately for each dyad to account for dyad-specific factors such as differences in conversational rate. The ITI duration was used only to determine the duration of the analysis window local to a floor exchange over which co-speech movement density was calculated (see *Turn Approach and Turn Receipt Analyses* below and **Figure 3**); ITI was not itself analyzed.

Three floor exchange types[3] (factor: EXCHANGE TYPE) were designated for data analysis, as follows (see **Figure 3**):

- Non-Overlapping Exchange (NOE): Exchanges in which one member of a dyad stops speaking, and after a pause, the other dyad member begins speaking.
- Overlapping Exchange (OE): Exchanges in which one dyad member begins speaking prior to the time when the other dyad member has stopped speaking.
- Non-Exchange-Adjacent Region (NEAR): A region of participant speech that does not fall within any other analysis window and which is not interrupted by the speech of the other dyad member. NEAR regions are considered a *baseline* region to which the other exchange-proximate regions of interest are compared; this level therefore serves as the reference level for the EXCHANGE TYPE factor.

Example TextGrids for each of the analyzed exchange types (NOE, OE and NEAR) are shown in **Table 1**, and a summary of exchange types for each dyad can be found in **Table 2**.

In addition to floor exchange types, the data were coded for two conversational roles (factor: ROLE):

- Speaker: leading up to a NOE (non-overlapping exchange), "speaker" is the dyad member speaking prior to the pause; after an NOE has just occurred, "speaker" is the person who takes the floor and begins speaking. At an OE (overlapping exchange), the "speaker" is the dyad member who is initially speaking before the other dyad member begins speaking. Speaker is used as the reference level of the ROLES factor.
- Listener: This is the dyad member who is *not* speaking during the analysis window, with the exception of the analysis region following an OE, in which case "listener" is the dyad member who is initially not speaking but who

[3]Two additional exchange types were identified in the dataset but were not included in further analysis. The first of these is a *turn-within-turn*, in which one dyad member's speech turn occurs entirely within the other dyad member's speech turn. The second exchange type excluded from analysis is a *non-consummated exchange*, in which one dyad member stops speaking and after a long pause (>500 ms) during which the other dyad member remains silent, the same dyad member begins speaking once again. These excluded exchange types are challenging to interpret as *turns-within-turns* could represent either backchanneling or a failed floor exchange, and *non-consummated exchanges* could represent a failed floor exchange or an exceptionally long pause.

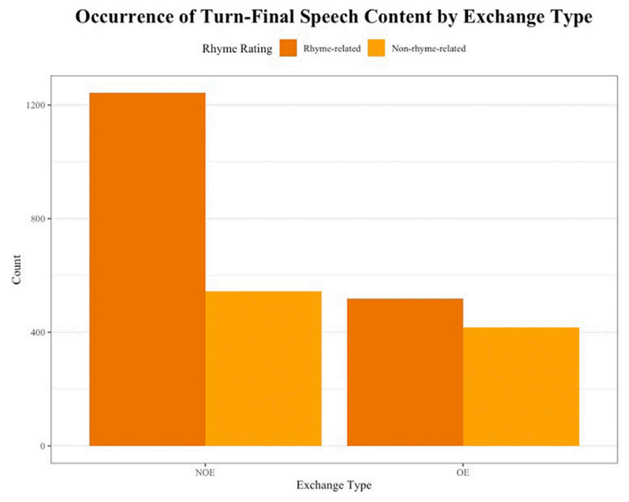


FIGURE 4 | Count of speech content occurrence by exchange type, pooled across speakers.

then begins speaking during the ongoing speech turn of the other dyad member.

Finally, both overlapping and non-overlapping exchanges were coded as being rhyme-related or non-rhyme-related (*speech content*). To perform this analysis, ITI durations described above were used to create an analysis window whose right edge aligned with the right edge of exchanges for each dyad. The text transcriptions of recorded speech occurring within this *turn-approach analysis window* were extracted and, to define the factor SPEECH CONTENT obtaining at the floor exchange, the three coauthors coded the extracted transcriptions as being either:

- Rhyme-related: Primarily lexical material associated with the nursery rhyme that is underway (whether correct words or not).[4] Rhyme-related speech was used as the reference level for the SPEECH CONTENT factor
- Non-Rhyme-related: Primarily lexical material that is not associated with the nursery rhyme that is underway

Among overlapping (OE) and non-overlapping (NOE) exchanges (n = 1706), 38% of SPEECH CONTENT was coded as non-rhyme-related and 62% was coded as rhyme-related. Average pairwise rater agreement was very strong at 93.24%; Fleiss' $\kappa = 0.852$ (interrater reliability was assessed using ReCal3 (Freelon, 2013) and R package irr (Gamer et al., 2019)).

Turn Approach and Turn Receipt Analyses

Two regions of analysis of co-speech movement are considered—the region immediately leading up to a floor exchange, denoted the Turn Approach, and the region immediately following a floor exchange, denoted the Turn Receipt.

In the Turn Approach Analysis for non-overlapping exchanges (NOE), the right edge of the analysis window is aligned with the right edge—or end—of a participant's speech turn, such that the analysis window covers the speech interval *leading up to the floor exchange*. For overlapping exchanges (OE), the right edge of the analysis window is aligned with the right edge (end) of the initial speaker's turn. Schematic representations of analysis window placement and length for each Turn Approach exchange type are given in **Figure 3A**.

Specifically, to compute the average inter-turn interval (ITI) for each dyad, we took the following steps. For non-overlapping exchanges (NOE), the ITI value is a positive number. For overlapping exchanges (OE), we considered the ITI value to simply be zero since there is no inter-turn interval or delay between when one speaker stops speaking and the other begins. We summed all the ITI values for a dyad and divided by the total number of exchanges for that dyad[5]. This procedure for calculating average ITI yielded analysis windows of: 858 ms for Dyad 1, 763 ms for Dyad 2, 485 ms for Dyad 3, and 730 ms for Dyad 4. Note that ITI was used *only* to define the duration of the analysis windows and was not itself the object of any analysis.

In the Turn Receipt analysis, the analysis window duration for each dyad is computed in the same manner as in the Turn Approach analysis. This analysis is complementary to the Turn Approach analysis, in that the Turn Receipt analysis focuses on the opposite "side" of speech turns from the Turn Approach analysis. Therefore, the placement of the analysis window is now aligned to the *left edge* (or onset) of a speech turn, such that the analysis window covers the portion of a turn immediately *following* a speaker exchange. In the case of an overlapping exchange (OE) this corresponds with the onset of the second speaker's turn. A schematic representation of the placement of analysis windows used in the Turn Receipt analysis is shown in **Figure 3B**.

Finally, the reference level for comparing movement density at floor exchanges was specified to be the NEAR (non-exchange adjacent region); see **Figure 3C**. The NEAR region is equivalent to the ITI duration centered on the midpoint of a turn, when turns were sufficiently long such that the NEAR region did not interfere with any other speaker's speech or any other possible analysis region (either Turn Approach or Turn Receipt). If there was not enough duration in a given speech turn to guarantee that the NEAR region did not overlap any other analysis region, the NEAR was not calculated for that turn.

[4]Coders also had available for reference the canonical text of each of the 24 nursery rhymes. In cases where both speakers were speaking during the analysis window, the instructions to the coders stated that speech should be coded as 'rhyme-related' if either one of the two speakers' transcriptions were primarily lexical material associated with the ongoing rhyme

[5]In addition to NOE and OE exchange types, *non-consummated exchanges* (as described in *Exchange Types, Conversational Roles and Speech Content Types*) are included in the ITI calculations so as to include all potential floor exchanges

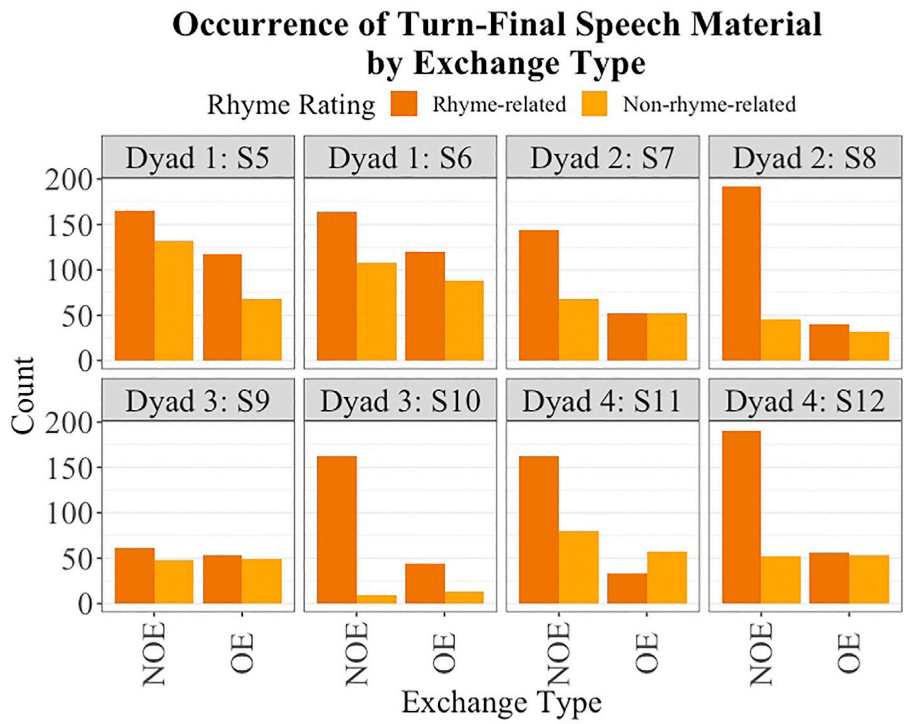


FIGURE 5 | Count of speech content occurrence by exchange type by speaker.

RESULTS

The co-speech movement density results presented here comprise visualization, descriptive analysis, and linear mixed effects modeling. Data processing was performed in MATLAB (MATLAB, 2018), and statistical analyses were performed in R version 4.1.1 (R Core Team, 2021). Data manipulation and organization was performed in R using package dplyr version 1.0.7 (Wickham et al., 2019). Visualizations were produced using R package ggplot2 version 3.3.5 (Wickham et al., 2019). Linear mixed effects models and associated statistics were produced using R packages lme4 version 1.1–27.1 (Bates et al., 2015), lmerTest version 3.1–3 (Kuznetsova et al., 2014), and afex version 1.0–1 (Singmann et al., 2021). Each Turn Approach and Turn Receipt analysis includes violin graph visualizations of mean velocity peaks per second (PPS) for brow and head movements by individual participant, descriptive statistics for the PPS measure (summarized over all participants), and linear mixed effects models detailed in the **Supplementary Materials**.

The first analysis probes the effect of SPEECH CONTENT during Turn Approach (the spoken material immediately preceding speech offset) on co-speech movement density (in peaks per second)[6]. We used the lmer() function in R package lme4 (Bates et al., 2015) to create a model containing the fixed effect of SPEECH CONTENT and random effects of ITEM and PARTICIPANT nested within DYAD. We estimated significance using the χ^2 tests and F-tests in the mixed() function of afex (Singmann et al., 2021).

Next, the analysis shifts to the primary goal of illuminating how CONVERSATIONAL ROLE and speech EXCHANGE TYPE affect co-speech gesturing rate for movements of the brow and head in both the Turn Approach and Turn Receipt analyses. The Turn Approach region considers co-speech movement behavior *leading up* to a floor exchange. The Turn Receipt analysis concerns co-speech behavior *immediately after* a speaker exchange has occurred. The fixed effects are conversational ROLE and speech EXCHANGE TYPE and their interaction, with random effects of ITEM (where each item is a particular nursery rhyme) and of PARTICIPANTS nested within DYADS. These models do not include random slopes because introduction of random slopes created convergence issues. The linear mixed effects models in these analyses all used the same fixed and random effects structure (PPS ~ ROLE * EXCHANGE TYPE + (1|DYAD/PARTICIPANT) + (1|ITEM). We used the lmer() function in R package lme4 (Bates et al., 2015) to create an initial (treatment-coded) model containing all effects of interest. Then, using the function mixed() in the R package afex (Singmann et al., 2021), we estimated the significance of all fixed effects entered in the interaction analyses using *F*-tests with the Kenward-Roger method[7] for approximating degrees of freedom. Finally, for

[6]Note that because speech content type was evaluated based only on the words at the offset of speech turns, this was analyzed only for Turn Approach and not for Turn Receipt analyses (because Turn Receipt analyses consider speech turn onset).

[7]This method provides good control against Type I errors in smaller datasets like the one presented here.

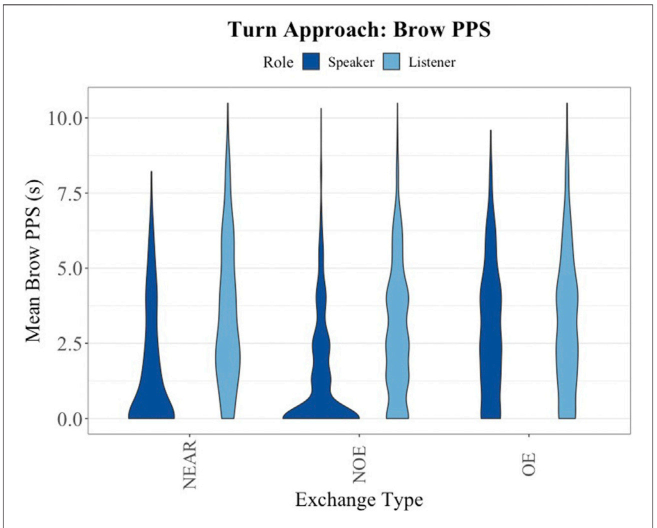


FIGURE 6 | Turn Approach brow PPS summary by exchange type and role.

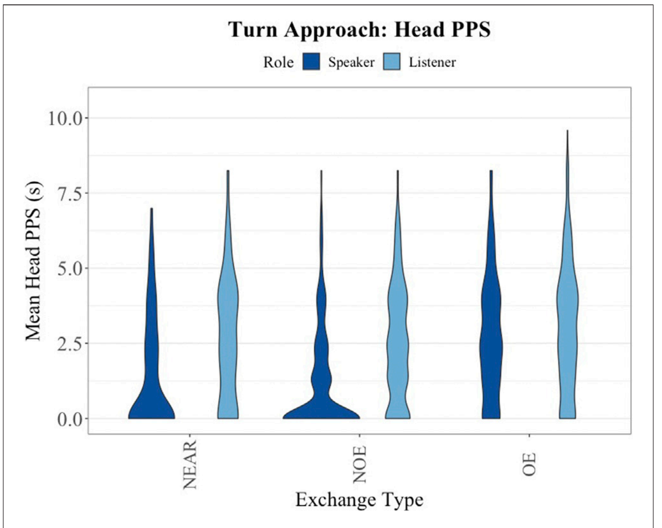


FIGURE 7 | Turn Approach Head PPS summary by exchange type and role.

TABLE 3 | Turn Approach Brow PPS summary statistics (pooled over participants).

Exchange type	Role	*N*	Mean PPS	Median PPS	Max PPS
NEAR	Speaker	195	1.58	1.17	8.22
NEAR	Listener	195	3.35	2.74	10.5
NOE	Speaker	893	1.32	0	10.3
NOE	Listener	893	2.86	2.62	10.5
OE	Speaker	467	2.94	2.74	9.59
OE	Listener	467	3.13	2.74	10.5

each ROLE * EXCHANGE TYPE analysis, we created two parallel models with contrast coding to examine the main effects of EXCHANGE TYPE and ROLE. All models are reported in the **Supplementary Materials** section.

Speech Content Analysis

Previous research suggests that lexical/semantic content is useful to speakers in predicting the end of a speech turn (De Ruiter et al., 2006; Garrod and Pickering, 2015), and while our design reflects no prediction regarding whether the speech content at floor exchanges is associated with unique movement behavior responses, it was prudent to determine whether the nature of the lexical content at the floor exchange (coded as *rhyme-related* or *non-rhyme-related* based on the immediately preceding lexical material) had an association with co-speech movement density. Recall that a large portion, more than a third, of the speech content defined in this way was conversational and not specific to nursery rhyme production.

First, a 2 × 2 contingency table was created comparing type of SPEECH CONTENT (rhyme-related or non-rhyme-related) and EXCHANGE TYPE (overlapping exchange or non-overlapping exchange), and a χ^2 analysis was performed to statistically assess the distributions. We found that SPEECH CONTENT was indeed non-randomly associated with EXCHANGE TYPE; χ^2 (df = 1, N = 2,720) = 52.627 ($p < 0.001$). Specifically, rhyme-related speech content was more likely to be found at non-overlapping exchanges than at overlapping exchanges (see **Figure 4**), though qualitatively, the strength of this association varied by speaker (see **Figure 5**).

To statistically test the speech content analysis, we specified linear mixed effects models with the same structure for both brow and head movement signals[8]; we only consider the Turn Approach region in this speech content analysis, as lexical material in only this region was the basis for the coding for SPEECH CONTENT. We included SPEECH CONTENT as a fixed effect; random intercepts were fitted for each of two random effects, ITEM and PARTICIPANT (participants are nested within dyad)[9]. The SPEECH CONTENT model for brow movements was found to differ significantly from a model without the SPEECH CONTENT effect ($\chi^2(1)$ = 4.835, p = 0.028). Non-rhyme-related speech was associated with significantly denser Brow movements than baseline rhyme-related speech (β = 0.203, SE = 0.091, t = 2.229, p = 0.026). The SPEECH CONTENT model for the head also differed from a model without the SPEECH CONTENT effect ($\chi^2(1)$ = 12.565, $p < 0.001$). Non-rhyme-related speech was also associated with significantly denser Head movements than baseline rhyme-related speech (β = 0.299, SE = 0.084, t = 3.585, $p < 0.001$). See **Supplementary Tables S2, S3** for model formula and complete model summaries.

Turn Approach Analysis

As described above, the Turn Approach analysis was designed to consider speaker and listener behavior at and just before the *offset* of a speech turn (**Figure 3A**). This provides insight into the ways

[8]We use treatment coding to report results for the single fixed effect in the model.
[9]The specification of random slopes caused convergence issues and were therefore not included; though it would be ideal to have enough data to estimate random slopes, the random effects structure described here is a very good representation of the experiment as performed.

TABLE 4 | Turn Approach Head PPS summary statistics (pooled over participants).

Exchange type	Role	*N*	Mean PPS	Median PPS	Max PPS
NEAR	Speaker	195	1.40	0	6.99
NEAR	Listener	195	2.60	2.33	8.25
NOE	Speaker	893	1.18	0	8.25
NOE	Listener	893	2.48	2.33	8.25
OE	Speaker	467	2.72	2.33	8.25
OE	Listener	467	2.99	2.74	9.59

TABLE 5 | Turn Receipt Brow PPS summary statistics (pooled over participants).

Exchange type	Role	*N*	Mean PPS	Median PPS	Max PPS
NEAR	Speaker	195	1.58	1.17	8.22
NEAR	Listener	195	3.35	2.74	10.5
NOE	Speaker	893	1.32	0	9.59
NOE	Listener	893	3.36	3.50	10.5
OE	Speaker	467	2.02	1.37	8.22
OE	Listener	467	3.43	3.50	10.5

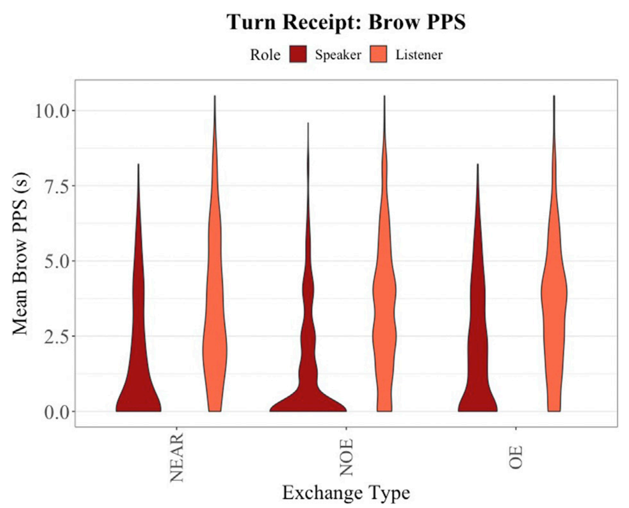


FIGURE 8 | Turn Receipt Brow PPS summary by exchange type and role.

that speakers and listeners may pattern their movement behavior *in anticipation* of an upcoming floor exchange. The brow movement data is considered first and then the head movement data.

Turn Approach Brow PPS

Figure 6 shows a violin plot of Turn Approach Brow PPS data in the two factors of interest, conversational ROLE and floor EXCHANGE TYPE. Descriptive statistics are summarized in **Table 3**. Recall that NEAR regions (non-exchange-adjacent regions) are utilized as a reference level for EXCHANGE TYPE and Speaker is the reference level for ROLE. Including the predictors ROLE and EXCHANGE TYPE and their interaction improved model fit (F (2) = 37.01, $p < 0.001$). A model summary for the final model is available in **Supplementary Table S4**. In the model testing the main effect of ROLE, Listeners were found to show significantly denser brow movements than Speakers ($\beta = 1.165$, SE = 0.089, t = 13.039, $p < 0.001$). In the model testing the main effect of EXCHANGE TYPE, significantly denser brow movement was attested in the OE region than in the NEAR region ($\beta = 0.614$, SE = 0.127, t = 4.850, $p < 0.001$). Conversely, significantly *less dense* brow movement was attested in the NOE region than in the NEAR region ($\beta = -0.275$, SE = 0.117, t = -2.342, $p = 0.019$). See **Supplementary Table S5** for summaries of the contrast-coded models used to report main effects. Finally, a significant interaction of Listener and OE region was observed, indicating that PPS values are affected by both the EXCHANGE TYPE and a dyad member's ROLE as speaker or listener ($\beta = -1.570$, SE = 0.250, t = −6.287, $p < 0.001$). The PPS parameter estimate for Listeners at OE (3.127 PPS) is qualitatively higher than that of Speakers at OE (2.934 PPS) and the PPS value for Speakers' brow movements at overlapping exchanges is more dense than their movements at NEAR (1.534), while this difference did not exist for Listeners (see regression table in **Supplementary Table S6**). The distinction in listener and speaker brow movement behavior in OE and NEAR regions drives the observed significant interaction value.

In sum, these results suggest that brow movements in Turn Approach regions are substantially more frequent for listeners than speakers, that brow movements are more dense at overlapping exchanges than in non-exchange adjacent regions (NEAR) of speech and less dense at non-overlapping exchanges than during NEAR speech. Additionally, EXCHANGE TYPE and ROLE jointly affect brow movement density driven by the fact that Speakers' co-speech brow movements are denser at overlapping exchanges in Turn Approach than they are in non-exchange adjacent regions.

Turn Approach Head PPS

Turning to head movement density at Turn Approach, key patterns are shown in **Figure 7**. Descriptive statistics are summarized in **Table 4**. Including the predictors EXCHANGE TYPE and ROLE and their interaction improved model fit (F (2) = 23.99, $p < 0.0001$). A summary of the full model is presented in **Supplementary Table S7**. In a contrast-coded model testing the main effect of ROLE, Listeners were found to show significantly denser head movements than Speakers ($\beta = 0.927$, SE = 0.081, t = 11.387, $p < 0.001$). In a contrast-coded model testing the main effect of EXCHANGE TYPE, OE regions are associated with significantly greater head movement density than NEAR regions ($\beta = 0.838$, SE = 0.115, t = 7.256, $p < 0.001$). See **Supplementary Table S8** for summaries of the contrast-coded models used to report main effects. A significant interaction of Listener and OE is obtained in the treatment-coded Turn Approach head movement model ($\beta = -0.937$, SE = 0.227, t = −4.125, $p < 0.001$). The PPS parameter estimate for Listeners at OE (2.951 PPS) is qualitatively greater than that of Speakers at OE (2.683 PPS), and the PPS value for Speakers at OEs is qualitatively denser than in NEAR intervals (1.376 PPS) but such a difference is not apparent for Listeners (see regression table in **Supplementary Table S9**), which drives the observed significant interaction value.

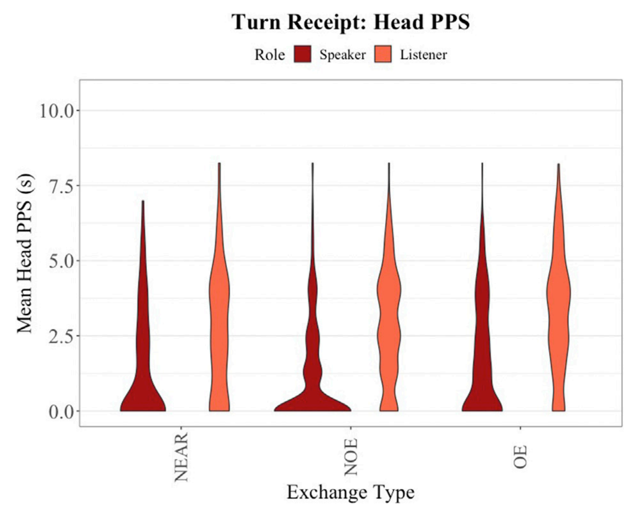


FIGURE 9 | Turn Receipt Head PPS summary by exchange type and role.

TABLE 6 | Turn Receipt Head PPS summary statistics (pooled over participants).

Exchange type	Role	N	Mean PPS	Median PPS	Max PPS
NEAR	Speaker	195	1.40	0	6.99
NEAR	Listener	195	2.60	2.33	8.25
NOE	Speaker	893	1.15	0	8.25
NOE	Listener	893	2.82	2.74	8.25
OE	Speaker	467	1.74	1.31	8.25
OE	Listener	467	3.24	3.50	8.22

In sum, these results suggest that head movements in Turn Approach are, like brow movements, substantially more frequent for listeners than speakers, and that head movements in OE regions are more dense than those in NEAR regions. Also as with brow, EXCHANGE TYPE and ROLE jointly affect brow movement density with Speakers' head movements being more dense at overlapping exchanges than at non-exchange adjacent regions.

Turn Receipt Analysis

The Turn Receipt analysis is complementary to the Turn Approach analysis, considering speaker and listener behavior at the *onset* and in the early moments of an initiated speech turn, when a new speaker has just begun speaking (see **Figure 3B**). The brow movement data is considered first, followed by the head movement data.

Turn Receipt Brow PPS

A graphical representation of Turn Receipt Brow PPS results is shown in **Figure 8**. Descriptive statistics are summarized in **Table 5**. Including the predictors EXCHANGE TYPE and ROLE and their interaction improved model fit (F (2) = 7.18, $p < 0.001$). In a contrast-coded model testing the main effect of ROLE, Listeners' brow movements at Turn Receipt were significantly denser than speakers' movements ($\beta = 1.736$, SE = 0.090, $t = 19.260$, $p < 0.001$). In a contrast-coded model testing the main effect of EXCHANGE TYPE, OE regions are associated with significantly greater head movement density than NEAR regions ($\beta = 0.372$, SE = 0.128, $t = 2.916$, $p = 0.004$). No significant interactions between levels of EXCHANGE TYPE and ROLE were observed for the Turn Receipt Brow model. See **Supplementary Table S10** for a summary of the full model, **Supplementary Table S11** for summaries of the contrast-coded models used to report main effects, and **Supplementary Table S12** for the regression table for this model.

In sum, for the Turn Receipt Brow model, we again observe significantly denser listener brow movement compared with speakers. As was seen in the Turn Approach models, brow movements in OE regions at Turn Receipt are significantly denser than brow movements in NEAR regions. No significant interactions between the two fixed effects in this model were observed.

Turn Receipt Head PPS

Turning to the head movement data at Turn Receipt, a graphical representation of PPS for individual participants is shown in **Figure 9**. Descriptive statistics are summarized in **Table 6**. Including the predictors EXCHANGE TYPE and their interaction improved model fit marginally (F (2) = 2.84, $p = 0.058$). In a contrast-coded model testing the main effect of ROLE, Listeners were found to show denser head movements at turn approach than Speakers ($\beta = 1.458$, SE = 0.079, $t = 18.472$, $p < 0.001$). Similar to all the models discussed so far, significantly more head movement was attested in the OE Turn Receipt region than in the NEAR region ($\beta = 0.461$, SE = 0.112, $t = 4.122$, $p > 0.001$). A significant interaction of Listener and *NOE* is also observed in this model ($\beta = 0.468$, SE = 0.204, $t = 2.289$, $p = 0.022$). Note that the model fit only marginally improves when including the interaction term, so these model results should not be over-interpreted. See **Supplementary Table S13** for a summary of the full Turn Receipt Head model, **Supplementary Table S14** for summaries of the contrast-coded models used to report main effects, and **Supplementary Table S15** for the full model regression table.

These results suggest that head movements in Turn Receipt regions are more frequent for listeners than speakers, and more frequent in OE regions than NEAR regions, as also observed for brow and head in Turn Approach and brow in Turn Receipt. A significant crossover interaction in NOE*Listener obtained, a result that is unique to the Turn Receipt head model.

Results Summary

The experiment protocol successfully provided a rich database of speech for four interacting dyads, with a variety of floor exchange types, speech both related and unrelated to the nursery rhyme prompts, and participants acting both as speakers and as listeners.

The analyses of the brow and head movement density signals revealed several similarities. Non-rhyme related SPEECH CONTENT was associated with greater movement density than rhyme-related speech content, for Turn Approach head and brow movements. Listeners consistently produced higher movement density than speakers for both brow and head movement across all turn types, both approaching and following a floor exchange. Overlapping exchange regions were consistently associated with denser movements of both brow and head in Turn Approach and

Turn Receipt than non-overlapping exchanges. Finally, *approaching an overlapping floor exchange*, speakers but not listeners displayed more dense movements of both brow and head relative to movement during speech remote from the floor exchange.

DISCUSSION

Consistency Across Movement Signals

One of the current findings that warrants highlighting is the similarities in behavior across brow and head movements measured in this research.[10] While a few researchers have considered both brow and head movements in the same study (e.g., Bolinger, 1983; Hadar et al., 1983; McClave, 2000; Clark and Krych, 2004; Munhall et al., 2004; Krahmer and Swerts, 2007; Kita, 2009; Kim et al., 2014; Prieto et al., 2015), previous research has not illuminated whether different effectors of co-speech movements pattern similarly or differently at floor exchanges. In the present study, there is a remarkable similarity in how brow and head behave in the vicinity before, after, and remote from a floor exchange. Given the inherent differences in range of motion, degrees of freedom and velocity of the signal types (and the known role of head movement in signaling semantic content such as agreement), this finding of systematic and similar patterning across the brow and head modalities stands to inform future investigations.

Speech Content and Movement Behavior

The central role of semantic and lexical content in successful conversational interaction is clear (De Ruiter et al., 2006). In this study we did not embark on a rigorous analysis of lexico-semantic characteristics of speech; we simply noted whether the spoken material immediately at the floor exchange was related or unrelated to the nursery rhyme verse and we tested whether that coded content had an association with movement behavior. A substantial number of studies have found that co-speech movement facilitates speech production—whether by facilitating thinking, reducing cognitive load, or facilitating lexical access (Alibali et al., 2000; Gillespie et al., 2014; Goldin-Meadow et al., 2001; Krauss, 1998; Melinger & Kita, 2007, though see Hoetjes et al., 2014 for possible evidence against this view). We would therefore have expected interlocutors to have higher co-speech movement density when executing the challenge of the rhyme task material, and furthermore, the rhythmic nature of the task (producing nursery rhymes) could have contributed to an increase of movement as well (for example an increase in movement associated with beat gestures). Instead, the reverse transpired for the Turn Approach region. It may be that the topical content of the non-rhyme related material was sufficiently concerned with the challenges of the collaborative task that it exhibited an uptick in co-speech movement density associated with heightened affect or load.

Exchange Types and Movement Behavior

This study sought to determine whether different types of conversational floor exchange events are associated with empirically distinct head or brow movement density. One clear result emerged across both analyses and signal types: overlapping exchanges were associated with speakers having substantially more dense head and brow movements than they did in non-exchange-adjacent regions of speech (**Figures 5**, **6**). **Figures 6–9** This finding can be considered in line with Duncan's suggestion that termination of manual co-speech gestures on the part of the speaker is a turn-yielding signal and the continuation of a manual co-speech gesture an "attempt suppressing" signal (Duncan Jr, 1972). While our approach differs from Duncan's, in that we do not analyze the timing of the end or the continuation of co-speech gestures but rather the density of movement, we think this increase in movement at overlapping exchanges by the speaker can be seen as further supporting this finding through a different measure, and now for brow and head movement. Alternatively, this increase in movement could be also due to the speaker and listener interacting concurrently and the speaker signaling cooperation in yielding the turn.

Movements at non-overlapping exchanges (NOEs) showed no consistent difference from non-exchange-adjacent (NEAR) speech for brow or head movements in the Turn Approach or Receipt regions. It is not entirely clear why movement behavior around non-overlapping exchange (NOE) speech is similar to baseline because co-speech movements could conceivably help with smooth turn-exchanges (e.g., Stivers et al., 2009; Trujillo et al., 2021). Nevertheless, our results indicate that interlocutors generally negotiate a NOE without an increase in their co-speech movement. (Cf. Duncan Jr, 1972 who finds the end of a manual gesture to be a signal for the end of the turn.) A question for future research is whether listeners actually use these movement signals to help predict the end of a current speaker's turn. An additional topic for future research concerns the functional role(s) of movement during the Turn Receipt, and whether qualitative rather than quantitative changes in movement behavior are meaningful.

Conversational Roles and Movement Behavior

Previous research has focused predominantly on co-speech movement behavior of a speaker. While there are a few previous works that have focused on the co-speech behavior of listeners versus speakers (Hilton, 2018), relatively little is known about listener movement behavior or simply the behavior associated with silent listening. Our study offers a novel consideration of empirical kinematic data collected simultaneously from both a speaker and a listener during interaction. One of the most consistent observations in the co-speech movement in our study was the more frequent head and brow gestures of listeners as compared to speakers. When participants were in the role of listeners, they moved their head and brow more frequently than they did when in the role of speakers, an observation that held true for both Turn Approach and Turn Receipt analysis regions. There are a number of interpretations of

[10]We can be sure in our study that the movement of the brow is not merely a consequence of the movement of the head, because head movement correction was performed on the brow movement trajectories (but not on the head movement trajectories). It is plausible that other future measures of the brow and head movements, such as displacement or duration, could yield differences.

why listener co-speech movement is more frequent than speaker movement in the present study. The uptick in co-speech movement density may indicate attentiveness or affiliation with the speaker (Clark and Krych, 2004; Latif et al., 2014), or it may indicate a listener's intent to start speaking (Duncan Jr, 1972; Hadar et al., 1985; Lee and Narayanan, 2010), or simply help the listener initiate their turn in some way (Hadar et al., 1983). Certainly, the across-the-board higher density of listener co-speech movement could be due to backchanneling that helps regulate turn-taking (McClave, 2000), but it may also be one way for the listener to engage in the interaction that does not intrude on the spoken contribution currently underway.

SUMMARY AND CONCLUSION

Using a novel interactive protocol designed to elicit many conversational floor exchanges within a structured, non-read dyadic speech interaction, this study examines hypotheses that the density of co-speech movements differs depending on exchange type and the participant's role as speaker or listener in the interaction. The results support the specific hypotheses. In brief, we find that co-speech movements of the brow and head are more dense for listeners as opposed to speakers, and that this is the case in both Turn Approach and Turn Receipt regions. Additionally, listeners display a higher rate of co-speech movement than speakers both at floor exchanges and remote from them. This patterning may be related to a listener's desire to signal interest, engagement or attention to the speaker without actually intruding on their interlocutor's speech signal, as well as possibly facilitating conversational turn-taking (Hadar et al., 1985; Holler et al., 2017).

Movement behavior is increased for speakers approaching overlapping exchanges (interruptions). Conversational role interacts with the type of floor exchange in its association with co-speech movement. Speakers who are approaching an interruptive exchange show an increase in their co-speech movement, possibly attempting to keep the floor or possibly creating a visual scenario that listeners see as ripe for interruption.

Overall, a high level of activation of interactional management and negotiation is exhibited in this dataset. We conclude that this interactional navigation may be facilitated in part by the patterning of co-speech movement across interlocutors that this study is able to analyze quantitatively for the first time. Furthermore, with the ability to examine both brow and head movements in conjunction, the kinematic data indicate that brow and whole-head movement densities tend to behave similarly across exchange types and conversational roles. Lastly, our findings based on large quantities of (non-read) dyadic speech have implications for the likelihood of any role of co-speech (non-manual) gesture in facilitating turn end prediction in that when approaching a floor exchange as the sole talker, no reliable changes in the amount of co-speech movement on the part of speakers are observed. Taken together, the study is an initial step in characterizing how speakers' and listeners' co-speech movements jointly pattern in dyadic conversational interaction.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: http://dx.doi.org/10.17632/jy5t72fd32.4 Gordon Danner, Samantha (2021), "Dataset for Co-speech Movement in Conversational Turntaking," Mendeley Data, V4, doi: 10.17632/jy5t72fd32.4.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the University of Southern California Institutional Review Board. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

SD: Conceptualization (supporting); Writing—original draft (lead), revisions (co-lead); Data curation (lead); Formal analysis (lead); Investigation (equal); Methodology (equal); Project administration (supporting); Resources (supporting); Software (lead); Supervision (supporting); Validation (lead); Visualization (lead). JK: Conceptualization (lead); Writing—original draft (supporting), revisions (co-lead); Formal analysis (supporting); Investigation (equal); Methodology (equal); Visualization (supporting). DB: Conceptualization (lead); Writing—original draft (supporting), revisions (co-lead); Formal analysis (supporting); Funding acquisition (lead); Investigation (equal); Methodology (equal); Project administration (lead); Resources (lead); Supervision (lead); Visualization (supporting).

FUNDING

This work was supported by NIH DC003172 (Byrd).

ACKNOWLEDGMENTS

We thank Mathias Barthel, James P. Trujillo, Nynaeve Perkins Booker, Lena Foellmer, Louis Goldstein, Sarah Harper, Cheyenne LaRoque, and Sungbok Lee for their assistance.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcomm.2021.779814/full#supplementary-material

REFERENCES

Alibali, M. W., Kita, S., and Young, A. J. (2000). Gesture and the Process of Speech Production: We Think, Therefore We Gesture. *Lang. Cogn. Process.* 15, 593–613. doi:10.1080/016909600750040571

Barbosa, A. v., Yehia, H. C., and Vatikiotis-Bateson, E. (2008). "Linguistically Valid Movement Behavior Measured Non-invasively," in *Auditory Visual Speech Processing*. Editors R. Gucke, P. Lucey, and S. Lucey (Queensland, Australia), 173–177.

Barkhuysen, P., Krahmer, E., and Swerts, M. (2008). The Interplay between the Auditory and Visual Modality for End-Of-Utterance Detection. *The J. Acoust. Soc. America* 123, 354–365. doi:10.1121/1.2816561

Barthel, M., Meyer, A. S., and Levinson, S. C. (2017). Next Speakers Plan Their Turn Early and Speak after Turn-Final "Go-Signals". *Front. Psychol.* 8, 1–10. doi:10.3389/fpsyg.2017.00393

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *J. Stat. Soft.* 67 (1), 1–48. doi:10.18637/jss.v067.i01

Bavelas, J. B., Chovil, N., Coates, L., and Roe, L. (1995). Gestures Specialized for Dialogue. *Pers Soc. Psychol. Bull.* 21, 394–405. doi:10.1177/0146167295214010

Bavelas, J. B., Chovil, N., Lawrie, D. A., and Wade, A. (1992). Interactive Gestures. *Discourse Process.* 15, 469–489. doi:10.1080/01638539209544823

Bavelas, J. B., Coates, L., and Johnson, T. (2002). Listener Responses as a Collaborative Process: The Role of Gaze. *J. Commun.* 52, 566–580. doi:10.1111/j.1460-2466.2002.tb02562.x

Bavelas, J., Gerwing, J., Sutton, C., and Prevost, D. (2008). Gesturing on the Telephone: Independent Effects of Dialogue and Visibility. *J. Mem. Lang.* 58, 495–520. doi:10.1016/j.jml.2007.02.004

Boersma, P., and Weenink, D. (2016). Praat: Doing Phonetics by Computer. Available at: http://www.praat.org/.

Bögels, S., and Torreira, F. (2015). Listeners Use Intonational Phrase Boundaries to Project Turn Ends in Spoken Interaction. *J. Phonetics* 52, 46–57. doi:10.1016/j.wocn.2015.04.004

Bolinger, D. (1983). Intonation and Gesture. *Am. Speech* 58, 156–174. doi:10.2307/455326

Borràs-Comes, J., Kaland, C., Prieto, P., and Swerts, M. (2014). Audiovisual Correlates of Interrogativity: A Comparative Analysis of Catalan and Dutch. *J. Nonverbal Behav.* 38, 53–66. doi:10.1007/s10919-013-0162-0

Clark, H. H., and Krych, M. A. (2004). Speaking while Monitoring Addressees for Understanding. *J. Mem. Lang.* 50, 62–81. doi:10.1016/j.jml.2003.08.004

Cummins, F. (2012). Gaze and Blinking in Dyadic Conversation: A Study in Coordinated Behaviour Among Individuals. *Lang. Cogn. Process.* 27, 1525–1549. doi:10.1080/01690965.2011.615220

Cvejic, E., Kim, J., Davis, C., and Gibert, G. (2010). "Prosody for the Eyes: Quantifying Visual Prosody Using Guided Principal Component Analysis," in Proceedings of the 11th Annual Conference of the International Speech Communication Association, Makuhari, Chiba, Japan, September 26-30, 2010, 1433–1436.

Danner, S. G. (2017). *Effects of Speech Context on Characteristics of Manual Gesture*. Doctoral dissertation Los Angeles, (CA): University of Southern California in Los Angeles.

Danner, S. G., Barbosa, A. V., and Goldstein, L. (2018). Quantitative Analysis of Multimodal Speech Data. *J. Phonetics* 71, 268–283. doi:10.1016/j.wocn.2018.09.007

Davies, M. (2008). The Corpus of Contemporary American English: 450 Million Words, 1990-present. Available at: http://corpus.byu.edu/coca/.

Dideriksen, C., Fusaroli, R., Tylén, K., Dingemanse, M., and Christiansen, M. H. (2019). "Contextualizing Conversational Strategies: Backchannel, Repair and Linguistic Alignment in Spontaneous and Task-Oriented Conversations," in *Proceedings of the 41st Annual Conference of the Cognitive Science Society : Creativity + Cognition + Computation*. Editors A. K. Goel, C. M. Seifert, and C. Freksa (Canada: Montreal), 261–267. doi:10.31234/osf.io/fd8y9

Duncan, S., Jr. (1972). Some Signals and Rules for Taking Speaking Turns in Conversations. *J. Personal. Soc. Psychol.* 23, 283–292. doi:10.1037/h0033031

Flecha-García, M. L. (2010). Eyebrow Raises in Dialogue and Their Relation to Discourse Structure, Utterance Function and Pitch Accents in English. *Speech Commun.* 52, 542–554. doi:10.1016/j.specom.2009.12.003

Freelon, D. (2013). ReCal OIR: Ordinal, Interval, and Ratio Intercoder Reliability as a Web Service. *Int. J. Internet Sci.* 8, 10–16. Available at: http://www.ijis.net/ijis8_1/ijis8_1_freelon_pre.html.

Fuchs, S., and Reichel, U. D. (2016). "On the Relationship between Pointing Gestures and Speech Production in German Counting Out Rhymes : Evidence from Motion Capture Data and Speech Acoustics," in *Proceedings of P&P 12*. Editors C. Draxler and F. Kleber Munich: Ludwig Maximilian University, 1–4.

Gamer, M., Lemon, J., Fellows, I., and Singh, P. (2019). Various Coefficients of Interrater Reliability and Agreement. Available at: https://cran.r-project.org/package=irr.

Garrod, S., and Pickering, M. J. (2015). The Use of Content and Timing to Predict Turn Transitions. *Front. Psychol.* 6, 1–12. doi:10.3389/fpsyg.2015.00751

Geluykens, R., and Swerts, M. (1992). Prosodic Topic- and Turn-Finality Cues. Proceedings of the IRCS Workshop on Prosody in Natural Speech, Netherlands, 63–70.

Gillespie, M., James, A. N., Federmeier, K. D., and Watson, D. G. (2014). Verbal Working Memory Predicts Co-speech Gesture: Evidence from Individual Differences. *Cognition* 132, 174–180. doi:10.1016/j.cognition.2014.03.012

Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., and Wagner, S. (2001). Explaining Math: Gesturing Lightens the Load. *Psychol. Sci.* 12, 516–522. doi:10.1111/1467-9280.00395

Gordon Danner, S., Krivokapić, J., and Byrd, D. (2021). Dataset for Co-speech Movement in Conversational Turn-Taking. *V*. doi:10.17632/jy5t72fd32.1

Goujon, A., Bertrand, R., and Tellier, M. (2015). *Eyebrows in French Talk-In-Interaction* Nantes, France.

Granum, M. (2017). NurseryRhymes.org - Nursery Rhymes with Lyrics and Music. Available at: https://www.nurseryrhymes.org/(Accessed November 6, 2018).

Guaïtella, I., Santi, S., Lagrue, B., and Cavé, C. (2009). Are Eyebrow Movements Linked to Voice Variations and Turn-Taking in Dialogue? an Experimental Investigation. *Lang. Speech* 52, 207–222. doi:10.1177/0023830909103167

Gullberg, M. (2010). Language-specific Encoding of Placement Events in Gestures. *Event Representation Lang. Cogn.* 11, 166–188. doi:10.1017/CBO9780511782039.008

Hadar, U., Steiner, T. J., and Clifford Rose, F. (1985). Head Movement during Listening Turns in Conversation. *J. Nonverbal Behav.* 9, 214–228. doi:10.1007/bf00986881

Hadar, U., Steiner, T. J., Grant, E. C., and Rose, F. C. (1983). Kinematics of Head Movements Accompanying Speech during Conversation. *Hum. Movement Sci.* 2, 35–46. doi:10.1016/0167-9457(83)90004-0

Hilton, K. (2018). What Does an Interruption Sound like? Stanford University

Hoetjes, M., Krahmer, E., and Swerts, M. (2014). Does Our Speech Change when We Cannot Gesture? *Speech Commun.* 57, 257–267. doi:10.1016/j.specom.2013.06.007

Holler, J., Kendrick, K. H., and Levinson, S. C. (2017). Processing Language in Face-To-Face Conversation: Questions with Gestures Get Faster Responses. *Psychon. Bull. Rev.* 25, 1900–1908. doi:10.3758/s13423-017-1363-z

Holler, J., Schubotz, L., Kelly, S., Hagoort, P., Schuetze, M., and Özyürek, A. (2014). Social Eye Gaze Modulates Processing of Speech and Co-speech Gesture. *Cognition* 133, 692–697. doi:10.1016/j.cognition.2014.08.008

Hömke, P., Holler, J., and Levinson, S. C. (2018). Eye Blinks Are Perceived as Communicative Signals in Human Face-To-Face Interaction. *PLoS ONE* 13, e0208030–13. doi:10.1371/journal.pone.0208030

Ishi, C. T., Ishiguro, H., and Hagita, N. (2014). Analysis of Relationship between Head Motion Events and Speech in Dialogue Conversations. *Speech Commun.* 57, 233–243. doi:10.1016/j.specom.2013.06.008

Kelly, S. D., Özyürek, A., and Maris, E. (2010). Two Sides of the Same Coin. *Psychol. Sci.* 21, 260–267. doi:10.1177/0956797609357327

Kendon, A. (1972). "Some Relationships between Body Motion and Speech," in *Studies In Dyadic Communication*. Editors A. W. Siegman and B. Pope (New York: Pergamon Press), 177–210. doi:10.1016/b978-0-08-015867-9.50013-7

Kim, J., Cvejic, E., and Davis, C. (2014). Tracking Eyebrows and Head Gestures Associated with Spoken Prosody. *Speech Commun.* 57, 317–330. doi:10.1016/j.specom.2013.06.003

Kita, S. (2009). Cross-cultural Variation of Speech-Accompanying Gesture: A Review. *Lang. Cogn. Process.* 24 (2), 145–167. doi:10.1080/01690960802586188

Kita, S., and Özyürek, A. (2003). What Does Cross-Linguistic Variation in Semantic Coordination of Speech and Gesture Reveal?: Evidence for an Interface Representation of Spatial Thinking and Speaking. *J. Mem. Lang.* 48, 16–32. doi:10.1016/S0749-596X(02)00505-3

Krahmer, E., and Swerts, M. (2007). The Effects of Visual Beats on Prosodic Prominence: Acoustic Analyses, Auditory Perception and Visual Perception. *J. Mem. Lang.* 57, 396–414. doi:10.1016/j.jml.2007.06.005

Krauss, R. M. (1998). Why Do We Gesture when We Speak? *Curr. Dir. Psychol. Sci.* 7, 54. doi:10.1111/1467-8721.ep13175642

Kuznetsova, A., Bruun Brockhoff, P., and Haubo Bojesen Christensen, R. (2014). lmerTest: Tests in Linear Mixed Effects Models. Available at: http://cran.r-project.org/package=lmerTest.

Latif, N., Barbosa, A. v., Vatiokiotis-Bateson, E., Castelhano, M. S., and Munhall, K. G. (2014). Movement Coordination during Conversation. *PLoS ONE* 9, e105036. doi:10.1371/journal.pone.0105036

Lee, C.-C., and Narayanan, S. (2010). "Predicting Interruptions in Dyadic Spoken Interactions," in Proceedings of the 2010 IEEE International Conference on Acoustics, Speech and Signal Processing, Dallas, TX, USA, 14-19 March 2010 (Dallas, Texas: IEEE), 5250–5253. doi:10.1109/ICASSP.2010.5494991

Lee, Y., Gordon Danner, S., Parrell, B., Lee, S., Goldstein, L., and Byrd, D. (2018). Articulatory, Acoustic, and Prosodic Accommodation in a Cooperative Maze Navigation Task. *PLOS ONE* 13, e0201444. doi:10.1371/journal.pone.0201444

Leonard, T., and Cummins, F. (2011). The Temporal Relation between Beat Gestures and Speech. *Lang. Cogn. Process.* 26, 1457–1471. doi:10.1080/01690965.2010.500218

Levelt, W. J. M., Richardson, G., and la Heij, W. (1985). Pointing and Voicing in Deictic Expressions. *J. Mem. Lang.* 24, 133–164. doi:10.1016/0749-596X(85)90021-X

Levinson, S. C., and Holler, J. (2014). The Origin of Human Multi-Modal Communication. *Phil. Trans. R. Soc. B* 369, 20130302. doi:10.1098/rstb.2013.0302

Levinson, S. C., and Torreira, F. (2015). Timing in Turn-Taking and its Implications for Processing Models of Language. *Front. Psychol.* 6, 731. doi:10.3389/fpsyg.2015.00731

Levitan, R., Beňuš, Š., Gravano, A., and Hirschberg, J. (2015). Entrainment and Turn-Taking in Human-Human Dialogue. *Turn-taking Coord. Human-Machine Interaction: Pap. 2015 AAAI Spring Symp.*, 44–51.

Loehr, D. P. (2004). Gesture and Intonation. PhD thesis Washington, (DC): Georgetown University.

Magyari, L., Bastiaansen, M. C. M., de Ruiter, J. P., and Levinson, S. C. (2014). Early Anticipation Lies behind the Speed of Response in Conversation. *J. Cogn. Neurosci.* 26, 2530–2539. doi:10.1162/jocn_a_00673

MATLAB (2018). *Version R2018a* Natick, Massachusetts: The MathWorks Inc..

McClave, E. Z. (2000). Linguistic Functions of Head Movements in the Context of Speech. *J. Pragmatics* 32, 855–878. doi:10.1016/S0378-2166(99)00079-X

Melinger, A., and Kita, S. (2007). Conceptualisation Load Triggers Gesture Production. *Lang. Cogn. Process.* 22, 473–500. doi:10.1080/01690960600696916

Mondada, L. (2007). Multimodal Resources for Turn-Taking. *Discourse Stud.* 9, 194–225. doi:10.1177/1461445607075346

Munhall, K. G., Jones, J. A., Callan, D. E., Kuratate, T., and Vatikiotis-Bateson, E. (2004). Visual Prosody and Speech Intelligibility. *Psychol. Sci.* 15, 133–137. doi:10.1111/j.0963-7214.2004.01502010.x

Munhall, K. G., Ostry, D. J., and Parush, A. (1985). Characteristics of Velocity Profiles of Speech Movements. *J. Exp. Psychol. Hum. Perception Perform.* 11, 457–474. doi:10.1037/0096-1523.11.4.457

Nota, N., Trujillo, J. P., and Holler, J. (2021). Facial Signals and Social Actions in Multimodal Face-To-Face Interaction. *Brain Sci.* 11, 1017. doi:10.3390/brainsci11081017

Ostry, D. J., Cooke, J. D., and Munhall, K. G. (1987). Velocity Curves of Human Arm and Speech Movements. *Exp. Brain Res.* 68, 37–46. doi:10.1007/BF00255232

Özyürek, A., Kita, S., Allen, S. E. M., Furman, R., and Brown, A. (2005). How Does Linguistic Framing of Events Influence Co-speech Gestures? *Gest* 5, 219–240. doi:10.1075/gest.5.1.15ozy

Prieto, P., Puglesi, C., Borràs-Comes, J., Arroyo, E., and Blat, J. (2015). Exploring the Contribution of Prosody and Gesture to the Perception of Focus Using an Animated Agent. *J. Phonetics* 49, 41–54. doi:10.1016/j.wocn.2014.10.005

R Core Team (2021). R: A Language and Environment for Statistical Computing. Available at: https://www.gbif.org/tool/81287/r-a-language-and-environment-for-statistical-computing. doi:10.1007/978-3-540-74686-7

Roberts, S. n. G., Torreira, F., and Levinson, S. C. (2015). The Effects of Processing and Sequence Organization on the Timing of Turn Taking: a Corpus Study. *Front. Psychol.* 6, 509. doi:10.3389/fpsyg.2015.00509

Rochet-Capellan, A., and Fuchs, S. (2014). Take a Breath and Take the Turn: How Breathing Meets Turns in Spontaneous Dialogue. *Phil. Trans. R. Soc. B* 369, 20130399. doi:10.1098/rstb.2013.0399

Ruiter, J.-P. d., Mitterer, H., and Enfield, N. J. (2006). Projecting the End of a Speaker's Turn: A Cognitive Cornerstone of Conversation. *Language* 82, 515–535. doi:10.1353/lan.2006.0130

Scobbie, J. M., Turk, A., Geng, C., King, S., Lickley, R., and Richmond, K. (2013). "The edinburgh Speech Production Facility Doubletalk Corpus," in Proceedings of the Annual Conference of the International Speech Communication Association, August 2013, 764–766.

Sikveland, R. O., and Ogden, R. (2012). Holding Gestures across Turns. *Gest* 12, 166–199. doi:10.1075/gest.12.2.03sik

Singmann, H., Bolker, B., Westfall, J., Aust, F., and Ben-Shachar, M. S. (2021). *Afex: Analysis of Factorial Experiments*. R package version 1.0-1. Available at: https://CRAN.R-project.org/package=afex.

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., et al. (2009). Universals and Cultural Variation in Turn-Taking in Conversation. *Proc. Natl. Acad. Sci.* 106, 10587–10592. doi:10.1073/pnas.0903616106

Tiede, M., Bundgaard-Nielsen, R., Kroos, C., Gibert, G., Attina, V., Kasisopa, B., et al. (2010). Speech Articulator Movements Recorded from Facing Talkers Using Two Electromagnetic Articulometer Systems Simultaneously. *J. Acoust. Soc. America* 128, 1–9. doi:10.1121/1.3508805

Trujillo, J. P., Levinson, S. C., and Holler, J. (2021). "Visual Information in Computer-Mediated Interaction Matters: Investigating the Association between the Availability of Gesture and Turn Transition Timing in Conversation," in *Human-Computer Interaction. Design and User Experience Case Studies. HCII 2021. Lecture Notes in Computer Science*. Editor M. Kurosu (Berlin/Heidelberg, Germany: Springer), 12764, 643–657. doi:10.1007/978-3-030-78468-3_44

Vilela Barbosa, A., Déchaine, R.-M., Vatikiotis-Bateson, E., and Camille Yehia, H. (2012). Quantifying Time-Varying Coordination of Multimodal Speech Signals Using Correlation Map Analysis. *J. Acoust. Soc. America* 131, 2162–2172. doi:10.1121/1.3682040

Voigt, R., Eckert, P., Jurafsky, D., and Podesva, R. J. (2016). Cans and Cants: Computational Potentials for Multimodality with a Case Study in Head Position. *J. Sociolinguistics* 20, 677–711. doi:10.1111/josl.12216

Wagner, P., Malisz, Z., and Kopp, S. (2014). Gesture and Speech in Interaction: An Overview. *Speech Commun.* 57, 209–232. doi:10.1016/j.specom.2013.09.008

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, J., et al. (2019). Welcome to the Tidyverse. *J. Open Source Softw.* 4 (43), 1686. doi:10.21105/joss.01686

Yehia, H., Rubin, P., and Vatikiotis-Bateson, E. (1998). Quantitative Association of Vocal-Tract and Facial Behavior. *Speech Commun.* 26, 23–43. doi:10.1016/s0167-6393(98)00048-x

Yuan, J., and Liberman, M. (2008). Speaker Identification on the SCOTUS Corpus. *Proc. Acoust. '08*, 6–9. doi:10.1121/1.2935783

Zellers, M., Gorisch, J., House, D., and Peters, B. (2019). "Hand Gestures and Pitch Contours and Their Distribution at Possible Speaker Change Locations : a First Investigation," in *Gesture and Speech in Interaction*. 6th edition (Paderborn: GESPIN), 93–98.

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

frontiers in Psychology

ORIGINAL RESEARCH
published: 05 January 2022
doi: 10.3389/fpsyg.2021.782765

Predictive Processing in Poetic Language: Event-Related Potentials Data on Rhythmic Omissions in Metered Speech

Karen Henrich[1†] and Mathias Scharinger[1,2,3†]*

[1] Department of Language and Literature, Max Planck Institute for Empirical Aesthetics, Frankfurt, Germany, [2] Research Group Phonetics, Philipps-University of Marburg, Marburg, Germany, [3] Center for Mind, Brain, and Behavior, Universities of Marburg and Giessen, Marburg, Germany

Predictions during language comprehension are currently discussed from many points of view. One area where predictive processing may play a particular role concerns poetic language that is regularized by meter and rhyme, thus allowing strong predictions regarding the timing and stress of individual syllables. While there is growing evidence that these prosodic regularities influence language processing, less is known about the potential influence of prosodic preferences (binary, strong-weak patterns) on neurophysiological processes. To this end, the present electroencephalogram (EEG) study examined whether the predictability of strong and weak syllables within metered speech would differ as a function of meter (trochee vs. iamb). Strong, i.e., accented positions within a foot should be more predictable than weak, i.e., unaccented positions. Our focus was on disyllabic pseudowords that solely differed between trochaic and iambic structure, with trochees providing the preferred foot in German. Methodologically, we focused on the omission Mismatch Negativity (oMMN) that is elicited when an anticipated auditory stimulus is omitted. The resulting electrophysiological brain response is particularly interesting because its elicitation does not depend on a physical stimulus. Omissions in deviant position of a passive oddball paradigm occurred at either first- or second-syllable position of the aforementioned pseudowords, resulting in a 2-by-2 design with the factors foot type and omission position. Analyses focused on the mean oMMN amplitude and latency differences across the four conditions. The result pattern was characterized by an interaction of the effects of foot type and omission position for both amplitudes and latencies. In first position, omissions resulted in larger and earlier oMMNs for trochees than for iambs. In second position, omissions resulted in larger oMMNs for iambs than for trochees, but the oMMN latency did not differ. The results suggest that omissions, particularly in initial position, are modulated by a trochaic preference in German. The preferred strong-weak pattern may have strengthened the prosodic prediction, especially for matching, trochaic stimuli, such that the violation of this prediction led to an earlier and stronger prediction error. Altogether, predictive processing seems to play a particular role in metered speech, especially if the meter is based on the preferred foot type.

Keywords: meter, speech, prediction, trochaic preference, ERP, omission MMN

OPEN ACCESS

Edited by:
Ingmar Brilmayer,
University of Cologne, Germany

Reviewed by:
Laura Hahn,
Charité University Medicine Berlin, Germany
Paul Compensis,
University of Cologne, Germany

****Correspondence:***
Mathias Scharinger
mathias.scharinger@aesthetics.mpg.de

[†]These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *24 September 2021*
Accepted: *29 November 2021*
Published: *05 January 2022*

Citation:
Henrich K and Scharinger M (2022) Predictive Processing in Poetic Language: Event-Related Potentials Data on Rhythmic Omissions in Metered Speech. Front. Psychol. 12:782765. doi: 10.3389/fpsyg.2021.782765

INTRODUCTION

Spoken language is based on »quasi-regular« properties, exemplified by physiological and articulatory processes such as the vibration pattern of the vocal folds or the repetitious sequence of consonants and vowels (Greenberg et al., 2003; Reetz and Jongman, 2008). It is not surprising, then, that these regularities are considered within models of speech processing that capitalize on predictions (e.g., Kutas et al., 2011; Pickering and Garrod, 2013; Schröger et al., 2015; Kuperberg and Jaeger, 2016). Predictive language processing is an umbrella term to subsume approaches that focus on context effects on all levels of the linguistic hierarchy. With the rise of frameworks related to the predictive coding theory of human brain function (Friston, 2003, 2005, 2008; Kiebel et al., 2009), these context effects were translated into prediction or expectation effects. Oftentimes, the terms prediction and expectation have been used synonymously. Here, we attempt to distinguish between the more general concept of an expectation as reflecting the anticipation of a higher-order linguistic unit, and the more concrete concept of a prediction as reflecting the temporal and content-based forecast of a specific linguistic unit. For instance, in the sentence "A salmon is a…," the expectation is that an animate noun is following, while the specific prediction is that the word will start with the sound [f] (in "fish").

Regularities in spoken language have a particular relation to predictions, because they allow for these predictions to be sharpened (Schröger et al., 2015; Scharinger et al., 2016). Aside from the quasi-regular properties of speech, specific forms of language use characteristically exploit these regularities. A prime candidate for such language use is poetic language, where in Western tradition, regularities hold on the level of timing (expressed in rhythm and meter) and on the level of phonological, segmental properties (expressed in assonance, consonance, alliteration, and rhyme; Jakobson, 1960; Menninghaus et al., 2017). A third level, that is also crucial for non-poetic language, concerns speech prosody, i.e., all supra-segmental properties of speech such as stress, intonation and melody. Prosodic frameworks allow to describe regular sequences of syllables on the basis of syllable weight (Nespor and Vogel, 1986; Selkirk, 1995). Here, a basic distinction has been made between the pattern of strong syllables followed by weak syllables (SW-pattern, or trochaic pattern), and the pattern of weak syllables followed by strong syllables (WS-pattern, or iambic pattern). Within the prosodic hierarchy, the combination of syllables instantiating these patterns is expressed in foot types, of which trochees and iambs are the most basic ones (Hayes, 1995).

Metrical prosodic structure in speech is of general relevance for segmentation, timing, stress, and lexical access (Jusczyk, 1999; Domahs et al., 2008, 2014; Schmidt-Kassow and Kotz, 2009; Bohn et al., 2013; Molczanow et al., 2013; Roncaglia-Denissen et al., 2013; Henrich et al., 2014; Magne et al., 2016). Violations of even subtle rhythmic preferences, as e.g., expressed by the Rhythm Rule in German, are taxing processing resources (Bohn et al., 2013; Henrich et al., 2014), while adherence to regular rhythm or meter may facilitate lexical access (Magne et al., 2007; Cason and Schön, 2012; Molczanow et al., 2013, 2019). In poetic language, regular meter and rhyme, next to further so-called »parallelistic« properties, can lead to a relative ease of processing and a simultaneous increase of aesthetic appreciation (Obermeier et al., 2013, 2015; Menninghaus et al., 2017).

Experiments investigating the neurophysiological bases of these processing consequences of regular or irregular prosody rely on event-related potentials (ERP) of the human electroencephalogram (EEG). Most of the aforementioned studies focused on a violation response that has been established in the early eighties as electrophysiological index of a semantic context effect (Kutas and Hillyard, 1980, 1984). It was then shown that semantically incongruous sentence endings elicit a distinct negative deflection in the ERP at around 400 ms after word onset. The correspondingly called N400 was initially considered to be an electrophysiological index of lexico-semantic integration, but soon received a broader interpretation in that it could also be elicited by contexts without semantic violations. In general, ease of (lexico-semantic) processing has been attributed to a decrease in N400 amplitude (Chwilla et al., 1995; Franklin et al., 2007; Lau et al., 2013).

Several studies have shown that ease of processing is not only determined by suitable semantic context but also by regular prosody (e.g., meter, see Rothermich et al., 2010; Rothermich and Kotz, 2013). A further important observation of these and similar studies is that certain prosodic patterns (such as SW vs. WS) are preferred in some, if not all languages. The SW-pattern in trochees is considered the preferred pattern or foot type in German (Wiese and Speyer, 2015). Next to preferences for a certain foot type, there also preferences as to how syllable weight determining the respective types is related to prosodic properties. Here, the so-called Iambic-Trochaic law (ITL) stipulates that rhythmic grouping strategies show a basic difference between iambs and trochees: while longer sounds or syllables tend to be assigned to group (i.e., foot) endings, louder sounds or syllables are rather assigned to group (i.e., foot) beginnings (Hay and Diehl, 2007; de la Mora et al., 2013; Crowhurst and Olivares, 2014; Crowhurst, 2020). Put differently, a typical trochee consists of a syllable with high intensity, followed by a syllable with less intensity, while a typical iamb consists of a shorter syllable followed by longer syllable. Depending on task and stimulus material, the marking of group beginnings can also be achieved by fundamental frequency (f0), or more precisely, a relative higher pitch (Crowhurst and Olivares, 2014; Crowhurst, 2020).

Electrophysiological studies focusing on the rhythmic structure of language rarely distinguish between different foot types. Of the few, Breen et al. (2019) analyzed violations of SW (trochaic) patterns as compared to WS (iambic) patterns in a reading study with EEG. Violations were realized by incongruencies between a couplet context and a target word. For trochaic violations, they found two negativities, one of which showed similarities to the N400. For iambic violations, only a positivity was elicited. This suggests that a violation of a trochaic expectancy resulted in enhanced processing effort, possibly caused by a stronger expectation in the trochaic as compared to the iambic case.

Brochard et al. (2003) were interested whether subjective accenting of identical tone sequences would yield a trochaic pattern and whether processing of stimulus changes in allegedly strong positions would differ from processing in allegedly weak positions. They employed a so-called oddball paradigm in which multiple identical tones were repeated (standards), interspersed by infrequent tones with decreased loudness in either odd- (i.e., strong) or even-numbered (i.e., weak) positions of the sequences (deviants). Oddball paradigms elicit typical ERP-responses to both deviants and standards, and an additional mismatch response to the deviant, best seen in the difference wave form between deviant ERP and standard ERP. This response is called Mismatch Negativity (MMN), typically elicited by physical stimulus changes as well as violations of higher-order regularities (Näätänen, 1995; Näätänen and Alho, 1997; Winkler, 2007). Brochard et al. (2003) demonstrated that the ERP response to deviants in odd-numbered (strong) positions (and thus, the MMN) was stronger compared to the response to deviants in even-numbered (weak) positions. Subjective accenting derived from a trochaic preference thus seems to modulate the prediction of prosodic properties (here: loudness).

The MMN has been interpreted within predictive coding frameworks, since its elicitation is thought to reflect the prediction error between the perceived stimulus and the internal model (aka the prediction), triggered by the repeating standard (Baldeweg, 2006; Winkler, 2007). As the MMN has been shown to be modulated by long-term experience with sounds in general and with speech sounds in particular (Näätänen et al., 1997; Dehaene-Lambertz et al., 2000), it is plausible to assume that prosodic preferences would similarly modulate the MMN. The study by Brochard et al. (2003) provides an important example in this respect. However, other than in the study by Brochard et al. (2003), an even more direct index of the assumed prediction error is the ERP response to a sound omission in predictive contexts. The so-called omission MMN was initially found to reflect the prediction error when a predicted tone was omitted (Tervaniemi et al., 1994; Yabe et al., 1997; Horváth et al., 2010; Salisbury, 2012), but later work showed that the omission of predicted speech sounds can also elicit the omission MMN (Bendixen et al., 2014; Scharinger et al., 2017). In the study by Bendixen et al. (2014), predictability of word-final [ks] and [ts] in the German noun "Lachs" (salmon) and "Latz" (bib) was modified by either presenting only "Lachs" or "Latz" in standard position of an oddball paradigm (predictive condition), or by randomly presenting "Lachs" and "Latz" with a 50% probability of either noun (unpredictive condition). Deviants consisted of word fragments of which the word-final consonants were omitted. The omission MMN differed between the predictive and unpredictive condition, and showed larger amplitudes in the predictive condition.

The latter study as well as previous experiments on long-term memory effects on the MMN provide the basis of our assumptions here. We hypothesize that the omission MMN between 100 and 200 post-stimulus onset (Bendixen et al., 2014; Scharinger et al., 2017) is not only modulated by segmental information, but also by prosodic information, and thus, can index violations of prosodic predictions. More concretely, on the basis of Brochard et al. (2003) we would assume that the omission of sounds in strong positions results in stronger omission responses than the omission of sounds in weak positions. We furthermore expect that the omission MMN is also sensitive to patterns of strong and weak syllables (i.e., higher-order regularities), and therefore we hypothesize that the omission of sounds in strong positions of trochaic patterns lead to the strongest omission response. Trochees can therefore instantiate the strongest metrical predictions that we intend to test by electrophysiological means, using disyllabic pseudowords with trochaic and iambic patterns and with syllable omissions occurring in either first or second position of these pseudowords. To be precise, we expect that this 2 × 2-design would show an interaction of the effects of position of omission (first syllable, second syllable) and foot type (trochee, iamb).

MATERIALS AND METHODS

Participants

Participants were twenty native speakers of German, recruited from the participant database of the Max Planck Institute (12 females, 8 males, average age 25 ± 5 years). The sample size was based on previous studies with similar designs (Colin et al., 2009). All participants were right-handed, with scores >90% on the Edinburgh Handedness Inventory (Oldfield, 1971). None of the participants reported a history of hearing or neurological problems and participated for monetary compensation (€ 10 per hour). The study was approved by the local Ethics Committee and in accordance with the declarations of Helsinki. Prior to the experiment, participants provided written informed consent and were informed about legal aspects of the study as well as data handling policies in written and spoken form.

Materials

Trochaic and iambic stimuli were disyllabic pseudowords, starting with the voiced velar stop [g] and followed by the round, back high vowel [u], i.e., "gugu." First, complete pseudowords were recorded in the carrier-sentence "Er soll nun gugu sagen (he shall say gugu now)," with "gugu" either pronounced with a strong initial syllable ($N = 10$) or a strong final syllable ($N = 10$). Carrier-sentences and pseudowords were spoken by a phonetically trained female speaker and recorded with 44.1 kHz temporal and 16 bit amplitude resolution in a silent recording chamber of the Max-Planck-Institute for Empirical Aesthetics in Frankfurt (Germany). From the entire set of 20 recordings, we selected those gu-syllables that had the most comparable pitch changes between strong and weak versions and showed the least difference in intensity. We decided to use stimuli that approximate typical trochaic and iambic disyllabic words without differing too much in acoustic terms, for any change of acoustic properties would modulate the omission MMN. We arrived at four syllables from one trochaic and one iambic pseudoword, of which the weak syllables had a very comparable pitch contour, differing from the strong counterparts by about 35 Hz in average pitch height. Final full-word stimuli were cross-spliced in that the original strong syllable from the selected trochaic pseudoword was combined with the weak syllable from the selected iambic pseudoword, resulting in a trochaic cross-spliced test stimulus. Vice versa, the weak syllable from the trochaic pseudoword was combined with the strong syllable of the iambic pseudoword, resulting in an iambic cross-spliced test stimulus. All syllables were trimmed to 250 ms with the phonetic software PRAAT, using the overlap-add algorithm. This was done in order to avoid MMN asymmetries that arise solely by differences in stimulus or stimulus part durations. Longer stimuli in deviant compared to standard position elicit a smaller MMN than vice versa, i.e., shorter stimuli in deviant compared to standard position (Takegata et al., 2008; Colin et al., 2009). Furthermore, all syllables were set to an internal intensity of 70 dB, corresponding to a comfortable listening level at ∼70 dB SPL when played during the experiment. Wave forms and pitch tracks of the experimental full-word stimuli are displayed in **Figure 1A**. Due to identical syllable durations, each disyllabic word had a duration of 500 ms.

We also analyzed the phonetic timing in the trochaic and iambic words. Due to the cross-splicing, this timing was identical across conditions. First, closure durations as measure from technical stimulus beginning until onset of the consonantal burst were 50 ms. Second, the time from the onset of the consonantal burst until the beginning of the vowels was 40 ms. In acoustic terms, this means that syllables were separated by 50 ms-pauses (corresponding to the consonantal closure durations).

Omissions were realized as syllable omissions. All omissions were created by truncating the cross-spliced pseudowords at their respective mid-points. For instance, an omission in first-syllable position of a trochee resulted in a weak syllable that originally stemmed from an iamb. In total, due to two foot types and two positions, four truncated pseudowords realized the four types of omissions.

Design

The stimulus material was arranged in a typical oddball paradigm, where stimuli could occur in standard or deviant position, distributed over several blocks. Blocks as displayed in **Figure 1B**. were further split in half in order to guarantee manageable experiment times. Thus, in 2 × 4 blocks, standards consisted of disyllabic (full) pseudowords (**Figure 1B**) and deviants of truncated pseudowords. Truncations resulted in either first-syllable omissions or second-syllable omissions. If syllables were omitted in first position, the deviant effectively started with silence. In each block, there were 350 standards and 50 deviants (translating into 87.5% standards and 12.5% deviants). The stimulus material was pseudo-randomized, with different randomization for each participant. Constraints on randomizations were as follows: (1) minimally four consecutive standards; (2) maximally 10 consecutive standards; (3) no immediate repetition of identical standard numbers, e.g., five standards and then five standards again. The first three standards per block and standards immediately following a deviant were discarded from further analyses. The stimulus material, arranged in 2 × 4 blocks, therefore constituted a 2 × 2 design, with two levels of foot type (trochee, iamb) and two levels of omission position (first syllable, second syllable). In order to match the number of occurring strong and weak syllables, we additionally included four blocks where standards were single syllables (strong and weak syllables from trochees and iambs) and deviants were full (i.e., disyllabic) pseudowords. These four blocks were not analyzed further.

All stimuli were presented with a constant inter-stimulus interval of 300 ms. This means that inter-stimulus differences were 300 ms (measured from the end of the second syllable of one disyllabic word to the beginning of the first syllable of the next disyllabic word). This translates into a Stimulus Onset Asynchrony (SOA) of 800 ms. Note that deviants with word-initial omissions effectively resulted in an SOA of 1,050 ms measured from the beginning of the standard immediately before the deviant and the beginning of the truncated deviant syllable.

Procedure

Stimuli were presented over open-field loudspeakers placed symmetrically 1 m in front of the participants. Participants were seated in electrically and acoustically shielded EEG-cabins. Next to the loudspeakers, a flat-screen was placed 1.2 m in front of the participants. This screen was used to display a silent movie during the passive oddball paradigm.

After EEG-setup, participants passively listened to the 12 blocks of standard-deviant trains. There was no task except the request to ignore the sounds as best as possible, while watching a silent movie (without subtitles). After each block, the experimenter allowed for a short break. In the middle of the experiment, the break was longer and the air in the EEG cabin was refreshed. Each block lasted for about 5 1/2 min; the entire experiment in the cabin about 65 min.

Electroencephalogram Recording

Continuous EEG was recorded from 64 Ag/AgCl electrodes, arranged on a nylon cap following the extended 10–20 system (Oostenveld et al., 2011). EEG signals were amplified with a BIOSEMI ActiveTwo amplifier. Two electrodes placed left and right posterior to Cz were used as online-reference and as ground during the recording. EEG signals were recorded with a sampling rate of 500 Hz and filtered between DC and 250 Hz within the ActiveView BIOSEMI software.

Electroencephalogram Pre-processing and Analysis

Electroencephalogram raw data were analyzed within fieldtrip (Oostenveld et al., 2011), running on Matlab (Mathworks, 2016). Electrophysiological responses were analyzed in time windows from 200 pre-stimulus onset to 800 ms post-stimulus onset. These epochs were defined on the basis of the full disyllabic words and underwent automatic artifact detection implemented within fieldtrip (Oostenveld et al., 2011). This involved detecting muscle and eye-movement (electro-oculogram) artifacts as well as epochs with amplitudes exceeding 150 μV (peak-to-peak). Automatic artifact detection led to the exclusion of individual epochs, but in no participant or condition did the exclusion rate exceed 25% of the total number of epochs (mean exclusion rate: 9.27%). Subsequently, epochs were band-pass filtered between

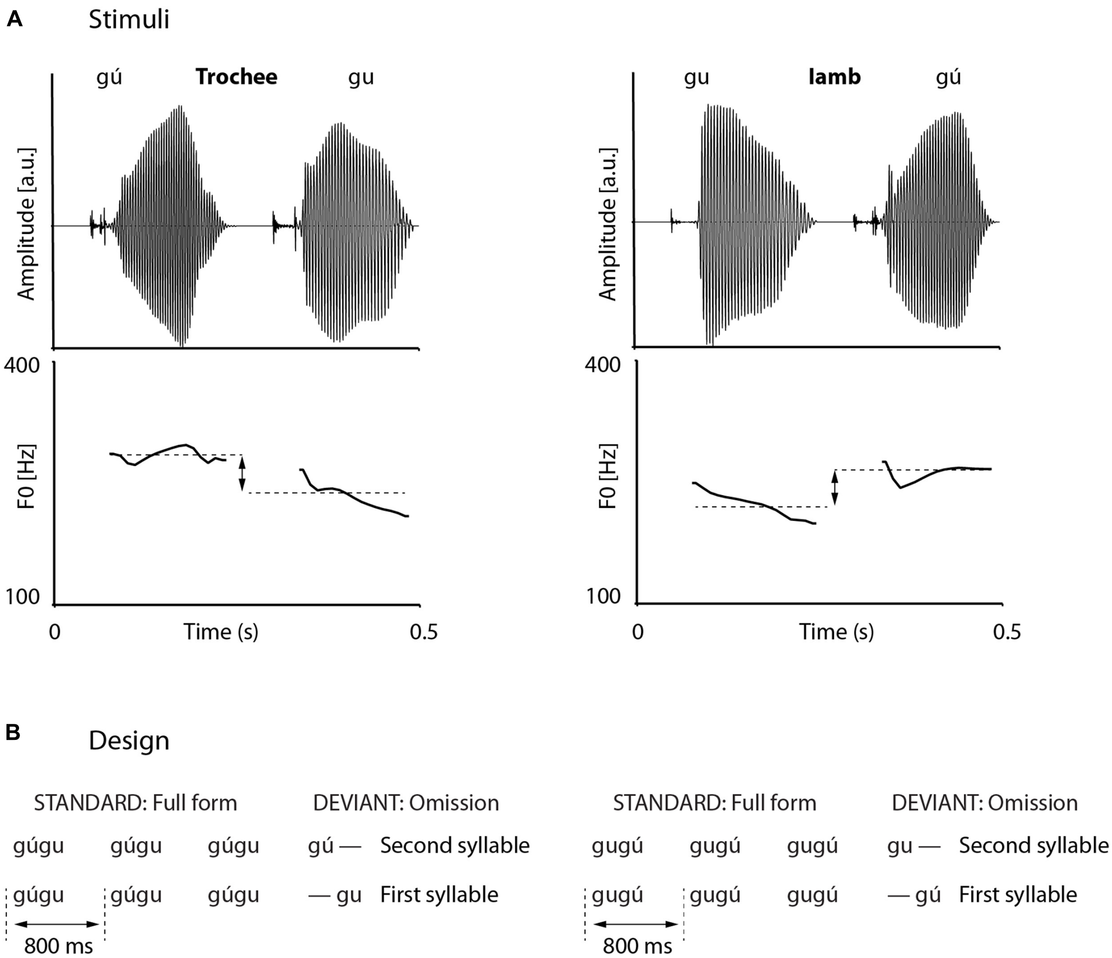


FIGURE 1 | Stimulus material and design. **(A)** Pseudoword stimuli consist of cross-spliced strong (gú) and weak (gu) syllables, combined to a trochee (left) and an iamb (right). Intensities and durations are normalized, pitch tracks are similar but mean pitch height between strong and weak syllables differs by about 35 Hz. **(B)** Oddball design. Full-word deviants were interspersed by truncated deviants (omissions). Omissions could occur in word-initial or word-final syllable, and in trochees or iambs.

0.3 and 30 Hz (Hamming-window digital Butterworth filter) and re-referenced to electrodes in close proximity to the mastoids (TP9, TP10) in order to approximate a linked-mastoid reference, as is common for MMN studies (Näätänen and Alho, 1997; Schröger, 2005; Winkler, 2007). For baseline correction, the mean amplitude of the pre-stimulus window (−200 to 0 ms) was subtracted from the epoch. Responses to standards and deviants in the first-syllable and second-syllable omission conditions were then averaged separately.

Statistical Analyses

The mismatch negativity is defined as the difference between deviant and standard responses. In order to establish electrodes and time-points at which differences between standard and deviant responses are indeed significant, we used a multi-level, non-parametric cluster statistics approach (Henry and Obleser, 2012; Strauß et al., 2014), implemented in fieldtrip (Oostenveld et al., 2011). At the first level, we calculated independent-samples t-tests between single-trial amplitude values for standards and single-trial amplitude values for deviants, separately for the first-syllable and the second-syllable omission conditions. We thereby obtained uncorrected by-participant t-values for all time points and all electrodes. These t-values were subsequently tested against zero using dependent-sample t-tests at the second, i.e., group level, of our cluster-analysis. We estimated type I-error controlled cluster significance probabilities (at $p < 0.05$) by a Monte-Carlo non-parametric permutation method with 1,000 randomizations. The resulting matrix of t-values (electrodes × time points) was then analyzed between 100 and 200 ms post word onset for the first-syllable omission condition, and between 350 and 450 ms post word onset for the second-syllable omission condition. These time windows represent the expected temporal location of the omission MMN, measured from stimulus onset (Bendixen et al., 2014; Scharinger et al., 2017). Within these time windows, electrodes-time point clusters were determined by neighboring electrodes and neighboring time points for which t-values were above the significance threshold ($p < 0.05$). In the first-syllable omission condition, this led to a cluster of 20 electrodes, showing significant standard-deviant differences between 130 and 180 ms post-stimulus onset (**Figure 1**). In the second-syllable omission condition, we obtained a cluster of 28 electrodes, yielding significant standard-deviant differences between 400 and 450 ms post-stimulus onset (**Figure 1**). Note that the latter time window corresponds to time points between 150 and 200 ms post-deviance onset. The final electrode selection for further analyses was then based on the intersection of the two electrode clusters, yielding 18 fronto-central electrodes (AF3, AF4, AF7, F1, F2, F3, F4, F5, F6, F7, FC1, FC2, FC3, FC4, FC5, FC6, FCz, and Fz).

Next, we calculated the omission MMN as difference between deviant and standard responses for the aforementioned electrodes, and in the two temporal regions as determined from the cluster statistics, separately for each participant and meter type (trochee, iamb). This resulted in mean MMN values for each participant, electrode, omission position and meter. Additionally, within the two time windows of the omission MMN, we determined the peak amplitude and the time point (latency) of this peak amplitude. Peak amplitudes were selected automatically by determining the minimum value of the deviant-standard difference in the respective time windows, and by manually inspecting the plausibility of the peaks. The automatic approach performed well, and only in three cases manual adjustment was necessary.

Cortical sources of the omission MMN were estimated using Variable Resolution Electromagnetic Tomography (VARETA; Bosch-Bayard et al., 2001; Scharinger et al., 2017). The VARETA algorithm attempts a reconstruction of cortical sources by looking for a discrete spline-interpolated solution to the EEG inverse problem. This is achieved by obtaining estimates of the spatially smoothest intracranial primary current density (PCD) distribution that is compatible with the observed scalp voltage distribution. Possible solutions are restricted to gray matter on the basis of the probabilistic brain tissue maps available from the Montreal Neurological Institute (MNI, Evans et al., 1993). First, possible sources are modeled as a pre-defined grid of voxels with 7 mm spacing. The 64 electrodes were co-registered with the average probabilistic brain atlas developed at the MNI, assuming a head radius of 85 mm. The difference ERPs of standards and deviants in the MMN time window as established by the cluster statistics were transformed into source space. Statistical parametric maps (SPMs) of the PCD estimates were then constructed based on a voxel-by-voxel Hotelling T^2 test against zero (with $df = 19$).

Omission MMN mean amplitudes, peak amplitudes and latencies were then submitted to linear-effects mixed models (LMMs), calculated with the statistical software R (R Development Core Team, Vienna, Version 3.2.2). Results are reported as mixed-effects analysis of variance (ANOVAs) with F-values that were estimated by the lmerTest package (Kuznetsova et al., 2014), using the Satterthwaite's method. These models used the fixed effects POSITION (omission of first syllable, omission of second syllable), FOOT TYPE (Trochee, Iamb), ELECTRODE (AF3, AF4, AF7, F1, F2, F3, F4, F5, F6, F7, FC1, FC2, FC3, FC4, FC5, FC6, FCz, and Fz) and the random effect SUBJECT in a full-factorial design (i.e., including all possible interactions).

RESULTS

Amplitudes

Omission MMNs were reliably elicited in the typical time windows between 100 and 200 ms after deviance onset (between 100–200 ms and 350–450 ms post-stimulus onset, **Figure 2**). When looking at each expression of the factors POSITION (first vs. second syllable) and FOOT TYPE (trochee, iamb), topographies of omission MMNs showed typical fronto-central distributions, with sources in left and right temporal areas, including primary and secondary auditory cortex, planum temporale and parts of superior and middle temporal gyrus (**Figure 3**).

Statistical analyses on amplitudes are summarized in **Table 1**.

For both mean and peak amplitudes, omission MMNs were larger for omissions in the second syllable (mean amplitude: -2.28 μV, peak amplitude: -3.06 μV) than in the first syllable (mean amplitude: -1.89 μV, peak amplitude: -2.84 μV). The interaction of the effects POSITION and FOOT TYPE also showed similar patterns for mean and peak amplitudes (**Figure 4**). Notably, in first position, trochees elicited larger MMN responses (mean amplitude: -2.05 μV, peak amplitude: -2.95 μV) than iambs (mean amplitude: -1.73 μV, peak amplitude: -2.73 μV), while in second position, iambs elicited larger MMN responses (mean amplitude: -2.50 μV, peak amplitude: -3.29 μV) than trochees (mean amplitude: -1.93 μV, peak amplitude: -2.82 μV). When the interaction was decomposed according to FOOT TYPE, Iambs [mean amplitude: $F_{(1,665)} = 38.97$, $p < 0.001$; peak amplitude: $F_{(1,665)} = 20.57$, $p < 0.001$], but not trochees (all Fs < 1, n.s.), showed higher amplitudes for omissions in second-syllable, compared to omissions in first-syllable position. All effects and interactions were independent of the electrodes.

Latencies

Naturally, MMN latencies differed between omissions of the first and omissions of the second syllable. Overall, iambs showed longer latencies than trochees; this, however, depended on the effect of POSITION, as seen from the decomposition of the interaction of the effects of POSITION and FOOT TYPE (**Table 2**).

The main effect of FOOT TYPE reflected on average an eight millisecond earlier omission MMN for trochees than for iambs. This effect was driven by foot type difference in the first-syllable omission condition, with significantly earlier latencies for trochees than for iambs. Here, omission MMNs occurred at 146 ms for trochees and at 161 ms for iambs. Latencies in the second-syllable omission condition did not differ between trochees and iambs (**Figure 5**).

DISCUSSION

The present study is the first omission MMN study focusing on meter perception in disyllabic speech-like structures. Its most important results on syllable omissions in regular trochaic and iambic contexts can be summarized as follows:

(a) Omissions in both first- and second-syllable position resulted in robust omission MMNs. This replicates the

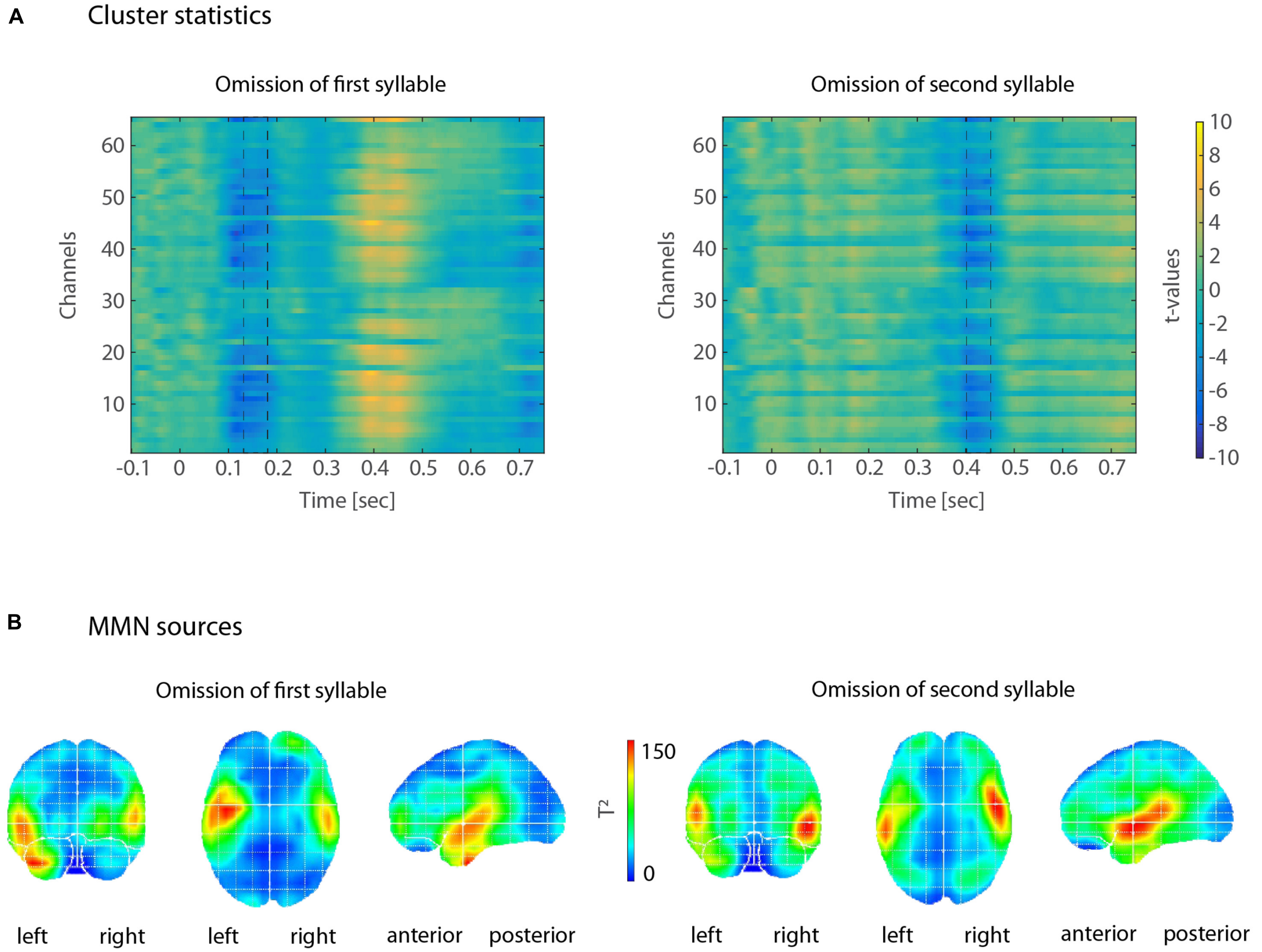


FIGURE 2 | (A) Illustration of cluster statistics by color-coded *t*-values. Left: In the first-syllable omission condition, several electrodes showed more negative responses for deviants than for standards, between 130 and 180 ms post-stimulus onset (indicated by dashed lines). Right: In the second-syllable omission condition, more negative responses for deviants than for standards occurred between 400 and 450 ms post-stimulus onset. Color-coding of *t*-values shows warmer colors for *t*-values > 0 and cooler colors for *t*-values <0. **(B)** Results of the VARETA source reconstructions for MMNs in response to first-syllable omissions (left) and second-syllable omissions (right). Warmer colors represent higher T^2-values. Sources are discernible in bilateral temporal cortices.

findings for omissions of speech sound sequences shorter than syllables (Bendixen et al., 2014; Scharinger et al., 2017) and extends the general feasibility of speech sound omissions to the level of the syllable.

(b) Omissions in second-syllable position resulted in a generally enhanced omission response compared to omissions in first-syllable position. This, however, depended on foot type and only held for iambs.

(c) Within first-syllable and second-syllable position, the main effect of FOOT TYPE indicated that first-syllable omissions resulted in larger MMNs for trochees than for iambs, and that second-syllable omissions resulted in larger MMNs for iambs than for trochees. This pattern corresponds to the weight-carrying syllable in the two foot types, with trochees consisting of a weight-carrying syllable in first position and with iambs consisting of a weight-carrying syllable in second position.

(d) The latency results suggest that an omission in first-syllable position was detected earlier for trochees than for iambs. Together with the amplitude patterns, trochees appear to imply stronger prosodic expectations, possibly caused by their preference in German (Wagner, 2012; Wiese and Speyer, 2015). The four points are further elucidated in the following sections.

Omission Responses to Syllables

The omission MMN has been identified as prediction error response to rare omissions in tone sequences whose inter-onset intervals would not exceed a specific temporal window of integration of about 125–150 ms (Yabe et al., 1997, 1998). Subsequent work has shown that the omission response can also be elicited by speech material (Bendixen et al., 2014) and in cases where the temporal window of integration is in fact exceeded (Scharinger et al., 2017). Here, we provide evidence

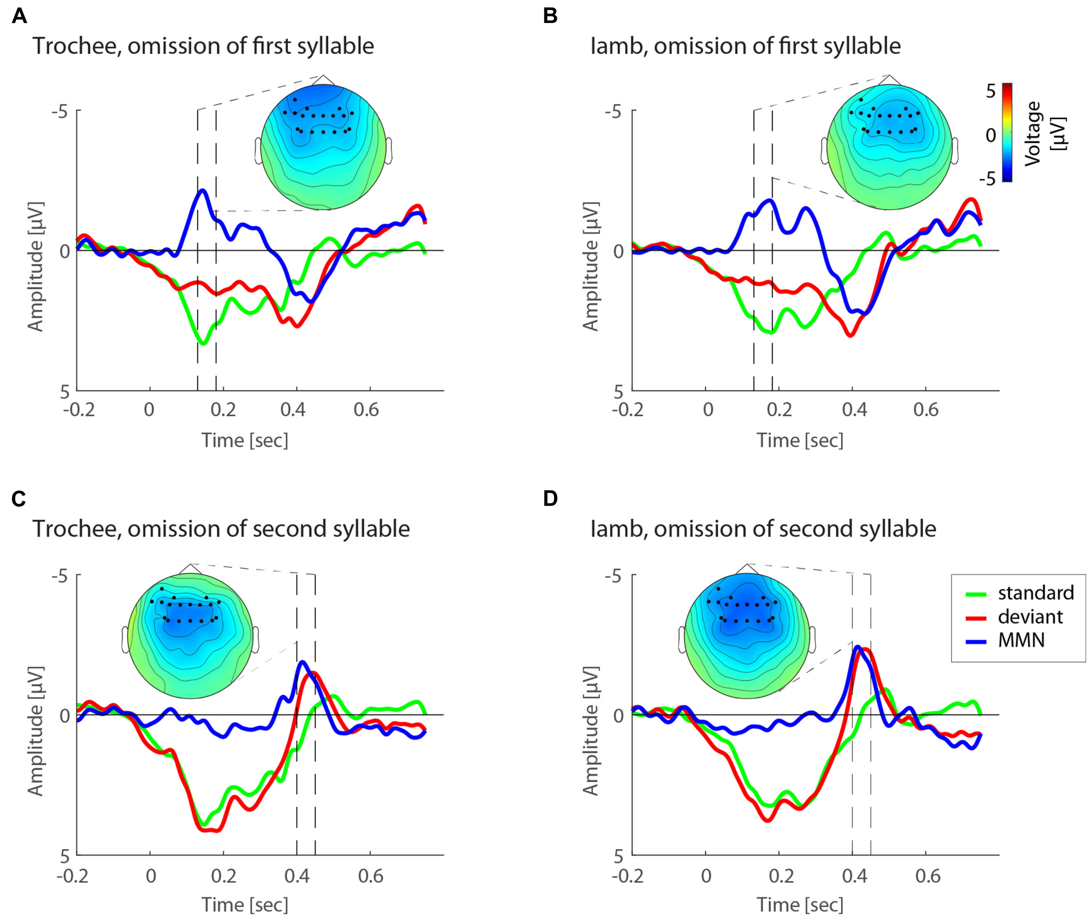


FIGURE 3 | Illustration of omission MMN effects. **(A)** Event-related potentials (ERP) to standard (green) and deviant (red) responses as well as difference wave forms (blue) for trochees in the first-syllable omission condition. **(B)** ERP responses for iambs in the first-syllable omission condition. **(C)** ERP responses for trochees in the second-syllable omission condition. **(D)** ERP responses for iambs in the second-syllable omission condition. Topographies highlight the electrodes and dashed lines the temporal windows established by the cluster statistics.

that omissions of syllables whose duration by far exceeded the 125–150 ms integration window can also elicit a robust omission MMN. This is the basis for our following interpretations, since we take the omission response to reflect a violation of a syllable-based prediction.

The difference between omissions in first- and second-syllable position in our experiment may—at first sight—be based on differences in temporal predictions. Second-syllable omissions are characterized by a violation of the word-internal timing. In all disyllabic pseudowords, the onset-to-onset interval of the two syllables is 250 ms. In addition to the prediction that the syllable in second position is a repeated version of the syllable in first position, there is also a strong temporal prediction that the onset of the second syllable is 250 ms after the onset of the first syllable. Temporal predictions in audition are particularly fostered by regular acoustic contexts, such as provided by oddball paradigms (Tavano et al., 2014; Auksztulewicz et al., 2018; Lumaca et al., 2019; Pinto et al., 2019). Tavano et al. (2014) and Auksztulewicz et al. (2018) explicitly refer to the need of temporal regularity for higher-order predictions, possibly supported by the brain's dynamic sensitivity to different processing frequencies (Arnal et al., 2014), related to motor-areas (Auksztulewicz et al., 2018) or subcortical, thalamo-cerebellar circuits (Schwartze et al., 2012). In our study, omissions in first and second position may differ on the basis of temporal predictability. While second-syllable omissions may rather be sensitive to word-internal temporal regularity, first-syllable omissions should be sensitive to between-word temporal regularity. However, since word-internal as well as word-external timing is constant throughout the experiment, the interpretation of the stronger effects in second position would be that a violation of within-word temporal regularity causes a stronger prediction error than a violation of between-word temporal regularity.

Foot Type Modulates Prosodic Predictions

A more plausible interpretation of the differences between first- and second position is based on the interaction of the effects

TABLE 1 | Summary of mixed-effects ANOVAs on mean amplitudes and peak amplitudes.

Factor	MeanSq	NumDF	DenDF	*F*-value	*P*	Sig
Mean amplitudes						
Position	38.57	1	1,349	8.34	0.004	**
Foot Type	5.48	1	1,349	1.19	0.276	n.s.
Electrode	3.41	17	1,349	0.74	0.766	n.s.
Position × Foot Type	72.16	1	1,349	15.61	0.000	***
Position × Electrode	1.76	17	1,349	0.38	0.989	n.s.
Foot Type × electrode	1.26	17	1,349	0.27	0.999	n.s.
Position × Foot Type × Electrode	1.96	17	1,349	0.42	0.981	n.s.
First-syllable omission						
Foot Type	18.93	1	665	10.90	0.001	**
Electrode	0.70	17	665	0.40	0.985	n.s.
Foot Type × Electrode	2.06	17	665	1.18	0.271	n.s.
Second-syllable omission						
Foot Type	58.71	1	665	9.61	0.002	**
Electrode	4.47	17	665	0.73	0.772	n.s.
Foot Type × Electrode	1.16	17	665	0.19	1.000	n.s.
Peak amplitudes						
Position	17.43	1	1,349	4.07	0.044	*
Foot Type	5.53	1	1,349	1.29	0.256	n.s.
Electrode	3.79	17	1,349	0.88	0.593	n.s.
Position × Foot Type	43.44	1	1,349	10.14	0.001	**
Position × Electrode	2.01	17	1,349	0.47	0.967	n.s.
Foot Type × Electrode	1.11	17	1,349	0.26	0.999	n.s.
Position × Foot Type × Electrode	2.32	17	1,349	0.54	0.933	n.s.
First-syllable omission						
Foot Type	8.98	1	665	5.71	0.017	*
Electrode	1.34	17	665	0.85	0.635	n.s.
Foot Type × Electrode	1.95	17	665	1.24	0.230	n.s.
Second-syllable omission						
Foot Type	39.99	1	665	7.03	0.008	**
Electrode	4.46	17	665	0.78	0.713	n.s.
Foot Type × Electrode	1.48	17	665	0.26	0.999	n.s.

When qualified by significant interaction, first- and second-syllable omission conditions are analyzed separately.
*Significance coding: ***p < 0.001; **p < 0.01; *p < 0.05.*
n.s., not significant.

TABLE 2 | Summary of mixed-effects ANOVAs on MMN latencies.

Factor	MeanSq	NumDF	DenDF	*F*-value	*P*	Sig
Position	26.21	1	1,349	129100.00	0.000	***
Foot Type	0.02	1	1,349	101.44	0.000	***
Electrode	0.00	17	1,349	0.63	0.874	n.s.
Position × Foot Type	0.02	1	1,349	79.13	0.000	***
Position × Electrode	0.00	17	1,349	1.12	0.331	n.s.
Foot Type × Electrode	0.00	17	1,349	0.11	1.000	n.s.
Position × Foot Type × Electrode	0.00	17	1,349	1.03	0.419	n.s.
First-syllable omission						
Foot Type	0.04	1	665	221.02	0.000	***
Electrode	0.00	17	665	1.00	0.452	n.s.
Foot Type × Electrode	0.00	17	665	0.53	0.939	n.s.
Second-syllable omission						
Foot Type	0.00	1	665	0.96	0.328	n.s.
Electrode	0.00	17	665	1.28	0.198	n.s.
Foot Type × Electrode	0.00	17	665	0.98	0.481	n.s.

When qualified by significant interaction, first- and second-syllable omission conditions are analyzed separately.
*Significance coding: ***p < 0.001.*
n.s., not significant.

of position and foot type. This interaction indicates that the position effect crucially depends on foot type: Only for iambs, the omission of the second syllable resulted in a larger omission MMN. That is, there is not a position effect *per se*, but rather a strong prediction of when strong syllables occur in either trochaic or iambic words. In iambic words, the strong syllable appears in second position, thus, the omission of the second syllable should result in a stronger prediction error, if the omission MMN is sensitive to prosodic properties such as syllable weight. This is supported by the results of our experiment, where indeed second-syllable omissions in iambs resulted in stronger MMNs than first-syllable position omissions. The same, complementary pattern, held for trochees: Here, omissions in first-position resulted in stronger omission MMNs than omissions in second-position. Put differently, the omission MMN is not only sensitive to syllable omissions and their temporal position, but also to the prosodic properties of these syllables, following the different foot types. This partially replicates the findings of Brochard et al. (2003) who demonstrated that the MMN depends on the subjective accenting of sound sequencing, with stronger MMNs in strong positions compared to weak positions. In our study, strong and weak positions are encoded in the acoustics of the experimental material. To this end, trochees consisted of initial syllables with a higher pitch than their final syllables, while iambs consisted of final syllables with a higher pitch than their initial syllables. In both cases, the syllables with higher pitch are likely to be interpreted as strong syllables, and the respective omission of the strong syllables resulted in a larger MMN than the omission of the corresponding weak syllables. Future studies may take this as a starting point when examining to what extent these prediction violations co-vary with higher-order, aesthetic processing. Existing studies strongly suggest an interactive effect of prosodic expectations and aesthetic appreciation (Obermeier et al., 2013, 2015; Menninghaus et al., 2015). The omission paradigm can offer a new way to quantify this correlation.

Trochaic Preferences

Finally, when looking at the MMN latencies, our patterns of results suggest that trochees take a specific role in that the omission of their (strong) first syllable results in an earlier MMN than the omission of the (weak) first syllable of iambs. Hence, the omission of a strong syllable in first position results in a particularly salient prediction violation. Of course, it is

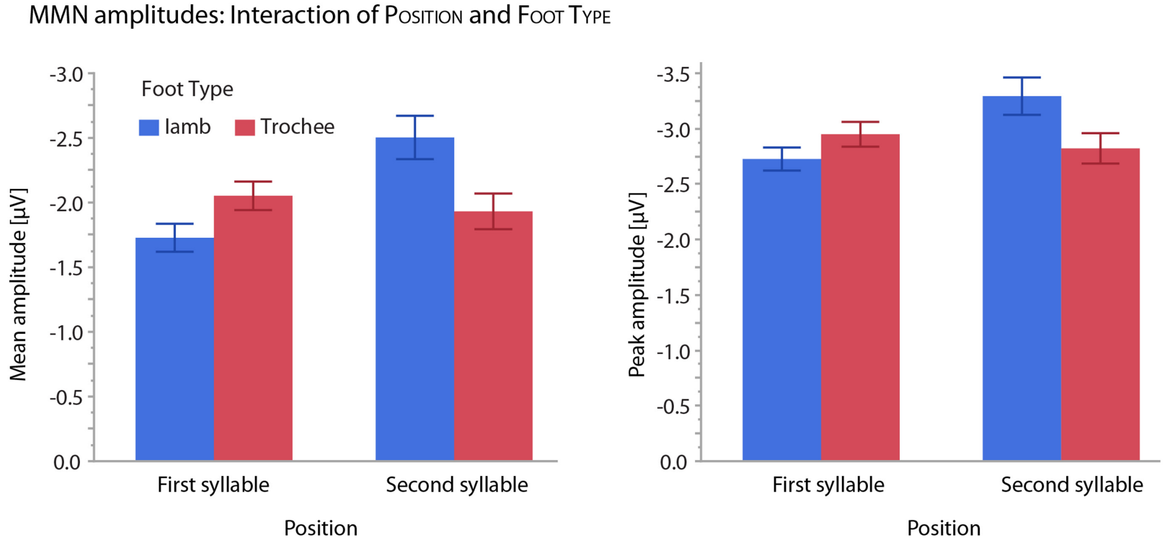


FIGURE 4 | Illustration of interaction patterns between the effects of Position and Foot Type. Left: mean amplitudes; right: peak amplitudes. Whiskers show standard errors of the mean.

impossible to base this effect on foot type because foot type and the position of strong syllable are confounded. To disentangle these effects, future work is necessary. However, in combination with the amplitude data, the conclusion seems warranted that trochees have a specific influence on the omission response in that this response is not only elicited at earlier latencies but also with a stronger amplitude when the strong syllable is omitted. Note that the omission of the strong syllable in iambs led to an even stronger MMN, indicating that at least the amplitude pattern does not depend on whether the syllable occurred word-initially or word-finally. Therefore, we conclude that the particular pattern elicited by trochees reflects their preferred status in German prosody (Wiese, 1996; Wagner, 2012; Wiese and Speyer, 2015). Furthermore, the latency effect in first-syllable position may also be driven by the Iambic-Trochaic Law (ITL) according to which foot beginnings are marked by higher pitch and/or higher syllable intensities, while foot endings are marked by longer syllable durations (Hay and Diehl, 2007; de la Mora et al., 2013; Crowhurst and Olivares, 2014; Crowhurst, 2020). Since we only modified pitch in our experiment, we cannot fully explore the ITL here, but suggest that earlier sensitivity to the omission of the higher-pitched syllable in trochees compared to the lower-pitched syllable in iambs is in accordance with this law. A likely articulatory explanation of this effect is that due to the respiratory cycle, word- and phrase initial syllables can be produced with higher intensities and higher pitch just because more air and more pressure is available after inhalation (see Tierney et al., 2011 for a similar explanation for song patterns in humans and non-humans).

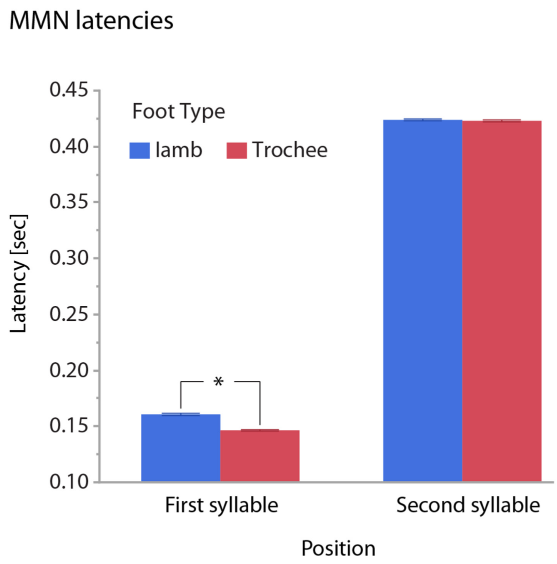


FIGURE 5 | Illustration of the interaction of the effects of Position and Foot Type on MMN latencies. Whiskers show standard errors of the mean. MMN latencies differ between trochees and iambs only in the first-syllable omission condition (significance marked by an asterisk, "*").

CONCLUSION

Audition benefits from local and global regularities, both temporally and phonologically (i.e., content-based). Regularities can generate strong predictions, whose violations lead to well-known electrophysiological responses. We here demonstrated the feasibility of the omission MMN to quantify foot type-based differences in prediction violations. This research can mark the starting point for further studies more concretely looking at the interplay of predictive processing and aesthetic evaluation.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of the Max-Planck-Society. The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

KH was involved in research question identification, study planning, protocol preparation, data analysis, data interpretation, and manuscript editing. MS was involved in research question identification, study planning, protocol preparation, data analysis, data interpretation, manuscript writing, editing, and reviewing. Both authors reviewed and approved the manuscript prior submission.

FUNDING

This study was supported by the Max-Planck-Gesellschaft (MPG).

ACKNOWLEDGMENTS

We would like to thank Winfried Menninghaus who gave us important feedback during the preparation of this study. We also would like to thank the members of the Language and Literature Department of the Max-Planck-Institute for Empirical Aesthetics, for their helpful comments and discussions. A particular thanks goes to Cornelius Abel and Miriam Riedinger for invaluable laboratory support.

REFERENCES

Arnal, L. H., Doelling, K. B., and Poeppel, D. (2014). Delta-beta coupled oscillations underlie temporal prediction accuracy. *Cereb. Cortex.* 25, 3077–3085. doi: 10.1093/cercor/bhu103

Auksztulewicz, R., Schwiedrzik, C. M., Thesen, T., Doyle, W., Devinsky, O., Nobre, A. C., et al. (2018). Not all predictions are equal: 'What' and 'When' predictions modulate activity in auditory cortex through different mechanisms. *J. Neurosci.* 38, 8680–8693. doi: 10.1523/JNEUROSCI.0369-18.2018

Baldeweg, T. (2006). Repetition effects to sounds: evidence for predictive coding in the auditory system. *Trends Cogn. Sci.* 10, 93–94. doi: 10.1016/j.tics.2006.01.010

Bendixen, A., Scharinger, M., Strauss, A., and Obleser, J. (2014). Prediction in the service of speech comprehension: modulated early brain responses to omitted speech segments. *Cortex* 53, 9–26. doi: 10.1016/j.cortex.2014.01.001

Bohn, K., Knaus, J., Wiese, R., and Domahs, U. (2013). The influence of rhythmic (ir)regularities on speech processing: evidence from an ERP study on German phrases. *Neuropsychologia* 51, 760–771. doi: 10.1016/j.neuropsychologia.2013.01.006

Bosch-Bayard, J., Valdos-Sosa, P., Virues-Alba, T., Aubert-Vazquez, E., John, E. R., Harmony, T., et al. (2001). 3D statistical parametric mapping of EEG source spectra by means of variable resolution electromagnetic tomography (VARETA). *Clin. Electroencephalogr.* 32, 47–61. doi: 10.1177/155005940103200203

Breen, M., Fitzroy, A. B., and Oraa Ali, M. (2019). Event-related potential evidence of implicit metric structure during silent reading. *Brain Sci.* 9:E192. doi: 10.3390/brainsci9080192

Brochard, R., Abecasis, D., Potter, D., Ragot, R., and Drake, C. (2003). The "Ticktock" of our internal clock: direct brain evidence of subjective accents in isochronous sequences. *Psychol. Sci.* 14, 362–366. doi: 10.1111/1467-9280.24441

Cason, N., and Schön, D. (2012). Rhythmic priming enhances the phonological processing of speech. *Neuropsychologia* 50, 2652–2658. doi: 10.1016/j.neuropsychologia.2012.07.018

Chwilla, D. J., Brown, C. M., and Hagoort, P. (1995). The N400 as a function of the level of processing. *Psychophysiology* 32, 274–285. doi: 10.1111/j.1469-8986.1995.tb02956.x

Colin, C., Hoonhorst, I., Markessis, E., Radeau, M., de Tourtchaninoff, M., Foucher, A., et al. (2009). Mismatch negativity (MMN) evoked by sound duration contrasts: an unexpected major effect of deviance direction on amplitudes. *Clin. Neurophysiol.* 120, 51–60. doi: 10.1016/j.clinph.2008.10.002

Crowhurst, M. J. (2020). The iambic/trochaic law: nature or nurture? *Lang. Linguist. Compass* 14, e12360. doi: 10.1111/lnc3.12360

Crowhurst, M. J., and Olivares, A. T. (2014). Beyond the Iambic-Trochaic Law: the joint influence of duration and intensity on the perception of rhythmic speech. *Phonology* 31, 51–94. doi: 10.1017/s0952675714000037

de la Mora, D. M., Nespor, M., and Toro, J. M. (2013). Do humans and nonhuman animals share the grouping principles of the Iambic - Trochaic Law? *Atten. Percept. Psychophys.* 75, 92–100. doi: 10.3758/s13414-012-0371-3

Dehaene-Lambertz, G., Dupoux, E., and Gout, A. (2000). Electrophysiological correlates of phonological processing: a cross-linguistic study. *J. Cogn. Neurosci.* 12, 635–647. doi: 10.1162/089892900562390

Domahs, U., Knaus, J. A., El Shanawany, H., and Wiese, R. (2014). The role of predictability and structure in word stress processing: an ERP study on Cairene Arabic and a cross-linguistic comparison. *Front. Psychol.* 5:1151. doi: 10.3389/fpsyg.2014.01151

Domahs, U., Wiese, R., Bornkessel-Schlesewsky, I., and Schlesewsky, M. (2008). The processing of German word stress: evidence for the prosodic hierarchy*. *Phonology* 25, 1–36. doi: 10.1017/S0952675708001383

Evans, A. C., Collins, D. L., Mills, S. R., Brown, E. D., Kelly, R. L., and Peters, T. M. (1993). "3D statistical neuroanatomical models from 305 MRI volumes," in *Proceedings of the Nuclear Science Symposium and Medical Imaging Conference, 1993 IEEE*, San Francisco, CA, 1813–1817. doi: 10.1109/NSSMIC.1993.373602

Franklin, M. S., Dien, J., Neely, J. H., Huber, E., and Waterson, L. D. (2007). Semantic priming modulates the N400, N300, and N400RP. *Clin. Neurophysiol.* 118, 1053–1068. doi: 10.1016/j.clinph.2007.01.012

Friston, K. (2003). Learning and inference in the brain. *Neural Netw.* 16, 1325–1352. doi: 10.1016/j.neunet.2003.06.005

Friston, K. (2005). A theory of cortical responses. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 815–815. doi: 10.1098/rstb.2005.1622

Friston, K. (2008). Hierarchical models in the brain. *PLoS Comput. Biol.* 4:e1000211. doi: 10.1371/journal.pcbi.1000211

Greenberg, S., Hitchcock, L., and Chang, S. (2003). Temporal properties of spontaneous speech - a syllable-centric perspective. *J. Phonet.* 31, 465–485. doi: 10.3389/fpsyg.2019.02019

Hay, J. S. F., and Diehl, R. L. (2007). Perception of rhythmic grouping: testing the iambic/trochaic law. *Percept. Psychophys.* 69, 113–122. doi: 10.3758/bf03194458

Hayes, B. (1995). *Metrical Stress Theory: Principles and Case Studies*. Chicago: The University of Chicago Press.

Henrich, K., Alter, K., Wiese, R., and Domahs, U. (2014). The relevance of rhythmical alternation in language processing: an ERP study on

English compounds. *Brain Lang.* 136, 19–30. doi: 10.1016/j.bandl.2014.07.003

Henry, M. J., and Obleser, J. (2012). Frequency modulation entrains slow neural oscillations and optimizes human listening behavior. *Proc. Natl. Acad. Sci. U.S.A.* 109, 20095–20100. doi: 10.1073/pnas.1213390109

Horváth, J., Müller, D., Weise, A., and Schröger, E. (2010). Omission mismatch negativity builds up late. *Neuroreport* 21, 537–541. doi: 10.1097/WNR.0b013e3283398094

Jakobson, R. (1960). "Linguistics and poetics," in *Style in Language*, ed. T. Sebeok (Cambridge: MIT Press), 350–377.

Jusczyk, P. W. (1999). How infants begin to extract words from speech. *Trends Cogn. Sci.* 3, 323–328. doi: 10.1016/s1364-6613(99)01363-7

Kiebel, S. J., Daunizeau, J., and Friston, K. J. (2009). Perception and hierarchical dynamics. *Front. Neuroinform.* 3:20. doi: 10.3389/neuro.11.020.2009

Kuperberg, G. R., and Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Lang. Cogn. Neurosci.* 31, 32–59. doi: 10.1080/23273798.2015.1102299

Kutas, M., DeLong, K. A., and Smith, N. J. (2011). "A look around what lies ahead: prediction and predictability in language processing," in *Predictions in the Brain*, ed. M. Bar (Oxford: Oxford University Press), 190–208. doi: 10.1093/acprof:oso/9780195395518.003.0065

Kutas, M., and Hillyard, S. A. (1980). Reading senseless sentences: brain potential reflect semantic incongruity. *Science* 207, 203–205. doi: 10.1126/science.7350657

Kutas, M., and Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature* 307, 161–163. doi: 10.1038/307161a0

Kuznetsova, A., Bruun Brockhoff, P., and Bojesen Christensen, R. (2014). *lmerTest: Tests for Random and Fixed Effects for Linear Mixed Effect Models (lmer Objects of lme4 Package). R Package Version 2.0-6.* Available online at: http://CRAN.R-project.org/package=lmerTest (accessed November 29, 2019).

Lau, E. F., Holcomb, P. J., and Kuperberg, G. R. (2013). Dissociating N400 effects of prediction from association in single-word contexts. *J. Cogn. Neurosci.* 25, 484–502. doi: 10.1162/jocn_a_00328

Lumaca, M., Trusbak Haumann, N., Brattico, E., Grube, M., and Vuust, P. (2019). Weighting of neural prediction error by rhythmic complexity: a predictive coding account using mismatch negativity. *Eur. J. Neurosci.* 49, 1597–1609. doi: 10.1111/ejn.14329

Magne, C., Astésano, C., Aramaki, M., Ystad, S., Kronland-Martinet, R., and Besson, M. (2007). Influence of syllabic lengthening on semantic processing in spoken French: behavioral and electrophysiological evidence. *Cereb. Cortex* 17, 2659–2668. doi: 10.1093/cercor/bhl174

Magne, C., Jordan, D. K., and Gordon, R. L. (2016). Speech rhythm sensitivity and musical aptitude: ERPs and individual differences. *Brain Lang.* 153–154, 13–19. doi: 10.1016/j.bandl.2016.01.001

Mathworks (2016). *MATLAB (Version 2016a)*. Natick, MA: Mathworks, Inc.

Menninghaus, W., Bohrn, I. C., Knoop, C., Kotz, S. A., Schlotz, W., and Jacobs, A. M. (2015). Rhetorical features facilitate prosodic processing while handicapping ease of semantic comprehension. *Cognition* 143, 48–60. doi: 10.1016/j.cognition.2015.05.026

Menninghaus, W., Wagner, V., Wassiliwizky, E., Jacobsen, T., and Knoop, C. A. (2017). The emotional and aesthetic powers of parallelistic diction. *Poetics* 63, 47–59. doi: 10.1016/j.poetic.2016.12.001

Molczanow, J., Domahs, U., Knaus, J., and Wiese, R. (2013). The lexical representation of word stress in Russian: evidence from event-related potentials. *Ment. Lex.* 8, 164–194. doi: 10.1075/ml.8.2.03mol

Molczanow, J., Iskra, E., Dragoy, I., Wiese, R., and Domahs, U. (2019). Default stress assignment in Russian: evidence from acquired surface dyslexia. *Phonology* 36, 61–90.

Näätänen, R. (1995). The mismatch negativity - a powerful tool for cognitive neuroscience. *Ear Hear.* 16, 6–18.

Näätänen, R., and Alho, K. (1997). Mismatch negativity (MMN) - the measure for central sound representation accuracy. *Audiol. Neurootol.* 2, 341–353.

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Ilvonen, A., et al. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature* 385, 432–434.

Nespor, M., and Vogel, I. (1986). *Prosodic Phonology*. Dordrecht: Foris.

Obermeier, C., Kotz, S. A., Jessen, S., Raettig, T., Koppenfels, M. V., and Menninghaus, W. (2015). Aesthetic appreciation of poetry correlates with ease of processing in event-related potentials. *Cogn. Affect. Behav. Neurosci.* 37, 1–7. doi: 10.3758/s13415-015-0396-x

Obermeier, C., Menninghaus, W., von Koppenfels, M., Raettig, T., Schmidt-Kassow, M., Otterbein, S., et al. (2013). Aesthetic and emotional effects of meter and rhyme in poetry. *Front. Psychol.* 4:10. doi: 10.3389/fpsyg.2013.00010

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113. doi: 10.1016/0028-3932(71)90067-4

Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J. M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intellig. Neurosci.* 2011:156869.

Pickering, M. J., and Garrod, S. (2013). An integrated theory of language production and comprehension. *Behav. Brain Sci.* 36, 329–347. doi: 10.1017/S0140525X12001495

Pinto, S., Tremblay, P., Basirat, A., and Sato, M. (2019). The impact of when, what and how predictions on auditory speech perception. *Exp. Brain Res.* 237, 3143–3153. doi: 10.1007/s00221-019-05661-5

Reetz, H., and Jongman, A. (2008). *Phonetics: Transcription, Productions, Acoustics and Perception*. Oxford: Wiley-Blackwell.

Roncaglia-Denissen, M. P., Schmidt-Kassow, M., and Kotz, S. A. (2013). Speech rhythm facilitates syntactic ambiguity resolution: ERP evidence. *PLoS One* 8:e56000. doi: 10.1371/journal.pone.0056000

Rothermich, K., and Kotz, S. A. (2013). Predictions in speech comprehension: fMRI evidence on the meter-semantic interface. *Neuroimage* 70, 89–100. doi: 10.1016/j.neuroimage.2012.12.013

Rothermich, K., Schmidt-Kassow, M., Schwartze, M., and Kotz, S. A. (2010). Event-related potential responses to metric violations: rules versus meaning. *Neuroreport* 21, 580–584. doi: 10.1097/WNR.0b013e32833a7da7

Salisbury, D. F. (2012). Finding the missing stimulus mismatch negativity (MMN): emitted MMN to violations of an auditory gestalt. *Psychophysiology* 49, 544–548.

Scharinger, M., Monahan, P. J., and Idsardi, W. J. (2016). Linguistic category structure influences early auditory processing: converging evidence from mismatch responses and cortical oscillations. *Neuroimage* 128, 293–301.

Scharinger, M., Steinberg, J., and Tavano, A. (2017). Integrating speech in time depends on temporal expectancies and attention. *Cortex* 93, 28–40.

Schmidt-Kassow, M., and Kotz, S. A. (2009). Event-related brain potentials suggest a late interaction of meter and syntax in the P600. *J. Cogn. Neurosci.* 21, 1693–1708. doi: 10.1162/jocn.2008.21153

Schröger, E. (2005). The mismatch negativity as a tool to study auditory processing. *Acta Acúst.* 91, 490–501.

Schröger, E., Marzecová, A., and SanMiguel, I. (2015). Attention and prediction in human audition: a lesson from cognitive psychophysiology. *Eur. J. Neurosci.* 41, 641–664. doi: 10.1111/ejn.12816

Schwartze, M., Tavano, A., Schröger, E., and Kotz, S. A. (2012). Temporal aspects of prediction in audition: cortical and subcortical neural mechanisms. *Int. J. Psychophysiol.* 83, 200–207. doi: 10.1016/j.ijpsycho.2011.11.003

Selkirk, E. (1995). "Sentence prosody: Intonation, stress, and phrasing," in *The Handbook of Phonological Theory*, ed. J. Goldsmith (Cambridge, MA: Blackwell Publishers), 550–569.

Strauß, A., Kotz, S. A., Scharinger, M., and Obleser, J. (2014). Alpha and theta brain oscillations index dissociable processes in spoken word recognition. *Neuroimage* 97, 387–395.

Takegata, R., Tervaniemi, M., Alku, P., Ylinen, S., and Näätänen, R. (2008). Parameter-specific modulation of the mismatch negativity to duration decrement and increment: evidence for asymmetric processes. *Clin. Neurophysiol.* 119, 1515–1523. doi: 10.1016/j.clinph.2008.03.025

Tavano, A., Widmann, A., Bendixen, A., Trujillo-Barreto, N., and Schröger, E. (2014). Temporal regularity facilitates higher-order sensory predictions in fast auditory sequences. *Eur. J. Neurosci.* 39, 308–318. doi: 10.1111/ejn.12404

Tervaniemi, M., Saarinen, J., Paavilainen, P., Danilova, N., and Näätänen, R. (1994). Temporal integration of auditory information in sensory memory as reflected by the mismatch negativity. *Biol. Psychol.* 38, 157–167.

Tierney, A. T., Russo, F. A., and Patel, A. D. (2011). The motor origins of human and avian song structure. *Proc. Natl. Acad. Sci. U.S.A.* 108, 15510–15515. doi: 10.1073/pnas.1103882108

Wagner, P. (2012). "Meter specific timing and prominence in German poetry and prose," in *Understanding Prosody, Language, Context and Cognition*, ed. O. Niebuhr (Berlin: Walter de Gruyter), 219–236.

Wiese, R. (1996). *The Phonology of German*. Oxford: Oxford University Press.

Wiese, R., and Speyer, A. (2015). Prosodic parallelism explaining morphophonological variation in German. *Linguistics* 53, 525–559. doi: 10.1515/ling-2015-0011

Winkler, I. (2007). Interpreting the mismatch negativity. *J. Psych. Phys.* 21, 147–163.

Yabe, H., Tervaniemi, M., Reinikainen, K., and Näätänen, R. (1997). Temporal window of integration revealed by MMN to sound omission. *Neuroreport* 8, 1971–1974. doi: 10.1097/00001756-199705260-00035

Yabe, H., Tervaniemi, M., Sinkkonen, J., Huotilainen, M., Ilmoniemi, R. J., and Näätänen, R. (1998). Temporal window of integration of auditory information in the human brain. *Psychophysiology* 35, 615–619. doi: 10.1017/S0048577298000183

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

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

frontiers in Communication

ORIGINAL RESEARCH
published: 09 February 2022
doi: 10.3389/fcomm.2021.637585

Developmental Language Disorder as Syntactic Prediction Impairment

Arild Hestvik[1], Baila Epstein[2], Richard G. Schwartz[3] and Valerie L. Shafer[3]*

[1]Department of Linguistics and Cognitive Science, University of Delaware, Newark, DE, United States, [2]Communication Arts, Sciences, and Disorders, Brooklyn College, Boylan Hall, Brooklyn, NY, United States, [3]PhD Program in Speech-Language-Hearing Sciences, The Graduate Center, City University of New York, New York, NY, United States

We provide evidence that children with Developmental Language Disorder (DLD) are impaired in predictive syntactic processing. In the current study, children listened passively to auditorily-presented sentences, where the critical condition included an unexpected "filled gap" in the direct object position of the relative clause verb. A filled gap is illustrated by the underlined phrase in "*The zebra that the hippo kissed the camel on the nose*...", rather than the expected "*the zebra that the hippo kissed [e] on the nose*", where [e] denotes the gap. Brain responses to the filled gap were compared to a control condition using adverb-relative clauses with identical substrings: "*The weekend that the hippo kissed the camel on the nose [e]*...". Here, the same noun phrase is not unexpected because the adverb gap occurs later in the structure. We hypothesized that a filled gap would elicit a prediction error brain signal in the form of an early anterior negativity, as we have previously observed in adults. We found an early (bilateral) anterior negativity to the filled gap in a control group of children with Typical Development (TD), but the children with DLD exhibited no brain response to the filled gap during the same early time window. This suggests that children with DLD fail to predict that a relativized object should correspond to an empty position after the relative clause verb, suggesting an impairment in predictive processing. We discuss how this lack of a prediction error signal can interact with language acquisition and result in DLD.

Keywords: syntax, gap-filling, prediction, event-related potentials, developmental language disorder, relative clauses, sentence processing

OPEN ACCESS

Edited by:
Phillip M. Alday, Max Planck Institute for Psycholinguistics, Netherlands

Reviewed by:
Sophie M. Hardy, University of Warwick, United Kingdom
Elisabeth Rabs, Saarland University, Germany

***Correspondence:**
Arild Hestvik
hestvik@udel.edu

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Communication

Received: *03 December 2020*
Accepted: *09 December 2021*
Published: *09 February 2022*

Citation:
Hestvik A, Epstein B, Schwartz RG and Shafer VL (2022) Developmental Language Disorder as Syntactic Prediction Impairment. Front. Commun. 6:637585. doi: 10.3389/fcomm.2021.637585

HIGHLIGHTS

- Typically developing children exhibit a very early brain response to prediction error during sentence processing
- Developmental Language Disorder children do not exhibit this brain response
- The finding suggest that Developmental Language Disorder involves a prediction impairment
- Results are interpreted within a model of development that relates language acquisition to parsing development

1. INTRODUCTION

1.1 Syntactic Displacement

Displacement is the perturbation of syntactic constituents in the service of various speech acts, such as asking a question, focusing on something, restricting the meaning of the referent, passivizing a verb, topicalizing a constituent, and so on. It is an indispensable grammatical mechanism in human language. In relative clauses, such as "The man that Bill saw yesterday", the relativized noun is related to a displaced direct object. During processing, a mechanism, called the parser, automatically generates a search for the origin of the displaced constituent, and generates predictions about where it will be found in the unfolding sentence structure (Crain and Fodor, 1985; but see Lewis and Vasishth, 2005; McElree, 2000 for alternative models). The current study examined whether children with Developmental Language Disorders (DLD) are impaired at predicting where the syntactic location of gaps should be, compared to their typically developing peers.

Several authors have observed that children with DLD are impaired in the use of Wh-questions (Deevy and Leonard, 2004; Marinis and van der Lely, 2007; Epstein et al., 2013) and relative clauses (Fonteneau and van der Lely, 2008; Friedmann and Novogrodsky, 2011; Hesketh, 2006; Hestvik et al., 2010; Schuele and Nicholls, 2000; Stavrakaki, 2001, 2002), and more generally with non-canonical word order (Montgomery and Evans, 2017). Different explanations for this have been offered in the literature, ranging from genetically caused impaired knowledge state (van der Lely and Pinker, 2014); impaired working memory resources (Weismer, 1996; Marton and Schwartz, 2003; Montgomery et al., 2017); slowed processing speed (Miller et al., 2001; Kail and Miller, 2006; Leonard et al., 2007), impaired sensory processing and speech perception leading ultimately to atypical morphosyntax and syntax (Leonard and Bortolini, 1998; Joanisse and Seidenberg, 2003), or impaired implicit learning (Evans et al., 2009; Plante et al., 2017). The aim of the current study is to investigate a previously unexplored possibility, namely that DLD has its root in prediction mechanisms (see Jones et al., 2021 for a recent discussion).

Prediction is increasingly recognized as a critical aspect of human cognition (Friston, 2005; Friston and Kiebel, 2009; Parr and Friston, 2018), and over the past decade, prediction has come to the forefront of psycholinguistic modeling and research (Levy, 2008; Rabagliati et al., 2016; Kuperberg and Jaeger, 2016; Gambi et al., 2018; Pickering and Gambi, 2018). Processing of filler-gap dependencies (the key component of relative clauses and Wh-questions) has long been known to involve predictions that arise from "active filler strategies" (Frazier and Fodor, 1978; Stowe, 1986; Stowe et al., 1991). We assume a model of filler-gap processing that includes the following assumptions: 1) An expression is recognized as a filler and is placed and maintained in working memory; 2) an "active" search for a gap position is initiated while the sentence representation is being incrementally built over time; 3) once a potential gap position is found, the filler is retrieved from memory and interpreted in this position—this is the step of "filling the gap" (Wagers and Phillips, 2013) or "antecedent reactivation" (Swinney et al., 1989; Love and Swinney, 1996). The active search stage involves predictions about how the sentence is likely to unfold; the parser predicts that it will encounter a position which can be interpreted as a gap in the sentence structure (Lau et al., 2006). This prediction in turn speeds up processing because predictions allow structure (and even lexical items) to be prebuilt before being encountered in the input stream. Pre-activation leads to faster integration of upcoming linguistic material (see Nieuwland and Kazanina 2020) for a recent review).

1.2 The Current Study: ERP Measure of Filler-Gap Processing in Developmental Language Disorders

Early work on gap-filling in typical populations focused on demonstrating that a filler is dynamically reactivated at the gap position, by using behavioral measures that tested for priming by the filler at the temporal juncture of the gap (Love and Swinney, 1996; Nicol et al., 2006). In Hestvik et al. (2010), we used a behavioral priming task with children with DLD, and found that they did not exhibit priming by the filler at the corresponding gap position, in contrast to a typical developing control group (see also Marinis and van der Lely 2007)). Using cross-modal priming, the control group of TD children exhibited priming at the gap position of stimuli related to the filler, but DLD children did not:

(1) The zebra$_{FILLER}$ that the hippo on the hill had kissed $[e]_{GAP}$ on the nose ran far away

One possible explanation for lack of priming is that children with DLD, due to reduced working memory capacity (Weismer and Thordardottir, 1996; Fiebach et al., 2001; Marton and Schwartz, 2003; Montgomery, 2003; Ouchikh et al., 2016), are unable to maintain the filler in working memory (Sprouse et al., 2012; Kim et al., 2020), and are therefore slower at reactivating the filler at the gap position. If children with DLD are slower at reactivating the filler, perhaps at a delay after the verb, priming should be observed further downstream from the earliest possible gap-position. However, we have no model that predicts how much further downstream a gap might be postulated, making priming experiments impractical, as a 2x2 design is required at every hypothetical reactivation position. In addition, cross-modal priming tasks are cognitively demanding (e.g., using dual-task paradigms). Children with DLD may perform poorly on these behavioral tasks due to weaknesses in skills other than grammar, such as poor reading skills and or poor working memory capacity.

The goal of the current study was therefore to instead use a continuous measure of gap filling, via a study of predictive processing. ERPs exhibit millisecond timing of neural processes time-locked to a stimulus of interest and can provide an indication of the timing of a "surprise" response if a gap prediction is violated. The ERP technique is well-suited to test sentence processing in children and in language impaired populations. ERPs can be recorded to auditory sentences (thus, not requiring reading skills), and can use a relatively simple task (simple listening for comprehension). Despite the advantages of ERPs, only one study to date has used these measures to test gap-filling in children with DLD. Fonteneau and van der Lely (2008) presented sentences like (2) and time-locked the ERP to the underlined nouns:

(2) a. Who did Barbie push the clown into the wall?
b. Who did Barbie push the ball into?

Their TD control group exhibited increased negativity over left anterior sites, within 300 ms of the onset of the "filled-gap"

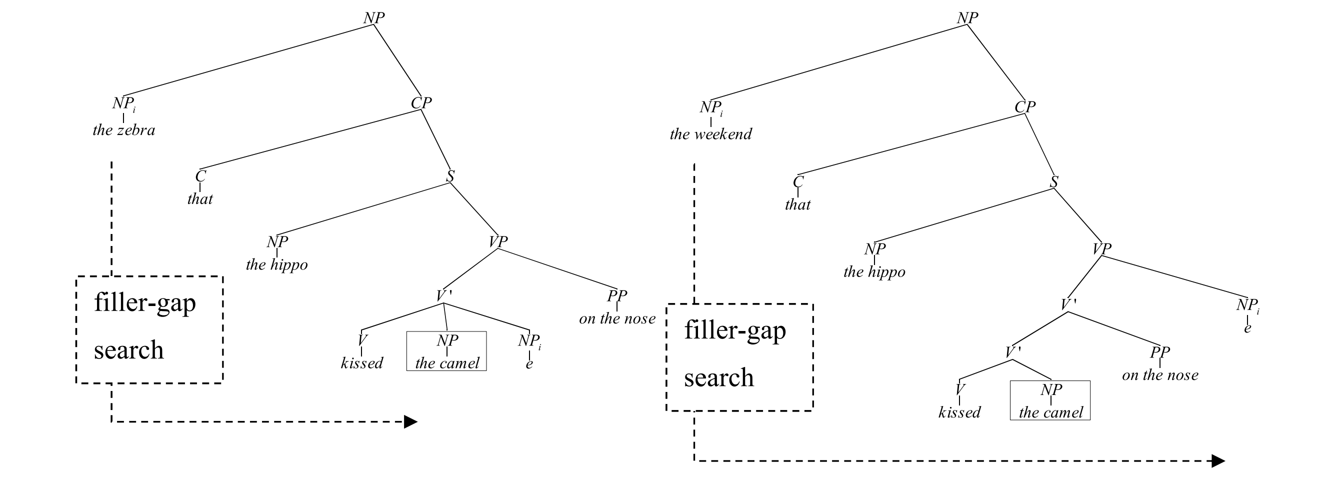


FIGURE 1 | Parse trees for experimental (filled gap) and control condition stimuli sentences.

underlined noun phrase in (2a), compared to the direct object in (2b). In contrast, the children with DLD showed a later negativity that the authors interpreted as an N400 effect reflecting that the noun was processed as being semantically anomalous or unexpected, rather than ungrammatical. However, the study had several limitations. The wide age range of the participants (10–21 years of age) makes interpretation of the results difficult because considerable developmental differences in the timing and polarity of ERPs to syntactic violations have been observed (Hahne et al., 2004). Also, the study did not control for matching noun phrases in test and control conditions, and therefore the early ERP difference could reflect processing of different lexical items rather than detection of an unexpected grammatical form (Steinhauer and Drury, 2012).

In the current study, we used a "filled gap" paradigm that controls for lexical factors to measure the effect of prediction violations during sentence comprehension. We contrasted test sentences like (3a) with control sentences like (3b) (these materials were also used in studies with adults in Hestvik et al., 2007; Hestvik et al., 2012)):

(3) a. The zebra that the hippo kissed the camel on the nose ran far away.
b. The weekend that the hippo kissed the camel on the nose, he ran far away.

The only difference between sentence (3a) and (3b) is in the probability of encountering "the camel" immediately after the verb. The relativized noun phrase in (3a) is a direct object argument of the verb, which makes a post-verbal NP highly unexpected (and the sentence is ultimately ungrammatical). The control sentence (3b) is perfectly grammatical, as a time adverb has been relativized. The relativized adverb also leads to a search for its gap. However, the gap is located at the right periphery of the verb phrase, as illustrated in **Figure 1** below; therefore, the occurrence of a noun phrase immediately following the verb is highly probable and not unexpected. Note that the two substrings and structures are otherwise identical. Thus, the only difference is in the grammatical function of the relativized noun, which predicts a direct object gap in (3a) but late adverb gap in (3b).

The experimental logic is illustrated in **Figure 1**. In both cases, we measured the brain response time-locked to the boxed NP "the camel": If a surprise response is generated by 'the camel' in (3a) but not in (3b), the only source of this response is that a gap is predicted in place of the NP in (3a) and not in (3b).

We predicted that the surprise should be reflected by an early Left Anterior Negativity (Hahne and Friederici, 1999). This prediction was based on previous studies with adults, where filled gaps was found to elicit early left anterior negativity (~200 ms), LAN (400–500 ms), and P600 (Felser and Jessen, 2020; Hestvik et al., 2012, 2007). Our first study with adults using the current paradigm revealed an early left anterior negativity to the filled gap (Hestvik et al., 2007). In Hestvik et al. (2012) we observed an early bilateral anterior negativity (EAN) in the same paradigm. Bilateral anterior negativities to syntactic violations have been observed in other studies (Kessler et al., 2004; Pakulak and Neville, 2011). We view the eLAN and EAN as belonging to a family of syntactic violation ERP responses.

We also assume that the eLAN/EAN does not directly reflect ungrammaticality (Friederici (2012), but rather reflects probabilistic processing. This is because the filled-gap NP in **Figure 1** does not make the sentence ungrammatical at the time point of its occurrence. The sentence could have a grammatical continuation, as in "The zebra that the hippo kissed the camel for." Thus, the EAN here reflects a low probability syntactic category "event" rather than ungrammaticality. The eLAN/EAN has been observed in grammatical expectation violation studies with a wide range of languages (Neville et al., 1991; Münte et al., 1993; Rösler et al., 1993; Knosche et al., 1999; Hinojosa et al., 2003; Kubota et al., 2003, 2018; Ye et al., 2006; Brunelliere et al., 2007; Isel et al., 2007). Our design compares identical word strings and identical syntactic structures in test and control conditions leading up to the critical word, and therefore meets the design requirements for appropriate controls that previous studies have been criticized for (Steinhauer and Drury, 2012).

For children with DLD, we predicted an absent or delayed brain response to the filled gap. A delayed anterior negativity would be consistent with the hypothesis that children with DLD experience a "generalized slowing" (Miller et al., 2001; Montgomery, 2004; Leonard et al., 2007) but are otherwise unimpaired. An absent EAN to the filled gap would be consistent with a lack of predictive processing of filler-gap constructions, which could be the result of poor working memory capacity (Epstein et al., 2013) (but see Discussion below), or a lack of grammatical knowledge of filler-gap relations (van der Lely, 2005; van der Lely and Pinker, 2014).

2 METHODS

2.1 Participants

Thirty children (8–13 years) were recruited and enrolled in the study, which took place in Manhattan, New York. In accordance with the Helsinki Declaration, the study was approved by the Graduate Center CUNY Internal Review Board. All children provided informed assent, and their caretakers provided informed consent. Fourteen of the children met the criteria for DLD. Seventeen age-matched typically developing (TD) children served as the control group. One child with DLD was later diagnosed with ADHD and excluded from the analysis. Among the remaining 13 children with DLD, 5 were female and 8 male (matching the prevalence of higher incidence of DLD for boys than girls); and among the 17 children with TD, 7 were female and 10 were male. We used age-matching of the control group, because language-matching would have introduced age-related confounding effects (Plante et al., 1993).

Left-handers were not excluded from the study (2 participants), as about 70% of left-handers still have left-lateralized language, and language lateralization is not predictable from handedness (Knecht et al., 2000; Corballis, 2014; Somers et al., 2015). There is also little evidence that DLD is related to handedness (Bishop, 2013). In addition, a recent study found that left-handers did not differ from right-handers in the P600 index of morpho-syntactic violations (Grey et al., 2017).

The study was representative of the ethnic and racial diversity of New York City: 37% of all participants were Hispanic or Latino (55% in the DLD group); 40% of all participants where Black or African American (45% in the DLD group); one child with DLD was Asian. 41% of the TD group was Black or African American and the remainder were White. All children reported English as their first language, and all were from households where English was the primary language.

The children in the study passed a pure-tone hearing screening at 20 dB HL, based upon the guidelines of the American Speech-Language-Hearing Association (1997). The children in the DLD group were all receiving speech pathology services in school at the time of the study. None of the children in the study had any history of frank neurological impairments, psychological or emotional disorders, attention deficit disorders or other neuro-developmental disorders (as reported by parent questionnaires). The children in both groups (except one child in the TD group) were tested on a battery of tests: The Clinical Evaluation of Language Fundamentals (CELF-4, Semel et al., 2004), the Test of Nonverbal Intelligence (TONI-3, Brown et al., 1997) and the Peabody Picture Vocabulary Test (PPVT, Dunn and Dunn, 2007). Children with DLD scored at least 1.25 standard deviations below the mean on at least two of the four core subtests of the CELF-4. **Table 1** provides means, standard deviations (SD) and ranges for these test scores and ages for each group. The mean expressive score on the CELF for the children in the DLD group was below 1.5 standard deviation of the population mean, but the mean PPVT score was within normal limits. Children in the TD group all scored within 1 SD of the mean on the CELF-4 and PPVT (see **Table 1**). Both groups of children scored within normal limits on the TONI-3.

As the descriptive statistics in **Table 1** show, the groups are matched on age and age variance, as well as on the TONI, meeting the standard description of DLD as being within normal range on non-verbal IQ. The DLD participants differed from the reference population with effect sizes between 1.5 and 2 standard deviations for each language-specific test: The DLD means on the CELF-R, CELF-E and PPVT were 1.5. SD, 2.0 SD and 1.5 SD below the population means, respectively.

2.2 Materials

The within-subject independent variable contained two levels: Filled gap vs. control. In addition, three other sentence types were used in the experiment to reduce predictability of stimuli and to prevent the children from engaging in strategies to predict filled gaps. Sixty-four stimuli were constructed for each the five sentence types, illustrated in **Table 2** (see the **Supplementary Appendix** for the full stimulus set).

2.2.1 Comprehension Questions

A set of comprehension questions was constructed for each of the 64 stimulus sentences in the Filled Gap, Adjunct Control, Declarative and Object Relative sentence types. The comprehension questions served multiple purposes. The primary purpose was to ensure that participants paid attention to and computed the meaning of the stimulus sentences. A secondary purpose was to measure whether DLD children exhibited Sustained Negativity between the filler and the gap in object Wh-questions compared to subject Wh-questions; these results are reported in Epstein et al. (2013).

Finally, the comprehension questions were used to measure whether the DLD children differed from TD children in their understanding of the stimuli. There were four question types: Object Wh-questions ("Who did the alligator tap?"), subject Wh-questions ("Who bumped the duck?"), Yes-No questions ("Did the hippo kiss the camel?") and a set of "easy" Yes-No questions ("Did you hear the word "road"?"). Question type was counterbalanced with the experimental condition type of the stimulus sentences (resulting in every question being asked four times over the entire experiment, but to different stimulus sentences). Thus, each subject heard 16 questions of each of the 4 question types, multiplied with 4 cells for a total of 256

TABLE 1 | Participant profiles with standard scores.

Group	Measure	Age	CELF-R	CELF-E	PPVT (-3 or -4)	TONI
DLD (N = 13)	Mean	10;1	79.92	76.38	85.54	98.23
	SD (months)	15 months	13.71	11.67	9.44	15.83
	Range	8;6–12;5	51–102	49–95	70–101	80–135
TD (N = 16*)	Mean	10;4	108.56	105.50	104.63	107.38
	SD (months)	12 months	12.13	13.42	12.15	12.12
	Range	8;5–12;3	88–125	89–133	86–129	90–130

*(*One TD participant did not take the CELF and PPVT tests; but was judged to have normal language development by a licensed speech language pathologist. For this reason, we report N = 16 in this table, but the ERP data are based on N = 17.)*

TABLE 2 | Sentence types.

Type	Label	Example
Test	Filled gap	The zebra that the hippo kissed the camel on the nose ran far away
Control	Adjunct	The weekend that the hippo kissed the camel on the nose he ran far away
Fillers	Object Relative	The zebra that the hippo kissed on the nose ran far away
Fillers	Declarative	The zebra said that the hippo kissed the camel on the nose and ran far away
Fillers	Temporal	The cockatoo squawked at the peacock before cleaning its feathers

questions. If DLD children failed to process sentences with filler-gap dependencies, they would be expected to exhibit guessing behavior for Adjunct Relatives and Object Relatives and should do worse on object Wh-questions than subject Wh-questions and Yes/No-questions, which do not involve long-distance dependencies.

To avoid asking comprehension questions after ungrammatical filled-gap sentences, each question was matched with two picture response options. One picture represented an object or character mentioned in the sentence. The other picture represented a question mark. Subjects were instructed to select the depicted object if it represented the answer, or alternatively the question mark if the depicted object did *not* represent the answer. Half the trials presented a picture depicting the correct answer, and the other half required choosing the *question mark* symbol. For the filled gap sentences, participants were expected to select the *question mark* response. This avoided asking a comprehension question to ungrammatical filled gap sentences.

Answers to comprehension questions were recorded by button press response and stored for analysis of accuracy and reaction time. An additional set of 38 "easy" filler sentences with heterogeneous structure (e.g., "The duckling and the chick that played near the barn ate all the seeds") were followed not by question but instead exclamations like "Is that so?", "You don't say", "Wow, ok", "I like that", and "That's really nice," so that not every sentence required a comprehension question (see the **Supplementary Appendix** for the full stimulus set).

2.2.2 Audio-Recording of Stimuli

The stimulus sentences and questions were digitally recorded by a female speaker (16-bit resolution and 22050 Hz sampling rate), and the comprehension questions were recorded by a different female speaker. The speaker of the test sentences was a trained linguist, who consciously avoided giving prosodic cues about the presence of a gap. Two recordings were made of each pair and the best was selected for use. Acoustic analyses of the critical stimuli were conducted to determine whether they contained unintended duration or pitch cues to upcoming gap positions (Nagel et al., 1994). The mean and standard deviations of the durations from verb onset to the determiner "the" following the verb were virtually identical (M = 413 ms, SD = 71 for adjunct control vs. M = 414 ms, SD = 72 for filled gap), thus containing no prosodic cue to a gap. In addition, the pitch contours of the filled gap and control sentences were determined to be virtually identical, by visual inspection.

2.3 Procedure

Participants were fitted with a 64 channel Electrical Geodesics Sensor Net (v2) containing silver/silver-chloride (Ag/AgCl) plated electrodes encased in electrolyte-wetted sponges. One electrode was placed under each eye to monitor eye movements and eye blinks (see the **Supplementary Appendix** for the full spatial layout of the electrode montage).

Participants were seated in a sound- and electrically shielded audiometric booth (International Acoustics Co.) that was dimly lit. Participants faced a computer screen positioned at eye level at a 70 cm distance. The stimulus presentation was controlled by a PC with Psychology Software Tools (PST) E-Prime software (Schneider et al., 2002), and behavioral responses were collected with a PST Serial Response Box. The sentences and questions were presented at 65 dB SPL with two free-field loudspeakers, one placed behind and one directly in front of the subject. Participants were instructed to position the index and fourth finger of their right hand on the response box with labeled buttons. A single sentence trial proceeded as follows: First, a picture of an eye, serving as a fixation point and a reminder not to blink, appeared in the center of the computer screen for 100 ms.

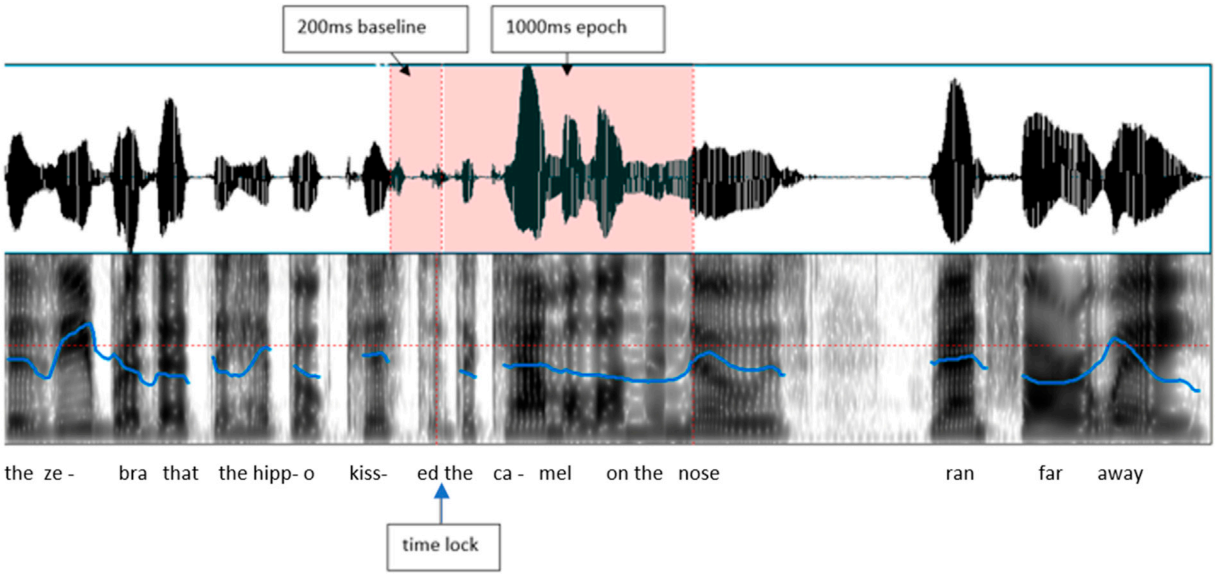


FIGURE 2 | A stimulus sentence example, indicating where the EEG was measured from (between offset of verb and onset of the article "the", and the baseline period.

This was followed by auditory presentation of the stimulus sentence, with the fixation picture remaining on the screen during the presentation. After a 1,000-ms pause, participants heard the comprehension question. Two response options were depicted on the screen for a maximum of 7,000 ms. One button represented each depicted response option. Accuracy feedback was provided after each question, as well as the cumulative accuracy, to encourage participants to take the questions seriously and give them motivation to track and monitor their own performance. A 1,000-ms pause followed before the next trial.

Each participant began with a set of practice trials followed by all the stimuli in two consecutive sessions. Each session was divided into four blocks of 32 trials, randomly drawn from each of the sentence types. Short breaks were given between each block, and a longer break between the two sessions. Participants were told to listen to the sentences for meaning and answer the comprehension questions. The entire recording session took between 1½ and 2 h.

2.4 EEG Recording, Artifact Correction and Principal Component Analysis/Independent Component Analysis Preprocessing

EEG was recorded with an Electrical Geodesics, Inc. NetAmps 200 system. Electrode impedances were below 60 kOhm, acceptable for high impedance amplifiers (Ferree et al., 2001). EEG was sampled at 200 Hz, with Cz as the reference, a 0.1–41.2 Hz bandpass filter, and digitized with 12-bit resolution. Stimulus onset markers were placed by E-Prime between the offset of the verb and the onset of "the" (example: "...the hippo kissed [MARK] the camel..."). The continuous EEG was segmented into 1,200 ms epochs, including a 200 ms pre-stimulus baseline and a 1,000 ms epoch duration, using EGI Netstation Waveform Tools, as illustrated in **Figure 2**.

The epoched data were then submitted to a semi-automatic artifact detection procedure using Netstation software. A channel in a single recording was marked as a bad channel if the fast average amplitude exceeded 200 μV; if the differential amplitude exceeded 100 μV; or if it had zero variance. A channel was considered a bad channel in all trials if it was a bad channel on 20 percent of the trials. A trial was excluded if it contained more than 10 bad channels, or if it contained lateral eye movements resulting in amplitudes greater than ±70 μV. Bad channels were deleted and replaced with data from the surrounding electrodes using spherical spline interpolation, as long as those channels contained good data. All trials with eye-blink activity were removed. We chose this procedure as an alternative to subtracting eyeblink activity via independent component analysis (ICA) decomposition, as our experience is that ICA eyeblink subtraction distorts the anterior negativity ERP. This agrees with Luck (2014, p. 215) who cautions against use of ICA when the ERP overlaps with blink topography, which was the case in the current study. Trials were then baseline corrected by subtracting the mean voltage of the 200 ms baseline pre-stimulus period from the entire segment; trials were finally averaged across conditions for each subject. The data were then re-referenced to the average voltage (Luu and Ferree, 2005).

2.5 Behavioral Data Analysis Plan

The proportion of correct answers after the four sentence types (except the "easy" filler stimuli which had no questions) were analyzed with a mixed factorial repeated measures ANOVA with four levels of Stimulus Type: Adjunct relative clause (control), object relative clause with a filled gap (test), and as fillers,

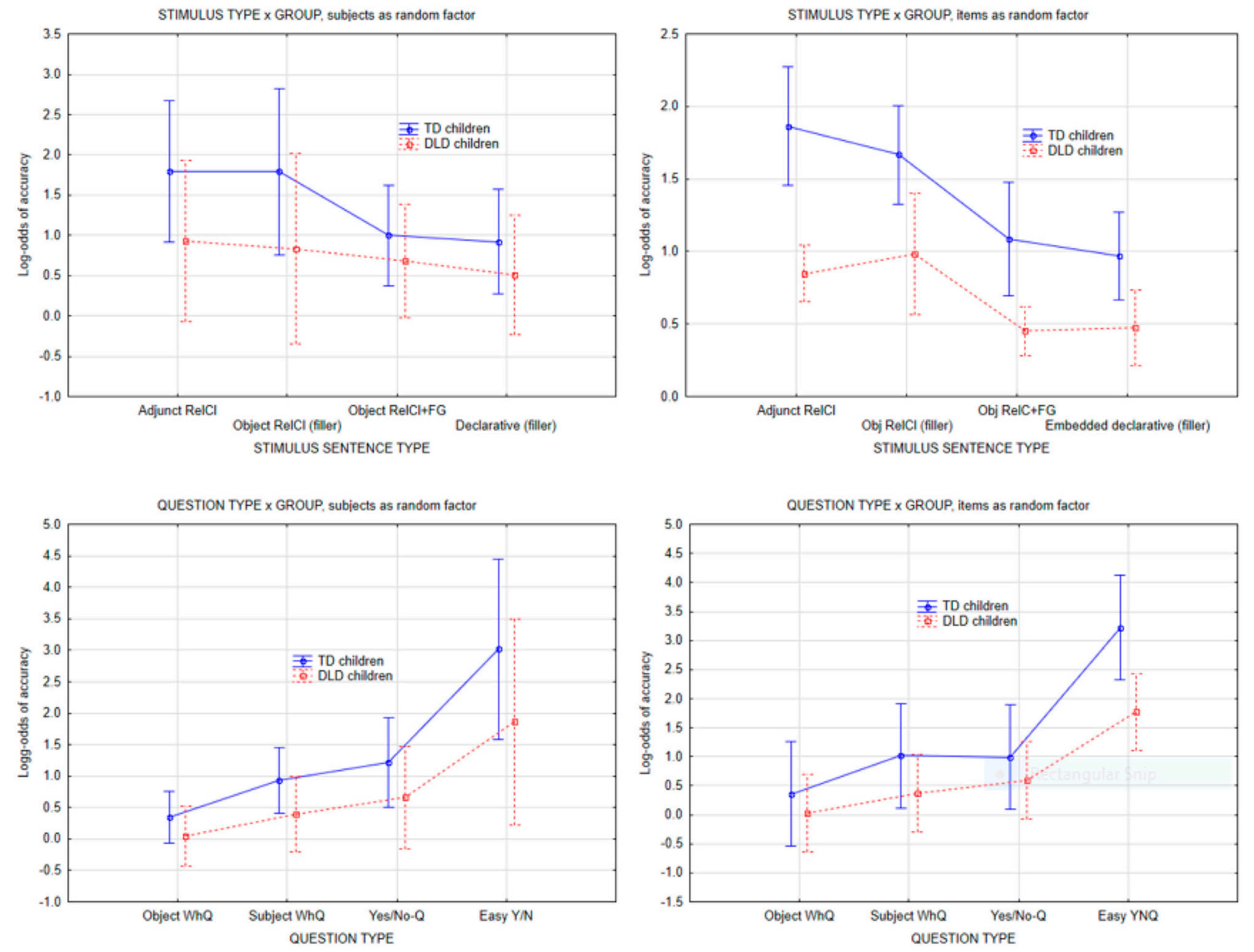


FIGURE 3 | Accuracy on comprehension questions by question type and stimulus sentence type, subject as random factor vs. item as random factor. Error bars indicate 95% confidence intervals.

declarative, object relative clause (filler), and a 2-clause embedded declarative (filler). Crossed with this was the four levels of Question Type: object Wh-question, subject Wh-questions, yes/no questions and "easy" yes/no questions, with Group as the between-subjects variable. Before analysis, the mean proportion of correct answers to each cell of the question type (4) x stimulus sentence type (4) was transformed to logits (the natural logarithm of the odds of a proportion) to approximate the ANOVA requirement of continuous and normally distributed variables. We conducted separate ANOVAs with subject versus item as random factor (Clark, 1973).

2.6 EEG/ERP Analysis Strategy

In order to determine which time windows and electrode regions to analyze in the EEG data, we used Principal Component analysis (PCA) (Gorsuch, 1983; Spencer et al., 2001; Dien and Frishkoff, 2005; Dien, 2010, 2012) and ICA (Delorme and Makeig, 2004; Jung et al., 2001; Makeig et al., 1997, 2002). This approach reduces experimenter bias related to selecting electrode channels and time windows (Luck and Gaspelin, 2017) and reduces multiple comparison problems, as it delivers data-driven constructs of time-windows and electrode regions. This method allows a more objective means of identifying regions and time windows of interest than subjective visual inspection of the 65 sites and 250 time points per site.

A sequential PCA/ICA procedure (Dien and Frishkoff, 2005; Dien, 2010, 2012) was applied to extract the temporal and spatial dynamics of the EEG response to the experimental conditions, using the ERP PCA toolkit in MatLab (Dien, 2010). The PCA/ICA solution was then used to guide and constrain the selection of time windows and electrode regions for constructing dependent measures for ANOVA. We did plan to analyze an early time window over anterior sites, based on our previous studies, but the PCA/ICA analysis allowed for an objective method in calculating this temporal-spatial component. As a first step in the analysis, the mean difference waves (filled gap minus control) per subject served as input to a temporal PCA using the covariance matrix and *promax* rotation (k = 3) with Kaiser loading weighting (Hendrickson and White, 1964; Richman, 1986; Tataryn et al., 1999). Following this step, temporal components were retained that accounted for at least 5% of the variance and fulfilled the Parallel Test and Scree Test (Horn, 1965). A spatial PCA was then conducted on each retained temporal factor using the INFOMAX rotation on the covariance matrix i.e., ICA (Bell and Sejnowski, 1995). Spatial factors for each of these temporal factors that accounted for at least 1% of the variance were then examined to determine which components best matched the temporal-spatial pattern of the AN. Note that the amount of variance accounted for by spatial factors is not relevant in determining the importance of a factor, because more focal effects will necessarily account for less variance than a more broadly distributed effect (which will be spread across more electrode sites). We refer the reader to tutorials for further explanation of the PCA approach (Dien, 2010, Dien, 2012, Dien, 2020).

The factors identified in the PCA/ICA that matched the temporal-spatial properties of AN (early in time, with anterior negativity) and their associated factor scores were assessed for significance by being used as dependent measures in mixed factorial repeated measures ANOVA, with group as a between-subjects variable (conducted separately for each of the five factors). Since the PCA/ICA factors were derived from difference waves, a significant intercept is analogous to a main effect of condition; and a main effect of group is analogous to an interaction between group and condition. The undecomposed, unweighted voltage data was then analyzed by using the temporo-spatial PCA/ICA region to select a voltage for each subject, condition and trial and analyzed with inferential statistics. Here, we used a linear mixed model, accounting for both subject and item variance. The analyses were carried out using Statistica (Statistica, 2017) and *lme4* (Bates et al., 2015) R (R Core Team, 2017) software.

3 RESULTS

3.1 Behavioral Comprehension Data

In the subject-as-random factor analysis, the independent variables were question type (4 levels), and stimulus sentence type (4 levels), with question type crossed with stimulus sentence type. There were 16 unique questions in each type: object Wh-questions, subject Wh-questions, yes/no-questions and "easy" yes/no-questions. Each question was posed once in each of the four stimulus conditions, resulting in a 4 x 4 within-subject design. This ANOVA resulted in main effects of group, question type, and stimulus sentence type but no interactions involving group. The main effect of group was caused by TD children having overall higher accuracy than DLD children (71 vs. 62%), $F(1,28) = 7.59$, $p = 0.01$). The main effect of question type ($F(3,84) = 111$, $p < 0.0001$) was due to Object Wh-questions being the hardest (54% accuracy), followed by Subject Questions (63%) and Yes/No-questions (67%), with the "easy" Yes/No-questions ("did hear the word X") having the highest accuracy (85%). There was no interaction between Question type and Group. There was also a main effect of stimulus sentence type, $F(3,84) = 8.8$, $p < 0.001$, such that the Adjunct control and relative clause filler had higher accuracy than the filled gap stimuli and the declarative clause fillers. Again, there was no interaction between Group and stimulus sentence type.

For the analysis with item (comprehension question) as random factor, the same 16 questions in each question type are now viewed as random samples of the infinite number of questions that could be formed within each type. The question type therefore becomes a grouping variable for questions (i.e., Wh-question, Yes/No-questions) and is in effect a "between-item" or grouping variable, with questions as the randomly sampled items that are being tested. Finally, participant group was added as a "within-item" variable for questions, because each question is tested repeatedly in both TD children and DLD children. The by-item analysis converged with the subject-as-random factor analysis in showing a main effect of group, and a main effect of question type and stimulus type. It differed from subject as random factor by exhibiting an interaction between question type and group. Inspection of the interaction plots revealed that this was driven by the "easy" Yes/No-questions ("did you hear the word "zebra"?") having higher accuracy in the TD group.

As **Figure 3** shows, both groups of children exhibited a similar pattern of accuracy. There was a main effect of group such that TD children had higher accuracy, but there was no interaction between group and question type or stimulus type, indicating that accuracy was not grammatically conditioned (see Discussion).

3.2 ERP Results

After artifact detection and correction, the mean proportion of good trials in the two experimental conditions for the TD group was 55% (SD = 19%, range: 18–84%), and 56% for the DLD group (SD = 18%, range: 35–99%). In terms of actual numbers of trials per condition, the TD group averaged 35 trials (SD = 12) for the control condition and 36 trials (SD = 12) for the filled gap condition. For the DLD group, the average was 35 trials (SD = 14) for the control condition and 35 trials (SD = 14) for the filled gap condition. Thus, the groups were descriptively similar in terms of how many trials were included per condition.

As stated in the Methods section, we chose to remove trials with eyeblinks, rather than using ICA to subtract blink activity. The current study started out with 64 delivered trials per condition, twice as many as in Hestvik et al. (2007); therefore, the remaining trial count after blinks were removed was still fairly high for this kind of experiment. Although some participants in each group still had a relatively low trial count in each cell, we decided to keep all participants due to the difficulty of finding and recruiting children with DLD; cf. Faul et al. (2007) who point out that one must compromise between single-subject statistical power and being able to serve clinical populations.

3.2.1 Descriptive ERP Results

Figure 4 shows the mean ERPs at Electrode site E14 (left anterior, near AF7 in the 10–10 system) and 84% confidence intervals (CIs) around the filled gap and control conditions. These graphs clearly show that the TD control group exhibited an early AN to the filled gap, between approximately 80 and 120 ms. In contrast,

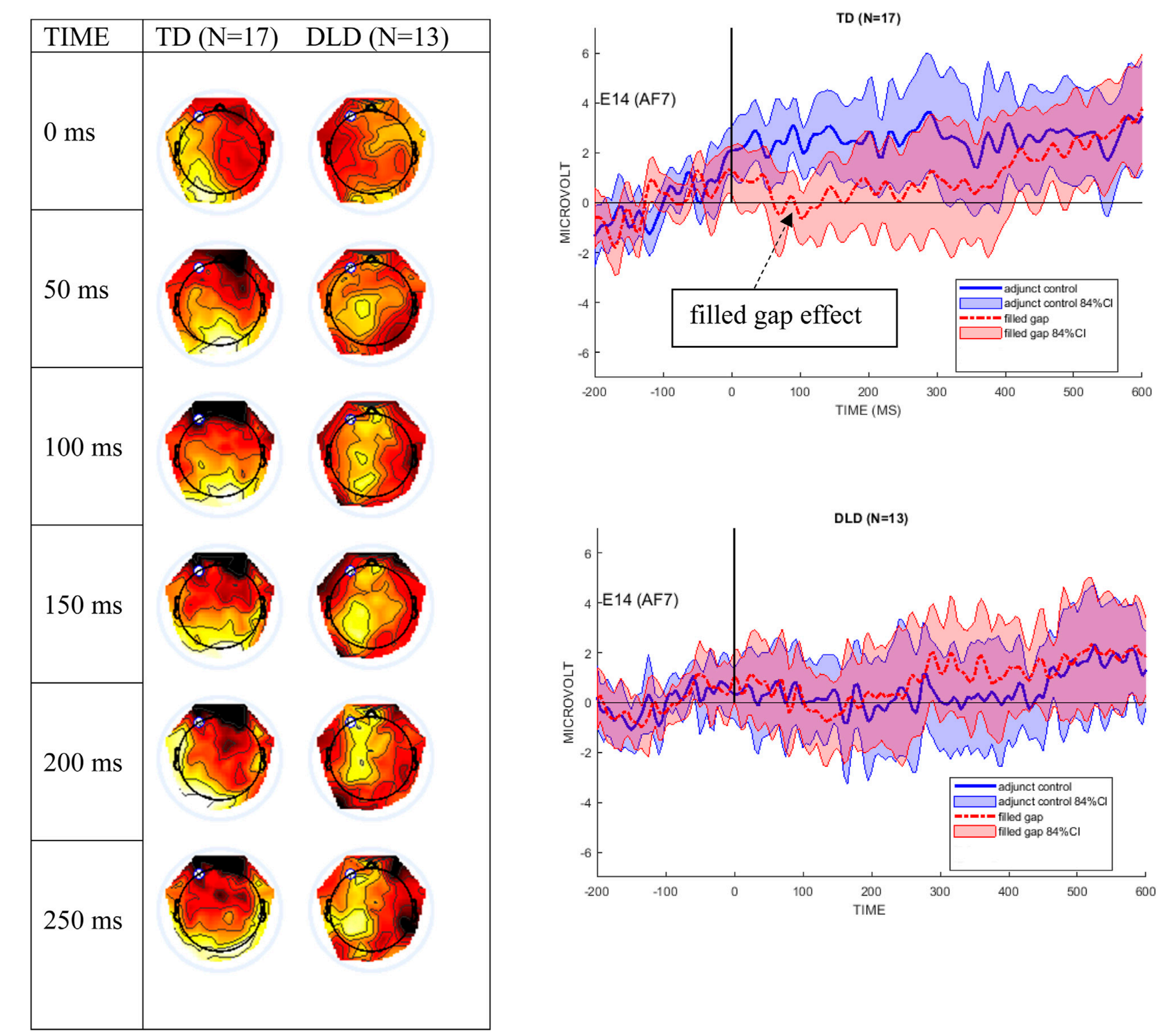


FIGURE 4 | The left panel shows difference wave topographical plots at 50-ms increments from onset of the filled gap NP, for the TD and the DLD groups. Scale for color min/max: −2 µV (dark red) to +2 µV (white). The white dot is electrode E14 (AF7). The right panel shows electrode E14 for the TD and the DLD group, with 84% confidence interval bands around the filled gap vs. the adjunct control waves. We here follow authors who argue that 95% CIs are too conservative for ERP designs (Schenker and Gentleman, 2001; Payton et al., 2003; Dien, 2020).

the DLD group showed no difference between conditions during this early time window. The left graph in **Figure 4** shows the difference wave topography (ERPs to the filled gap minus control) at 50-ms intervals from stimulus onset for both groups. AF7 was the site where the largest effect was observed in previous studies reporting eLAN in a time range of 100–200 ms (Friederici et al., 1993; Hahne and Friederici, 1999), and was therefore chosen to illustrate the effect as waveforms (right panels). As shown, the confidence intervals separate conditions during the eLAN time window, suggesting a meaningful difference.

The DLD group shows an apparent late condition effect from 500–700 ms after stimulus onset. This pattern was characterized by an anterior positivity/right-posterior negativity and is shown in the difference topographical plots in **Figure 5**; the right panel graphs display the mean waveforms at electrode E45 (PO8 in the 10-10 system) with 84% CIs, revealing the greatest difference between conditions from 500–600 ms.

3.2.2 Temporo-Spatial Principal Component Analysis/ Independent Component Analysis Analysis

To determine an objective measure of the temporal and spatial dynamics of the brain response to the filled gap, we first conducted a PCA decomposition of the effects, as outlined in the *Methods* section. The temporal PCA of the difference wave

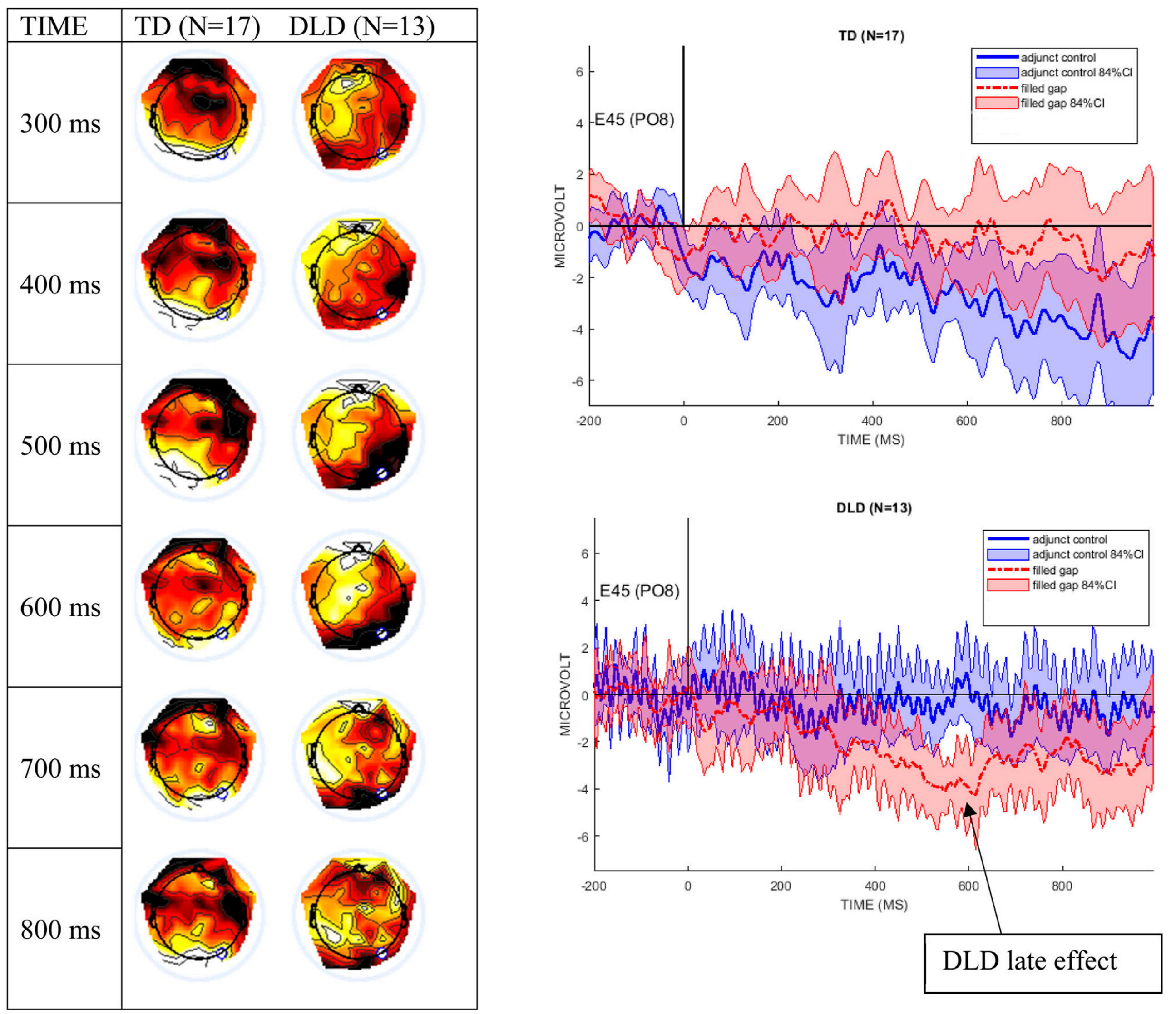


FIGURE 5 | The left panel shows difference wave topographical plots at 100-ms increments between 300 and 800 ms from the onset of the filled gap NP, for the TD controls and the DLD group. Scale for color min/max: −2 μV (dark red) to +2 μV (white). The white dot is electrode E45 (PO8). The right panel shows electrode E45 for the DLD and the TD group, with 85% CI around the filled gap (red) vs. the adjunct control (blue) waves.

(filled gap minus control) resulted in five retained temporal factors for further analysis, based on the criterion of selecting factors that accounted for at least 5% variance. Temporal factor 1 (TF1, peaking at 915 ms) accounted for 40% of the variance; TF2 (485 ms) accounted for 23% of the variance, TF3 (660 ms) accounted for 7%, TF4 (95 ms) accounted for 5% of the variance, and TF5 (235 ms) accounted for 5% of the variance. The spatial PCA step resulted in retaining 5 spatial factors for each temporal factor. The first spatial subfactor in each temporal factor accounted for most of the spatial variance (TF1SF1: 9.8%; TF2SF1: 5.1%; TF3SF1: 2.2%; TF4SF1: 1.6%; TF5SF1: 1.4%). The combined temporo-spatial factors accounted for 63% of the total variance in the data. **Figure 6** below shows the five temporal/spatial components and the peak channel for the difference wave factors for each group, and a topographical plot for the main effect difference wave at the peak latencies.

As shown in **Figure 6**, four of the five factors exhibited an anterior negativity/posterior positivity pattern, from an early time window (TF04SF1, 95 ms) to a late time window (TF01SF1, 915 ms). (We performed the analysis also with linked mastoids as the reference, which did not affect the overall results.) The anterior negativity topographies were strongly driven by the TD control group of children, as can be seen in the figures. In contrast, a late factor (TF03SF1, 660 ms peak latency) exhibited the opposite polarity pattern and was more strongly driven by the DLD group of children. As we will interpret TF04SF1 as the early anterior negativity response to a syntactic category violation, we will

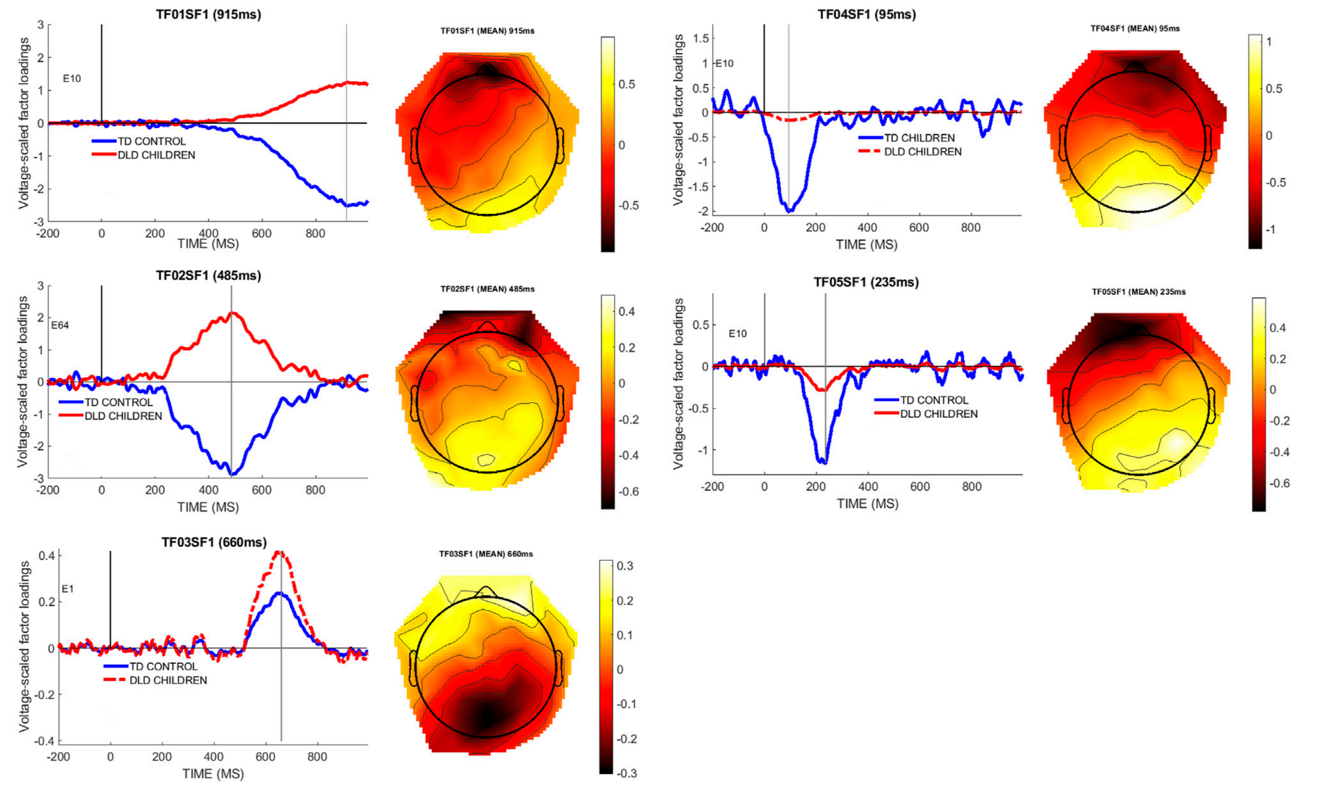


FIGURE 6 | Temporal factors (peak latency indicated) and their spatial distribution; left panels show the microvolt-scaled factor loading waveforms by group. The right panels show the spatial distribution of the effects by using the mean (main effect) spatial component. Note that groups only differ from each other in the amplitude domain for the temporo-spatial factor so the spatial distribution of the factor is identical for all participants, as the PCA/ICA analysis is conducted on the pooled data. The second vertical line indicates the peak latency of the temporal component.

henceforth label it as the "EAN (TF04SF1)" component, to differentiate it from the corresponding voltage component "EAN (voltage-ERP)" derived from this temporo-spatial factor (see below). The reason for this ambiguous denotation is that the PCA component and the voltage ERP represents two different approaches to analyze the same effect in the data.

Preliminary ANOVAS were performed separately for each of the five factors to determine their significance; using the mean factor scores per subject as input, the dependent variable was the factor score for the difference wave used as input to the PCA. Only TF04SF1 (95 ms), and no other temporo-spatial component, exhibited a statistically significant effect of group ($F(1,28) = 4.99$, $p = 0.03$, $\eta^2 = 0.15$. In orthogonal contrast analysis comparing each group mean against zero, using dummy coding (1 for the group to be tested, 0 for the group to leave out), the difference wave was significantly different from 0 for the TD group (estimate = −2.27, t = −3.03, $p = 0.005$), but not for the DLD group (estimate = 0.377, $t = 0.411$, $p = 0.68$), thus explaining the interaction.

To verify that the PCA/ICA factor matched the effect seen in raw data, we compared the EAN (TF04SF1) wave to the difference wave obtained from the raw voltage data, illustrated by the electrode where this PCA/ICA component was largest, specifically, E10 (FPz). **Figure 7** shows the mean voltage waveforms for the control condition, filled gap condition, and the difference waveform, with the EAN (TF04F1) factor waveform overlaid (black dotted line), and illustrates that the temporal-spatial factor models the early negativity in the undecomposed voltage data.

3.2.3 Voltage Analysis Constrained by the Early Bilateral Anterior Negativity (TF04SF1) Component

To analyze the early anterior negativity using a more traditional approach, but that is guided by the PCA/ICA results, we used the method suggested in (Dien, 2012) by selecting a voltage "window" constrained by the PCA solution. We first selected a time window defined by the time points with EAN (TF04SF1) temporal factor loadings exceeding 0.6. This resulted in a 45-160 ms time window, as shown in the left graph of **Figure 8**, left panel. Next, an electrode region was selected by including electrodes that exceeded a factor loading of 0.6 for the EAN (TF04SF1) component. These were sites E1, E2, E3, E6, E7, E8, E10, E11, E12, E61, E62, E58, E59, as shown in **Figure 8**, right panel.[1]

This time/space voltage construct, derived from the temporal and spatial weighting of the PCA/ICA-component TF04SF1, will be labeled "EAN (voltage-ERP)", to express that it derives from

[1]0.6 is an arbitrary treshhold but corresponds to a set of samples where the factor is highly weighted and delineates the effect in time and space.

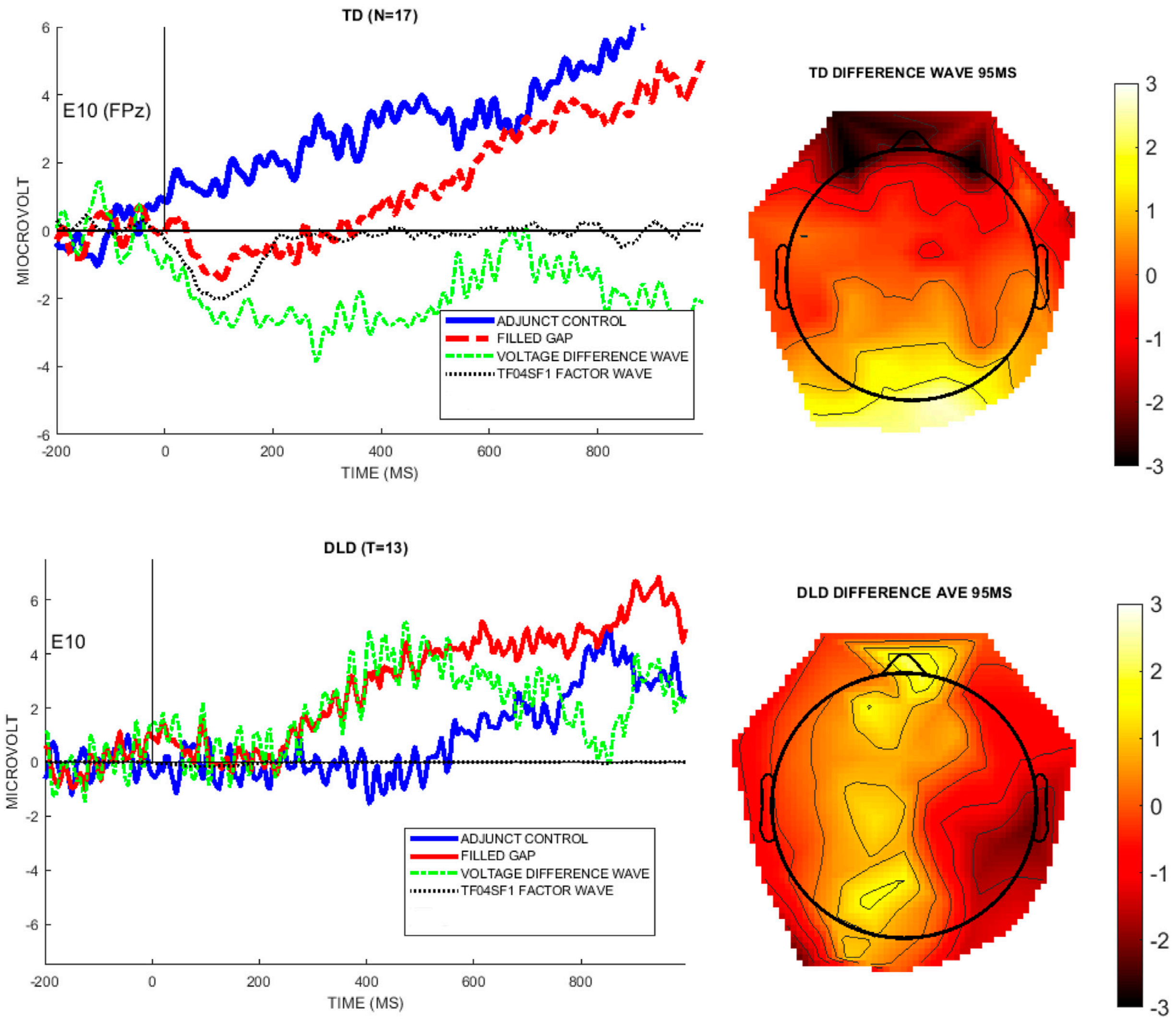


FIGURE 7 | The left graphs show the voltage for each group at the channel with the highest weighting for the EAN component TF04SF1 (E10/FPz). The right images are topographical plots of the raw voltage difference between the Adjunct and Filled gap conditions at the peak latency of TF04 (95 ms), for each group.

the temporo-spatial PCA factor, and represents the same effect (but in unweighted voltage space) as the PCA factor "EAN (TF04SF1)." The mean waveform of the region consisting of these 13 sites is shown per condition and group in **Figure 9**.

The mean voltage for the 45–160 ms time-window and electrode region for each participant and trial in the filled gap and control condition were used as the dependent measures in mixed model statistical analysis.

We performed a linear mixed-effects analysis using R (version 4.1.2) and *lme4* (version 1.1.27; Bates et al., 2015). The input data were the voltage values for each of the trials remaining in each condition after artifact correction, thus varying by subject and cell. We started with the maximal random effects structure and gradually reduced the random effects until the model converged. The fixed effects were Group (typical vs. DLD), Condition (control vs. filled gap) and their interaction. The model converged when we included Subject as a random intercept. We report the model's standardized coefficients after constructing orthogonal contrasts for the fixed effects, using the model parameters function from the *parameters* package (Lüdecke, Ben-Shachar, Patil, and Makowski, 2020), cf. **Table 3**.

The overall effect of each factor was estimated with Type III Wald chi-square tests using the *Anova* function from the *car* package (Fox & Weisberg, 2019). The effect was non-significant for Group ($\chi^2 = 0.865$, $p = 0.352$), Condition ($\chi^2 = 1.872$, $p = 0.171$), and the interaction term was not significant ($\chi^2 = 3.619$, $p = 0.057$); cf. **Figure 10**.

Based on our previous findings for adults using the same paradigm (Hestvik et al., 2012, 2007) and findings that typically developing children exhibit adult-like brain responses to syntactic violations from around 7 years of age (Hahne et al., 2004), we conducted the experiment with the expectation that the TD group should exhibit an eLAN or a similar early anterior negativity. We also expected the experiment to reveal whether DLD children did or did not show this effect. As shown in the interaction plot in **Figure 10**, the

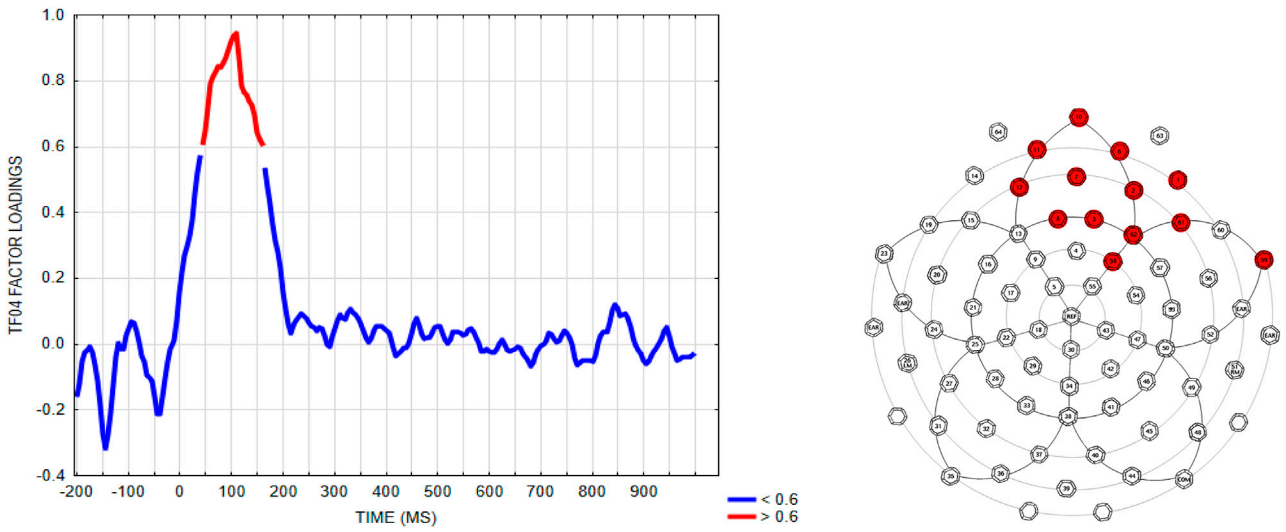


FIGURE 8 | Time samples and electrodes exceeding 0.6 factor loadings selected from the EAN (TF04SF1) component, for deriving a voltage measure from this time window and electrode region.

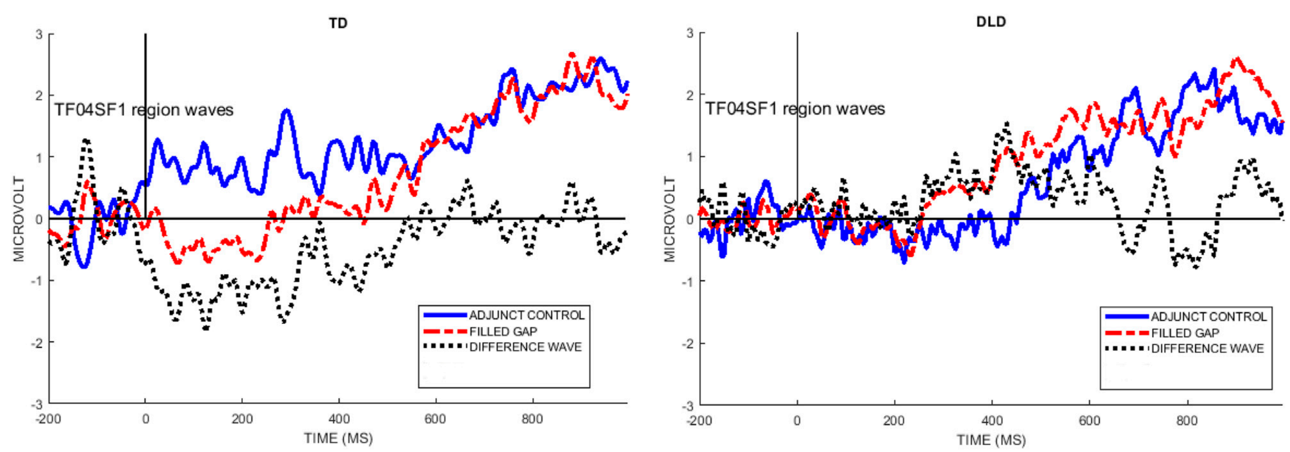


FIGURE 9 | EAN (voltage-ERP); mean voltage waveforms for the EAN-region by group.

TABLE 3 | Results of linear mixed model analysis.

Parameter	Coefficient	SE	95% CI	*t*	*p*
(Intercept)	−0.01	0.03	[−0.06, 0.04]	−0.38	0.701
Group	−0.05	0.06	[−0.16, 0.06]	−0.93	0.352
Condition	−0.06	0.04	[−0.15, 0.03]	−1.37	0.171
Group * Cond	0.17	0.09	[−0.01, 0.34]	1.90	0.057

expectation for the TD group appears to be borne out, while the DLD group shows a flat response.

We therefore set orthogonal contrasts to compare filled gap vs. control in each level of Group. The standardized model coefficients corresponding to the simple effects revealed a significant effect of Condition for the TD control group (b = 0.14, 95% CI = [0.03, 0.26], t = 2.49, $p = 0.013$) but not for the DLD group (b = −0.02, 95% CI = [−0.15, 0.11], t = −0.35, $p = 0.723$). This bears out the prediction that TD children should show an eLAN-like brain response to the prediction violation and reveals that the DLD children do not respond to the violation in this early time window.

3.2.4 Exploratory Analysis of Developmental Language Disorder Late Effect

Although the temporo-spatial factor TF03SF1 did not contain a statistically significant difference between the filled gap and control condition in the factor score analysis, it was the only factor that showed a DLD-specific response to the filled gap. The effect was also visible as a late right-posterior negativity combined with an anterior positivity in the grand average undecomposed voltage data, with a separation of conditions roughly in the 500–700 ms time window,

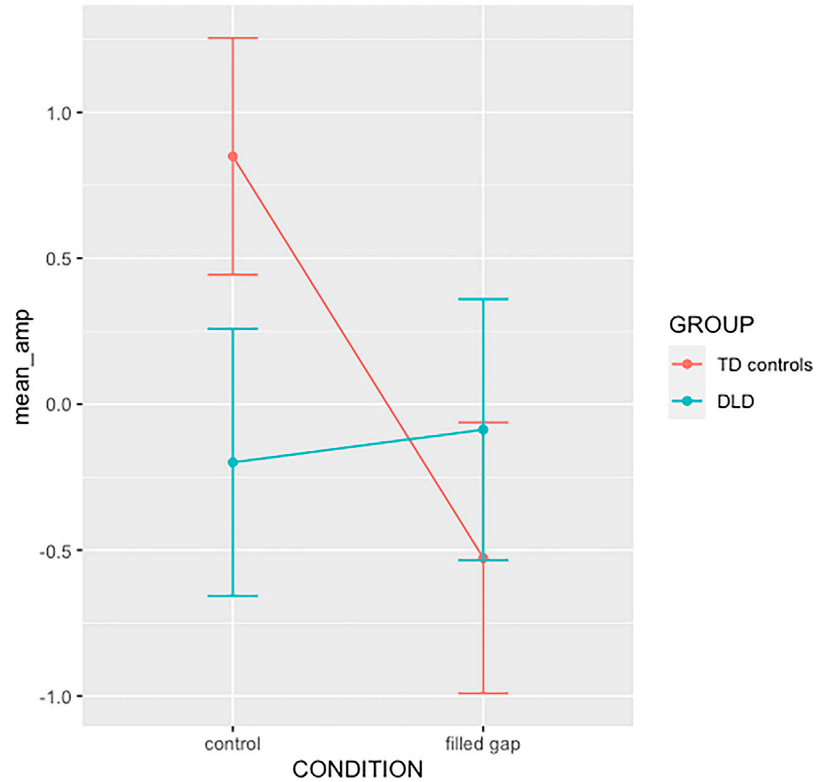


FIGURE 10 | Interaction plot for the 2x2 design GROUP x Condition interaction in TF04SF1. Error bar indicate standard error.

using an 84% confidence interval (cf. **Figure 5**). Given this, as well as previous literature that have reported observing N400 to filled gaps in DLD children (Fonteneau and van der Lely, 2008), we conducted an exploratory PCA analysis limited to the DLD children to ascertain whether there was evidence indicating differential processing of the filled gap and control condition in the brain response

Using the difference wave (filled gap minus adjunct control) as input, the initial temporal PCA retained 12 factors, accounting for 89% of the total variance. The first three temporal factors each accounted for at least 5% of the variance and were selected for analysis. TF01 (980 ms) accounted for 39% variance, TF02 (600 ms) accounted for 26% of the variance, and TF03 (275 ms) account for 6% of the variance. Visual inspection indicated that TF02 in the DLD-only analysis, peaking at 600 ms, captured the same component as TF03 in the analysis with all children pooled, cf. **Figure 5**. The follow-up spatial ICA decomposition of each of the temporal factors retained 4 spatial factors for each temporal factor, based on the criteria used above. These combined temporo-spatial factors accounted for 70% of the DLD data. Among its spatial subfactors, TF02SF2 had the largest factor loadings (mean factor score = 4.8, SD = 8.5). The voltage waveforms for the electrode showing peak positivity for TF02SF2 are shown in **Figure 11**, along with the overall TF2SF2 topography at 600 ms.

Temporal factor loadings exceeding 0.6 were used to construct a time window of 345–635 ms for analysis of the voltage data. All spatial factor loadings were below 0.6; we therefore simply computed the mean voltage in the time window for the peak positive channels (E10/FPz) and tested whether it was different from zero with a *t*-test. This anterior positivity was not significantly different from zero (mean = 4 μV, standard error = 2.33, $t(12) = 1.72$, $p = 0.11$).

The individual participants' mean factor scores (expressing the experimental effect in TF02SF2) are shown in **Figure 12** (following practice recommended in Rousselet et al. (2016)). This reveals heterogeneity in brain responses, suggesting individual differences among DLD children in how their parser responds to the stimuli.

4 DISCUSSION

The aim of this study was to measure whether children with DLD predict gap-positions after encountering fillers, in comparison to their typically developing peers. The current findings revealed that when typically developing children listen to relative clauses like "the zebra that the hippo kissed…", they generate the expectation that there should be no direct object after the verb (because it instead contains a gap). When this prediction is violated by an "unexpectedly filled gap," this triggers an early anterior negativity after about 100 ms after encountering the acoustic signal of an unexpected noun phrase (the word "the"). This brain response is strikingly similar to the anterior negativity observed in the same paradigm with adults (Hestvik et al., 2012, 2007) and suggests that 9–12 year old children with typical development are already showing mature patterns of

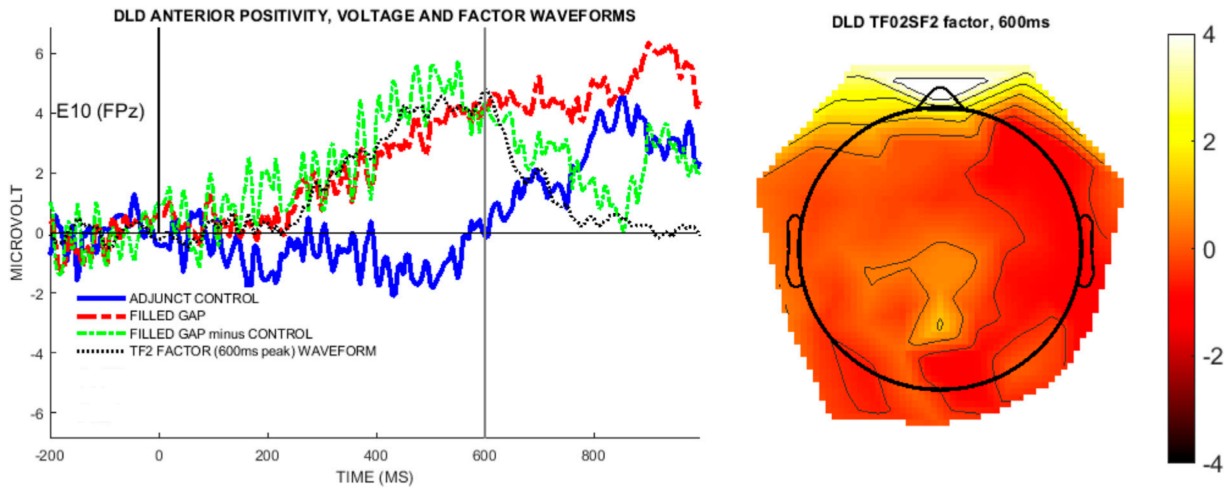


FIGURE 11 | The left panel shows the temporal factor waveform derived from the voltage difference waveform (filled gap minus control), overlaid with the undecomposed voltage waveforms for the filled gap and control condition waveforms, for E10 (FPz), the peak positive channel in TF02SF2. The right topoplot shows the temporospatial factor TF02SF2 at 600 ms.

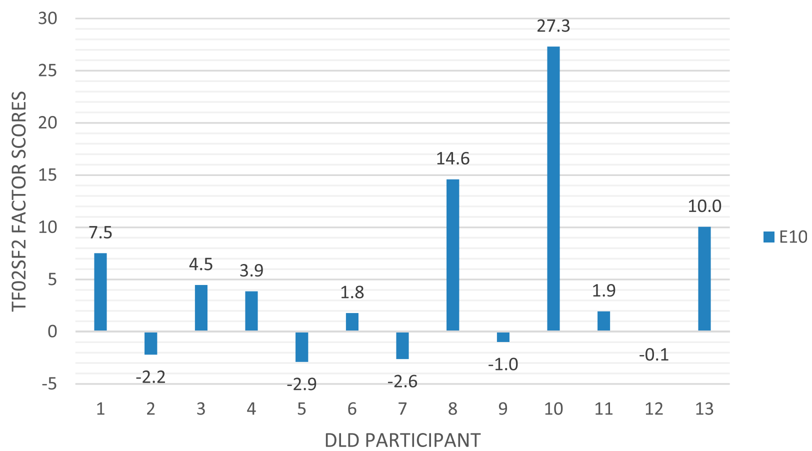


FIGURE 12 | Distribution of individual factor scores (representing the strength to which the individual participant contributed to the late ERP response to the filled gap).

sentence processing, at least for these structures; the same was found by Hahne et al. (2004). In contrast, our data suggest that children with DLD are not processing these structures in a mature fashion, and, in fact, exhibit a complete absence of a filled gap response. We interpret this to mean that children with DLD do not make filler-gap predictions during sentence comprehension. We next discuss several questions arising from this finding.

4.1 Why Does Early AN Provide Evidence of Prediction?

Why does early AN reflect prediction specifically, rather than an integration effect? As noted in the literature, the same ERP pattern can reflect both integration effects and prediction effects (Mantegna et al., 2019). We adopt the view in Dikker et al. (2009) that the earliness of the eLAN itself is a sign of prediction (see also Lau et al., 2006). It is early because top-down grammatical expectations translate into sensory-level predictions of phonetic form (DeLong et al., 2014a; DeLong et al., 2014b; DeLong et al., 2019; Delong et al., 2021). Specifically, a filler predicts a verb phrase with an absent NP. This prediction can be viewed as resulting in pre-activation of a hypothesized parse tree with no NP after the verb. When the parser encounters “the” which indeed introduces a NP, this phonetic signal is therefore highly unexpected. The salience of this phonetic signal plausibly generates a clear surprise response for several reasons. First, the definite determiner is the most frequent word in English (Aiden et al., 2014). Second, it is phonetically unusual, as one of only a handful of function words starting with the voiced dental fricative [ð]. The early nature of the filled gap response is also consistent with recent findings that the brain responds to words around 50 ms after acoustic information is processed (MacGregor et al., 2012). Donhauser and Baillet (2020) using auditory stimuli also found that higher level grammatical predictions translate into predictions at the phonetic level. If a filled gap were to be introduced by a determiner-less NP (such as the bare plural "camels"), this should give rise to a later response, because there is no unique phonetic signal of a bare plural NP. We have examined this prediction elsewhere (Bradley and Hestvik, 2010).

4.2 Lateralization of the Early AN

We observed a bilateral early anterior negativity that was slightly larger over the right than the left sites. Several other studies have also found bilateral early anterior negativity instead of eLAN to syntactic violations in adults (Kessler, 2003; Kessler et al., 2004; Pakulak and Neville, 2011) as well as in children (Sabisch et al., 2009). We interpret the bilateral anterior negativity in our study as functionally equivalent to the eLAN, indicating surprisal for an unexpected syntactic category. We do not assume a strict mapping between neurocognitive processes and the specific ERP latency and topography, but rather that there is a family of ERP responses indicating syntactic processing and syntactic anomaly detection.˙ Alternatively, the eLAN may be bilaterally distributed and the finding of asymmetry is related to other factors that modulate the topography. Shafer et al. (2000) observed an attenuated frontal positivity over left sites time-locked to the onset of grammatical utterances that started with "the" for children with DLD compared to those with TD. In addition, processing of the right frontal sites was enhanced in children with DLD. This pattern suggests an alternative processing route that engages right hemisphere sites. It will be important in future studies to explore such hemispheric differences in processing both grammatical and ungrammatical sentences and in relation to DLD.

4.3 Relationship Between Prediction Impairment and Comprehension

Our behavioral data did not suggest a difference in comprehension between TD and DLD conditional on gap-filling. If children with DLD fail to predict where a gap for a filler should be located, how can they interpret and understand such sentences? It has been suggested that these children interpret filler-gap sentences via alternative processing mechanisms, such as "direct semantic association" (Pickering and Barry, 1991). According to this model, the filler is associated directly with the argument structure of a verb without the syntactic mediation of a gap (Friedmann and Novogrodsky, 2007). If so, the filled gap NP might be analyzed as a referent that cannot be integrated into the argument structure of an already-saturated verb, which predicts a lexical/semantic integration violation and an N400 response (Frisch et al., 2004; Raettig et al., 2010).

Some indication supporting this idea is, as we have shown, that some DLD children did exhibit a later latency ERP effect. However, the observed late DLD ERP response to filled gaps did not reach significance, which could be due to the small sample size ($N = 13$), or be due due to individual differences among the children with DLD, as such heterogeneity has been observed in other studies (Shafer et al., 2007, Shafer et al., 2011).

Behavioral support for the idea that the DLD children interpret filler-gap sentences via alternative routes comes from our results of the comprehension question part of the current experiment. The children were tasked with interpreting grammatical filler-gap stimuli and grammatical filler-gap Wh-questions about the stimuli. In this task, we only observed a main effect of group, such that DLD children had an overall 8% lower accuracy. Crucially, there was no interaction between the sentence type of the stimulus sentence and group: The DLD children exhibit the same accuracy pattern for sentences with filler-gap dependencies vs. no filler-gap dependency. For example, filler-gap sentences were harder to answer correctly than non-filler gap stimulus sentences for both groups. If children with DLD failed to compute the meaning of sentences with filler-gap dependencies, they should exhibit significantly lower accuracy on object relative clause stimulus sentences than children with TD. Similarly, the DLD children exhibited the same pattern of accuracy as a function of whether the question itself contains an object-gap vs. a subject gap. Object gaps require the construction of a filler-gap dependency and they are typically harder to answer correctly than subject Wh-questions and yes/no-questions. Again, DLD children did not perform significantly worse on object Wh-questions vs. subject Wh-questions, than the TD children. These results provide an indication that that DLD children can calculate the meaning of sentences with filler-gap dependencies via alternative processing mechanisms (Friedmann and Novogrodsky, 2007).

4.4 Relationship Between Prediction and Language Acquisition

The "failure to predict" proposed here could play a key role in explaining why some children develop impaired grammatical knowledge. Recent theoretical work on typical language acquisition has emphasized the link between development of syntactic parsing and syntactic acquisition: children must "learn to parse" in order to analyze input and acquire syntax (Trueswell and Gleitman, 2007; Phillips and Ehrenhofer, 2014; Omaki and Lidz, 2015; Pozzan and Trueswell, 2015; Rabagliati et al., 2016). Current acquisition models also emphasize the reliance on error-signals tied to prediction (Dell et al., 2000, 2014; Montgomery and Evans, 2009). The developing child learns by adjusting the parser (probably, at an implicit level) in response to error signals. If children with DLD fail to predict and therefore fail to generate error signals, error-signal driven acquisition mechanisms will not succeed.

4.5 Why do Developmental Language Disorders Children Not Predict?

The current article does not address the underlying cause for the lack of prediction. One possible explanation lies in the lower verbal working memory resources often observed in DLD (Marton and Schwartz, 2003). Elsewhere, we have reported on the Sustained Anterior Negativity ERP as an index of working memory in filler-gap processing, which was elicited to the questions in the current study (Epstein et al., 2013). In their brain response to the comprehension questions, children with typical language development (TD) were expected to show a sustained anterior negativity, reflecting holding the Wh-word in memory until reaching the gap position (Fiebach et al., 2001; Phillips and Ehrenhofer 2014). Adults in Epstein et al. (2013) showed the predicted sustained anterior negativity, whereas children with TD showed a sustained positivity. Children with DLD showed no effect. This suggests that poor performance in long-distance dependencies in children with DLD may be related to low working memory capacity.

In Hestvik et al. (2012) we addressed the relationship between the filled gap response and working memory resources. We conducted a study with typically developed adults and examined whether these participants also exhibited a WM span modulation of the filled gap response. We found a bilateral early anterior negativity (AN) and a P600 to the filled gap, as well as an interaction with verbal memory span such that low span participants exhibit a delayed onset latency of the AN and P600 by about 200 ms (Hestvik et al., 2012). However, the adults with low memory span exhibited the same AN/P600 pattern as high-span listeners, unlike children with DLD who exhibited an absence of early anterior negativity. Therefore, low WM span typical adults do not model DLD children. It is therefore still unclear if reduced working memory explains the complete lack of a filled gap ERP effect in children with DLD, and the underlying cause of lack of prediction during sentence comprehension in this population requires further studies (Jones et al., 2021).

4.6 Limitation and Future Directions

A limitation of the current study is the relatively low number of DLD participants and consequently low statistical power for detecting true effects. This was in large part due to the challenges of recruiting and finding participants that meets the inclusion criteria, despite the reported high prevalence of 7% in the population (Leonard, 2017). While the absence of an early anterior negativity in the DLD group is clear, this makes the interpretation of the observed late ERP response in the DLD group suggestive at this time, and future studies with increased power are needed to replicate this effect and determine whether it generalizes to the population.

5 CONCLUSION

The current study revealed that children with Developmental Language Disorder (DLD) are not using the same neuro-parsing routines in processing long-distance dependencies as children with typical development (TD). Children with TD exhibited an early anterior negativity to a filled gap expectation violation in object relative clauses, which indicates predictive processing. Children with DLD show no similar early brain response, suggesting lack of predictive processing. The DLD children appear to still compute the meaning of relative clauses which suggests that they may use a variety of different strategies to process these sentences, despite their prediction impairment.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article is available at https://osf.io/m84jb/.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the CUNY Internal Review Board. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

AH: Conception, experimental design, stimulus development, software programming, data collection, signal processing, statistical analysis, writing, funding BE: Experimental design, stimulus development, data collection, clinical evaluation of participants, writing RS: Interpretation, writing, funding VS: Interpretation, writing, laboratory resources.

FUNDING

This work was supported by the National Institutes of Health, NIDCD 5R03DC006175 (AH, P.I.) and 5T32DC00039-09 (RS, P.I.).

ACKNOWLEDGMENTS

The authors wish to thank Dr. Nathan Maxfield for his essential contributions, including the idea that gap-filling should be studied in children with a passive listening paradigm, and that ERPs should be analyzed with temporo-spatial PCA. Thanks to Ray Jackendoff for design discussions, and to Joseph Dien for developing and providing support for the ERP PCA Toolkit. Thanks to J. P. Laurencau and Chao Han for statistical consulting.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcomm.2021.637585/full#supplementary-material

REFERENCES

Aiden, E., Michel, J.-B., and Michel, J.-B. (2014). Uncharted: Big Data as a Lens on Human Culture. *Choice Rev. Online* 52 (Issue 01), 52–0314. doi:10.5860/choice.52-0314

Bates, D., Mächler, M., Bolker, B. M., and Walker, S. C. (2015). Fitting Linear Mixed-Effects Models Using lme4 *Journal of Statistical Software* 67 (1). doi:10.18637/jss.v067.i01

Bell, A. J., and Sejnowski, T. J. (1995). An Information-Maximization Approach to Blind Separation and Blind Deconvolution. *Neural Comput.* 7 (6), 1129–1159. doi:10.1162/neco.1995.7.6.1129

Bishop, D. V. M. (2013). Cerebral Asymmetry and Language Development: Cause, Correlate, or Consequence? *Science* 340 (June), 1302–1310. doi:10.1126/science.1230531

Bradley, E. D., and Hestvik, A. (2010). *Testing the Sensory Hypothesis of the Early Left Anterior Negativity with Auditory Stimuli*. LSA Annual Meeting Extended Abstracts. doi:10.3765/exabs.v0i0.524

Brown, L., Sherbenou, R. J., Johnson, S., and Johnsen, S. K. (1997). in *Test of Nonverbal Intelligence*. 3rd ed. PRO-ED.

Brunellière, A., Franck, J., Ludwig, C., and Frauenfelder, U. H. (2007). Early and Automatic Syntactic Processing of Person Agreement. *Neuroreport* 18 (6), 537–541. doi:10.1097/wnr.0b013e3280b07ba1

Clark, H. H. (1973). The Language-As-Fixed-Effect Fallacy: A Critique of Language Statistics in Psychological Research. *J. Verbal Learn. Verbal Behav.* 12 (4), 335–359. doi:10.1016/S0022-5371(73)80014-3

Schuele, C. M., and Nicholls, L. (2000). Relative Clauses: Evidence of Continued Linguistic Vulnerability in Children with Specific Language Impairment. *Clin. Linguistics Phonetics* 14 (8), 563–585. doi:10.1080/026992000750048116

Corballis, M. C. (2014). Left Brain, Right Brain: Facts and Fantasies. *Plos Biol.* 12 (1), e1001767. doi:10.1371/journal.pbio.1001767

Crain, S., and Fodor, J. D. (1985). "How Can Grammars Help Parsers," in *Natural Language Parsing: Psychological, Computational and Theoretical Perspectives*. Editors D. Dowty, L. Karttunen, and A. Zwicky (Cambridge University Press).

Deevy, P., and Leonard, L. B. (2004). The Comprehension of Wh- Questions in Children with Specific Language Impairment. *J. Speech, Lang. Hearing Res.* 47 (August), 802–815. doi:10.1044/1092-4388(2004/060)

Dell, G. S., Chang, F., and B, P. T. R. S. (2014). *Comprehension and Acquisition the P-Chain : Relating Sentence Production and its Disorders to Comprehension and Acquisition Author for Correspondence : December 2013.*

Dell, G. S., Reed, K. D., Adams, D. R., and Meyer, A. S. (2000). Speech Errors, Phonotactic Constraints, and Implicit Learning: a Study of the Role of Experience in Language Production. *J. Exp. Psychol. Learn. Mem. Cogn.* 26 (6), 1355–1367. doi:10.1037/0278-7393.26.6.1355

DeLong, K. A., Quante, L., and Kutas, M. (2014a). Predictability, plausibility, and two late ERP positivities during written sentence comprehension. *Neuropsychologia* 61, 150–162. doi:10.1016/j.neuropsychologia.2014.06.016

DeLong, K. A., Troyer, M., and Kutas, M. (2014b). Pre-Processing in Sentence Comprehension: Sensitivity to Likely Upcoming Meaning and Structure. *Lang. Ling. Comp.* 8 (12), 631–645. doi:10.1111/lnc3.12093

DeLong, K. A., Chan, W., and Kutas, M. (2019). Similar Time Courses for Word form and meaning Preactivation During Sentence Comprehension. *Psychophysiology* 56 (4). doi:10.1111/psyp.13312

DeLong, K. A., Chan, W., and Kutas, M. (2021). Testing Limits: ERP Evidence for Word form Preactivation During Speeded Sentence Reading. *Psychophysiology*, 1–7. doi:10.1111/psyp.13720

Delorme, A., and Makeig, S. (2004). EEGLAB: An Open Source Toolbox for Analysis of Single-Trial EEG Dynamics Including Independent Component Analysis. *J. Neurosci. Methods* 134 (1), 9–21. doi:10.1016/j.jneumeth.2003.10.009

Dien, J. (2012). Applying Principal Components Analysis to Event-Related Potentials: A Tutorial. *Developmental Neuropsychol.* 37 (6), 497–517. doi:10.1080/87565641.2012.697503

Dien, J. (2020). *ERP PCA Toolkit 2.90 Tutorial.*

Dien, J., and Frishkoff, G. A. (2005). "Principal Components Analysis of ERP Data," in *Event-Related Potentials: A Methods Handbook*. Editor T. Handy (MIT Press).

Dien, J. (2010). The ERP PCA Toolkit: An Open Source Program for Advanced Statistical Analysis of Event-Related Potential Data. *J. Neurosci. Methods* 187, 138–145. doi:10.1016/j.jneumeth.2009.12.009

Donhauser, P. W., and Baillet, S. (2020). Two Distinct Neural Timescales for Predictive Speech Processing. *Neuron* 105 (2), 385–393. doi:10.1016/j.neuron.2019.10.019

Dunn, L. M., and Dunn, D. M. (2007). *Peabody Picture Vocabulary Test*. 4th ed. Bloomington, MN: Pearson, Inc.

Epstein, B., Hestvik, A., Shafer, V. L., and Schwartz, R. G. (2013). ERPs Reveal Atypical Processing of Subject versus objectWh-Questions in Children with Specific Language Impairment. *Int. J. Lang. Commun. Disord.* 48 (4), 351–365. doi:10.1111/1460-6984.12009

Evans, J. L., Saffran, J. R., and Robe-Torres, K. (2009). Statistical Learning in Children with Specific Language Impairment. *J. Speech Lang. Hear. Res.* 52 (2), 321–335. doi:10.1044/1092-4388(2009/07-0189)

Faul, F., Erdfelder, E., Lang, A.-G., and Buchner, A. (2007). G*Power 3: A Flexible Statistical Power Analysis Program for the Social, Behavioral, and Biomedical Sciences. *Behav. Res. Methods* 39, 175–191. doi:10.3758/bf03193146

Felser, C., and Jessen, A. (2020). "Brain Responses Elicited by Implausible Fillers and Filled Object Gaps in German," in *Typical and Impaired Processing in Morphosyntax*. Editor V. Torrens (John Benjamins Publishing Company), 75–90. doi:10.1075/lald.64.04fel

Ferree, T. C., Luu, P., Russell, G. S., and Tucker, D. M. (2001). Scalp Electrode Impedance, Infection Risk, and EEG Data Quality. *Clin. Neurophysiol.* 112 (3), 536–544. doi:10.1016/S1388-2457(00)00533-2

Fiebach, C. J., Schlesewsky, M., and Friederici, A. D. (2001). Syntactic Working Memory and the Establishment of Filler-gap Dependencies: Insights from ERPs and fMRI. *J. Psycholinguistic Res.* 30 (3), 321–338. doi:10.1023/a:1010447102554

Fonteneau, E., and van der Lely, H. K. J. (2008). Electrical Brain Responses in Language-Impaired Children Reveal Grammar-specific Deficits. *PLoS One* 3 (3), e1832. doi:10.1371/journal.pone.0001832

Frazier, L., and Fodor, J. D. (1978). The Sausage Machine: A New Two-Stage Parsing Model. *Cognition* 6 (4), 291–325. doi:10.1016/0010-0277(78)90002-1

Friederici, A. D., Pfeifer, E., and Hahne, A. (1993). Event-related Brain Potentials during Natural Speech Processing: Effects of Semantic, Morphological and Syntactic Violations. *Cogn. Brain Res.* 1 (3), 183–192. doi:10.1016/0926-6410(93)90026-2

Friederici, A. D. (2012). The Cortical Language Circuit: From Auditory Perception to Sentence Comprehension. *Trends Cogn. Sci.* 16 (5), 262–268. doi:10.1016/j.tics.2012.04.001

Friedmann, N., and Novogrodsky, R. (2007). Is the Movement Deficit in Syntactic SLI Related to Traces or to Thematic Role Transfer? *Brain Lang.* 101 (1), 50–63. doi:10.1016/j.bandl.2006.09.006

Friedmann, N., and Novogrodsky, R. (2011). Which Questions Are Most Difficult to Understand? *Lingua* 121 (3), 367–382. doi:10.1016/j.lingua.2010.10.004

Frisch, S., Hahne, A., and Friederici, A. D. (2004). Word Category and Verb-Argument Structure Information in the Dynamics of Parsing. *Cognition* 91 (3), 191–219. doi:10.1016/j.cognition.2003.09.009

Friston, K. (2005). A Theory of Cortical Responses. *Phil. Trans. R. Soc. B* 360, 815–836. doi:10.1098/rstb.2005.1622

Friston, K., and Kiebel, S. (2009). Predictive Coding under the Free-Energy Principle. *Phil. Trans. R. Soc. B* 364 (1521), 1211–1221. doi:10.1098/rstb.2008.0300

Gambi, C., Gorrie, F., Pickering, M. J., and Rabagliati, H. (2018). The Development of Linguistic Prediction: Predictions of Sound and Meaning in 2- to 5-Year-Olds. *J. Exp. Child Psychol.* 173, 351–370. doi:10.1016/j.jecp.2018.04.012

Gorsuch, R. L. (1983). *Factor Analysis*. Erlbaum.

Grey, S., Tanner, D., and van Hell, J. G. (2017). How Right Is Left? Handedness Modulates Neural Responses during Morphosyntactic Processing. *Brain Res.* 1669, 27–43. doi:10.1016/j.brainres.2017.05.024

Hahne, A., Eckstein, K., and Friederici, A. D. (2004). Brain Signatures of Syntactic and Semantic Processes during Children's Language Development. *J. Cogn. Neurosci.* 16 (7), 1302–1318. doi:10.1162/0898929041920504

Hahne, A., and Friederici, A. D. (1999). Electrophysiological Evidence for Two Steps in Syntactic Analysis: Early Automatic and Late Controlled Processes. *J. Cogn. Neurosci.* 11 (2), 194–205. doi:10.1162/089892999563328

Hendrickson, A. E., and White, P. O. (1964). Promax: a Quick Method for Rotation to Oblique Simple Structure. *Br. J. Stat. Psychol.* 17 (1), 65–70. doi:10.1111/j.2044-8317.1964.tb00244.x

Hesketh, A. (2006). The Use of Relative Clauses by Children with Language Impairment. *Clin. Linguist Phon* 20 (7), 539–546. doi:10.1080/02699200500266398

Hestvik, A., Bradley, E., and Bradley, C. (2012). Working Memory Effects of Gap-Predictions in Normal Adults: An Event-Related Potentials Study. *J. Psycholinguist Res.* 41 (6), 425–438. doi:10.1007/s10936-011-9197-8

Hestvik, A., Maxfield, N., Schwartz, R. G., and Shafer, V. (2007). Brain Responses to Filled Gaps. *Brain Lang.* 100 (3), 301–316. doi:10.1016/j.bandl.2006.07.007

Hestvik, A., Schwartz, R. G., and Tornyova, L. (2010). Relative Clause Gap-Filling in Children with Specific Language Impairment. *J. Psycholinguist Res.* 39 (5), 443–456. doi:10.1007/s10936-010-9151-1

Hinojosa, J., Martín-Loeches, M., Casado, P., Muñoz, F., and Rubia, F. (2003). Similarities and Differences between Phrase Structure and Morphosyntactic Violations in Spanish: An Event-Related Potentials Study. *Lang. Cogn. Process.* 18 (2), 113–142. doi:10.1080/01690960143000489

Horn, J. L. (1965). A Rationale and Test for the Number of Factors in Factor Analysis. *Psychometrika* 30 (2), 179–185. doi:10.1007/bf02289447

Isel, F., Hahne, A., Maess, B., and Friederici, A. D. (2007). Neurodynamics of Sentence Interpretation: ERP Evidence from French. *Biol. Psychol.* 74 (3), 337–346. doi:10.1016/j.biopsycho.2006.09.003

Joanisse, M. F., and Seidenberg, M. S. (2003). Phonology and Syntax in Specific Language Impairment: Evidence from a Connectionist Model. *Brain Lang.* 86 (1), 40–56. doi:10.1016/S0093-934X(02)00533-3

Jones, S. D., Westermann, G., Camarata, S. M., and Nittrouer, S. (2021). Predictive Processing and Developmental Language Disorder. *J. Speech Lang. Hear. Res.* 64 (1), 181–185. doi:10.1044/2020_JSLHR-20-00409

Jung, T.-P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., and Sejnowski, T. J. (2001). Analysis and Visualization of Single-Trial Event-Related Potentials. *Hum. Brain Mapp.* 14 (3), 166–185. doi:10.1002/hbm.1050

Kail, R. V., and Miller, C. A. (2006). Developmental Change in Processing Speed: Domain Specificity and Stability during Childhood and Adolescence. *J. Cogn. Development* 7 (1), 119–137. doi:10.1207/s15327647jcd0701_6

Kessler, K. L. (2003). *ERP Correlates of Morphosyntactic and Word-Order Phenomena in Native Speakers and Second-Language Learners of English.* New York, NY: CUNY Graduate Center.

Kessler, K. L., Martaharjono, G., and Shafer, V. L. (2004). "ERP Correlates of Age and Proficiency in L2 Processing of Syntactic and Inflectional Information," in Proceedings of the 28th Annual Boston University Conference on Language Development, 294–305.

Kim, N., Brehm, L., Sturt, P., and Yoshida, M. (2020). How Long Can You Hold the Filler: Maintenance and Retrieval. *Lang. Cogn. Neurosci.* 35 (1), 17–42. doi:10.1080/23273798.2019.1626456

Knecht, S., Dräger, B., Deppe, M., Bobe, L., Lohmann, H., Flöel, A., et al. (2000). Handedness and Hemispheric Language Dominance in Healthy Humans. *Brain* 123 (12), 2512–2518. doi:10.1093/brain/123.12.2512

Knösche, T. R., Maess, B., and Friederici, A. D. (1999). Processing of Syntactic Information Monitored by Brain Surface Current Density Mapping Based on MEG. *Brain Topogr* 12 (2), 75–87. doi:10.1023/a:1023442426799

Kubota, M., Ferrari, P., and Roberts, T. P. L. (2003). Magnetoencephalography Detection of Early Syntactic Processing in Humans: Comparison between L1 Speakers and L2 Learners of English. *Neurosci. Lett.* 353 (2), 107–110. doi:10.1016/j.neulet.2003.09.019

Kubota, M., Ono, Y., Ishiyama, A., Zouridakis, G., and Papanicolaou, A. C. (2018). Magnetoencephalography Reveals Mismatch Field Enhancement from Unexpected Syntactic Category Errors in English Sentences. *Neurosci. Lett.* 662, 195–204. doi:10.1016/j.neulet.2017.07.051

Kuperberg, G. R., and Jaeger, T. F. (2016). What Do We Mean by Prediction in Language Comprehension? *Lang. Cogn. Neurosci.* 31 (1), 32–59. doi:10.1080/23273798.2015.1102299

Lau, E., Stroud, C., Plesch, S., and Phillips, C. (2006). The Role of Structural Prediction in Rapid Syntactic Analysis. *Brain Lang.* 98 (1), 74–88. doi:10.1016/j.bandl.2006.02.003

Leonard, L. B., and Bortolini, U. (1998). Grammatical Morphology and the Role of Weak Syllables in the Speech of Italian-Speaking Children with Specific Language Impairment. *J. Speech Lang. Hear. Res.* 41 (6), 1363–1374. doi:10.1044/jslhr.4106.1363

Leonard, L. B., Ellis Weismer, S., Miller, C. A., Francis, D. J., Tomblin, J. B., and Kail, R. V. (2007). Speed of Processing, Working Memory, and Language Impairment in Children. *J. Speech Lang. Hear. Res.* 50 (2), 408–428. doi:10.1044/1092-4388(2007/029)

Leonard, L. B. (2017). "Specific Language Impairment," in *Oxford Research Encyclopedia of Psychology* (Oxford University Press), 1–24. doi:10.1093/acrefore/9780190236557.013.64

Levy, R. (2008). Expectation-based Syntactic Comprehension. *Cognition* 106 (3), 1126–1177. doi:10.1016/j.cognition.2007.05.006

Lewis, R. L., and Vasishth, S. (2005). An Activation-Based Model of Sentence Processing as Skilled Memory Retrieval. *Cogn. Sci.* 29 (3), 375–419. doi:10.1207/s15516709cog0000_25

Love, T., and Swinney, D. (1996). Coreference Processing and Levels of Analysis in Object-Relative Constructions; Demonstration of Antecedent Reactivation with the Cross-Modal Priming Paradigm. *J. Psycholinguist Res.* 25 (1), 5–24. doi:10.1007/BF01708418

Luck, S. J., and Gaspelin, N. (2017). How to Get Statistically Significant Effects in Any ERP experiment (And Why You Shouldn't). *Psychophysiol* 54 (1), 146–157. doi:10.1111/psyp.12639

Luu, P., and Ferree, T. C. (2005). *Determination of the HydroCel Geodesic Sensor Nets' Average Electrode Positions and Their 10-10 International Equivalents*. Portland, OR: Electrical Geodesics, Inc.

MacGregor, L. J., Pulvermüller, F., van Casteren, M., and Shtyrov, Y. (2012). Ultra-rapid Access to Words in the Brain. *Nat. Commun.* 3, 711. doi:10.1038/ncomms1715

Makeig, S., Jung, T.-P., Bell, A. J., Ghahremani, D., and Sejnowski, T. J. (1997). Blind Separation of Auditory Event-Related Brain Responses into Independent Components. *Proc. Natl. Acad. Sci.* 94 (20), 10979–10984. doi:10.1073/pnas.94.20.10979

Makeig, S., Westerfield, M., Jung, T.-P., Enghoff, S., Townsend, J., Courchesne, E., et al. (2002). Dynamic Brain Sources of Visual Evoked Responses. *Science* 295 (5555), 690–694. doi:10.1126/science.1066168

Mantegna, F., Hintz, F., Ostarek, M., Alday, P. M., and Huettig, F. (2019). Distinguishing Integration and Prediction Accounts of ERP N400 Modulations in Language Processing through Experimental Design. *Neuropsychologia* 134 (September), 107199. doi:10.1016/j.neuropsychologia.2019.107199

Marinis, T., and van der Lely, H. K. J. (2007). On-line Processing of Wh-questions in Children with G-SLI and Typically Developing Children. *Int. J. Lang. Commun. Disord.* 42 (5), 557–582. doi:10.1080/13682820601058190

Marton, K., and Schwartz, R. G. (2003). Working Memory Capacity and Language Processes in Children with Specific Language Impairment. *J. Speech Lang. Hear. Res.* 46 (5), 1138–1153. doi:10.1044/1092-4388(2003/08910.1044/1092-4388(2003/089)

McElree, B. (2000). Sentence Comprehension Is Mediated by Content-Addressable Memory Structures. *J. Psycholinguistic Res.* 29 (2), 111–123. doi:10.1023/A:1005184709695

Miller, C. A., Kail, R., Leonard, L. B., and Tomblin, J. B. (2001). Speed of Processing in Children with Specific Language Impairment. *J. Speech Lang. Hear. Res.* 44 (2), 416–433. doi:10.1044/1092-4388(2001/034)

Montgomery, J. W., and Evans, J. L. (2009). Complex Sentence Comprehension and Working Memory in Children with Specific Language Impairment. *J. Speech Lang. Hear. Res.* 52 (2), 269–288. doi:10.1044/1092-4388(2008/07-0116)

Montgomery, J. W., and Evans, J. L. (2017). "Whatdunit?" Sentence Comprehension Abilities of Children with SLI: Sensitivity to Word Order in Canonical and. *J. Speech Lang. Hearing Res.* August, 1–16. doi:10.1044/2017

Montgomery, J. W., Gillam, R. B., Evans, J. L., and Sergeev, A. V. (2017). "Whatdunit?" Sentence Comprehension Abilities of Children with SLI: Sensitivity to Word Order in Canonical and Noncanonical Structures. *J. Speech Lang. Hear. Res.* 60 (9), 2603–2618. doi:10.1044/2017_JSLHR-L-17-0025

Montgomery, J. W. (2004). Sentence Comprehension in Children with Specific Language Impairment: Effects of Input Rate and Phonological Working Memory. *Int. J. Lang. Commun. Disord.* 39 (1), 115–133. doi:10.1080/13682820310001616985

Montgomery, J. W. (2003). Working Memory and Comprehension in Children with Specific Language Impairment: what We Know So Far. *J. Commun. Disord.* 36 (3), 221–231. doi:10.1016/S0021-9924(03)00021-2

Münte, T. F., Heinze, H.-J., Mangun, G. R., Munte, T. F., Heinze, H.-J. J., and Mangun, G. R. (1993). Dissociation of Brain Activity Related to Syntactic and Semantic Aspects of Language. *J. Cogn. Neurosci.* 5 (3), 335–344. doi:10.1162/jocn.1993.5.3.335

Neville, H., Nicol, J. L., Barss, A., Forster, K. I., and Garrett, M. F. (1991). Syntactically Based Sentence Processing Classes: Evidence from Event-Related Brain Potentials. *J. Cogn. Neurosci.* 3 (2), 151–165. doi:10.1162/jocn.1991.3.2.151

Nicholas Nagel, H., Shapiro, L. P., and Nawy, R. (1994). Prosody and the Processing of Filler-gap Sentences. *J. Psycholinguist Res.* 23 (6), 473–485. doi:10.1007/BF02146686

Nicol, J., Swinney, D., Love, T., and Hald, L. (2006). The On-Line Study of Sentence Comprehension: an Examination of Dual Task Paradigms. *J. Psycholinguist Res.* 35 (3), 215–231. doi:10.1007/s10936-006-9012-0

Nieuwland, M. S., and Kazanina, N. (2020). The Neural Basis of Linguistic Prediction: Introduction to the Special Issue. *Neuropsychologia* 146, 107532. doi:10.1016/j.neuropsychologia.2020.107532

Omaki, A., and Lidz, J. (2015). Linking Parser Development to Acquisition of Syntactic Knowledge. *Lang. Acquisition* 22 (2), 158–192. doi:10.1080/10489223.2014.943903

Ouchikh, Y., Pazuelo, L., Wadhera, D., Gazman, Z., Aramridth, T., Scheuer, J., et al. (2016). *Working Memory Updating and Interference Control in Children with SLI and Their TLD Peers*, 2.

Pakulak, E., and Neville, H. J. (2011). Maturational Constraints on the Recruitment of Early Processes for Syntactic Processing. *J. Cogn. Neurosci.* 23 (10), 2752–2765. doi:10.1162/jocn.2010.21586

Parr, T., and Friston, K. J. (2018). The Anatomy of Inference: Generative Models and Brain Structure. *Front. Comput. Neurosci.* 12. doi:10.3389/fncom.2018.00090

Payton, M. E., Greenstone, M. H., and Schenker, N. (2003). Overlapping Confidence Intervals or Standard Error Intervals: What Do They Mean in Terms of Statistical Significance? *J. Insect Sci.* 3 (34), 1–6. doi:10.1673/031.003.3401

Phillips, C., and Ehrenhofer, L. (2014). The Role of Language Processing in Language Acquisition. *Linguistic Approaches to Bilingualism* 5, 409–453. doi:10.1075/lab.5.4.01phi

Pickering, M., and Barry, G. (1991). Sentence Processing without Empty Categories. *Lang. Cogn. Process.* 6, 229–259. doi:10.1080/01690969108406944

Pickering, M. J., and Gambi, C. (2018). Predicting while Comprehending Language: A Theory and Review. *Psychol. Bull.* 144 (10), 1002–1044. doi:10.1037/bul0000158

Plante, E., Patterson, D., Sandoval, M., Vance, C. J., and Asbjørnsen, A. E. (2017). An fMRI Study of Implicit Language Learning in Developmental Language Impairment. *NeuroImage: Clin.* 14, 277–285. doi:10.1016/j.nicl.2017.01.027

Plante, E., Swisher, L., Kiernan, B., and Restrepo., M. A. (1993). Language Matches. *J. Speech Lang. Hear. Res.* 36 (4), 772–776. doi:10.1044/jshr.3604.772

Pozzan, L., and Trueswell, J. C. (2015). Revise and Resubmit: How Real-Time Parsing Limitations Influence Grammar Acquisition. *Cogn. Psychol.* 80, 73–108. doi:10.1016/j.cogpsych.2015.03.004

R Core Team (2017). R: A language and environment for statistical computing (Version 3.4.3). Vienna, Austria: R Foundation for Statistical Computing. Available at: http://www.R- project.org/

Rösler, F., Pütz, P., Friederici, A., and Hahne, A. (1993). Event-Related Brain Potentials while Encountering Semantic and Syntactic Constraint Violations. *J. Cogn. Neurosci.*, 5(3), 345–362. doi:10.1162/jocn.1993.5.3.345

Rabagliati, H., Gambi, C., and Pickering, M. J. (2016). Learning to Predict or Predicting to Learn? *Lang. Cogn. Neurosci.* 31 (1), 94–105. doi:10.1080/23273798.2015.1077979

Raettig, T., Frisch, S., Friederici, A. D., and Kotz, S. A. (2010). Neural Correlates of Morphosyntactic and Verb-Argument Structure Processing: An EfMRI Study. *Cortex*, 46(5), 613–620. doi:10.1016/j.cortex.2009.06.003

Richman, M. B. (1986). Rotation of Principal Components. *J. Climatol.* 6, 293–335. doi:10.1002/joc.3370060305

Rousselet, G. A., Foxe, J. J., and Bolam, J. P. (2016). A Few Simple Steps to Improve the Description of Group Results in Neuroscience. *Eur. J. Neurosci.* 44, 2647–2651. doi:10.1111/ejn.13400

Sabisch, B., Hahne, C. A., Glass, E., von Suchodoletz, W., and Friederici, A. D. (2009). Children with Specific Language Impairment: The Role of Prosodic Processes in Explaining Difficulties in Processing Syntactic Information. *Brain Res.* 1261, 37–44. doi:10.1016/J.BRAINRES.2009.01.012

Schenker, N., and Gentleman, J. F. (2001). On Judging the Significance of Differences by Examining the Overlap between Confidence Intervals. *The Am. Statistician* 55 (3), 182–186. doi:10.1198/000313001317097960

Schneider, W., Eschman, A., and Zuccolotto, A. (2002). *E-prime Reference Guide*. Psychology Software Tools, Inc.

Semel, E., Wiig, E. H., and Secord, W. A. (2004). *Clinical Evaluation of Language Fundamentals*. 4th ed. Toronto, ON, Canada: The Psychological Corporation/A Harcourt Assessment Company.

Shafer, V. L., Ponton, C., Datta, H., Morr, M. L., and Schwartz, R. G. (2007). Neurophysiological Indices of Attention to Speech in Children with Specific Language Impairment. *Clin. Neurophysiol.* 118 (6), 1230–1243. doi:10.1016/j.clinph.2007.02.023

Shafer, V. L., Schwartz, R. G., and Martin, B. (2011). Evidence of Deficient central Speech Processing in Children with Specific Language Impairment: the T-Complex. *Clin. Neurophysiol.* 122 (6), 1137–1155. doi:10.1016/j.clinph.2010.10.046

Shafer, V. L., Schwartz, R. G., Morr, M. L., Kessler, K. L., and Kurtzberg, D. (2000). Deviant Neurophysiological Asymmetry in Children with Language Impairment. *Neuroreport* 11 (17), 3715–3718. doi:10.1097/00001756-200011270-00025

Somers, M., Aukes, M. F., Ophoff, R. A., Boks, M. P., Fleer, W., de Visser, K. L., et al. (2015). On the Relationship between Degree of Hand-Preference and Degree of Language Lateralization. *Brain Lang.* 144, 10–15. doi:10.1016/J.BANDL.2015.03.006

Spencer, K. M., Dien, J., and Donchin, E. (2001). Spatiotemporal Analysis of the Late ERP Responses to Deviant Stimuli. *Psychophysiology* 38 (2), 343–358.

Sprouse, J., Wagers, M., and Phillips, C. (2012). A Test of the Relation between Working-Memory Capacity and Syntactic Island Effects. *Language* 88 (1), 82–123. doi:10.1353/lan.2012.0004

Statistica (2017). *TIBCO Software Inc.*

Stavrakaki, S. (2002). "A-bar Movement Constructions in Greek Children with SLI," in *Clinical Linguistics: Theory and Applications in Speech Pathology and Therapy*. Editor E. Fava (John Benjamins), 131–153. doi:10.1075/cilt.227.12sta

Stavrakaki, S. (2001). Comprehension of Reversible Relative Clauses in Specifically Language Impaired and Normally Developing Greek Children. *Brain Lang.* 77 (3), 419–431. doi:10.1006/brln.2000.2412

Steinhauer, K., and Drury, J. E. (2012). On the Early Left-Anterior Negativity (ELAN) in Syntax Studies. *Brain Lang.* 120 (2), 135–162. doi:10.1016/j.bandl.2011.07.001

Stowe, L. A. (1986). Parsing WH-Constructions: Evidence for On-Line gap Location. *Lang. Cogn. Process.* 1 (3), 227–245. doi:10.1080/01690968608407062

Stowe, L. A., Tanenhaus, M. K., and Carlson, G. N. (1991). Filling Gaps On-Line: Use of Lexical and Semantic Information in Sentence Processing. *Lang. Speech* 34 (4), 319–340. doi:10.1177/002383099103400402

Swinney, D., Ford, M., and Bresnan, J. (1989). "On the Temporal Course of gap-filling and Antecedent Assignment during Sentence Comprehension," in *Language Structure and Processing*. Editors B. Grosz, R. Kaplan, M. Macken, and I. Sag (Stanford, CA: CSLI).

Tataryn, D. J., Wood, J. M., and Gorsuch, R. L. (1999). Setting the Value of K in Promax: A Monte Carlo Study. *Educ. Psychol. Meas.* 59 (3), 384–391. doi:10.1177/00131649921969938

Trueswell, J. C., and Gleitman, L. (2007). "Learning to Parse and its Implications for Language Acquisition," in *Oxford Handbook of Psycholinguistics*. Editor G. Gaskell. doi:10.1093/oxfordhb/9780198568971.013.0039

van der Lely, H. K. J. (2005). Domain-specific Cognitive Systems: Insight from Grammatical-SLI. *Trends Cogn. Sci.* 9 (2), 53–59. doi:10.1016/j.tics.2004.12.002

van der Lely, H. K. J., and Pinker, S. (2014). The Biological Basis of Language: Insight from Developmental Grammatical Impairments. *Trends Cogn. Sci.* 18 (11), 586–595. doi:10.1016/j.tics.2014.07.001

Wagers, M. W., and Phillips, C. (2014). Going the Distance: Memory and Control Processes in Active Dependency Construction. *Q. J. Exp. Psychol.* 67 (0), 1274–1304. doi:10.1080/17470218.2013.858363

Weismer, S. E. (1996). Capacity Limitations in Working Memory. *Top. Lang. Disord.* 17, 33–44. doi:10.1097/00011363-199611000-00005

Weismer, S. E., and Thordardottir, E. T. (1996). *Verbal Working Memory Abilities of School-Age Children with and without Language Impairment*, 2.

Ye, Z., Luo, Y.-j., Friederici, A. D., and Zhou, X. (2006). Semantic and Syntactic Processing in Chinese Sentence Comprehension: Evidence from Event-Related Potentials. *Brain Res.* 1071 (1), 186–196. doi:10.1016/j.brainres.2005.11.085

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

APPENDIX

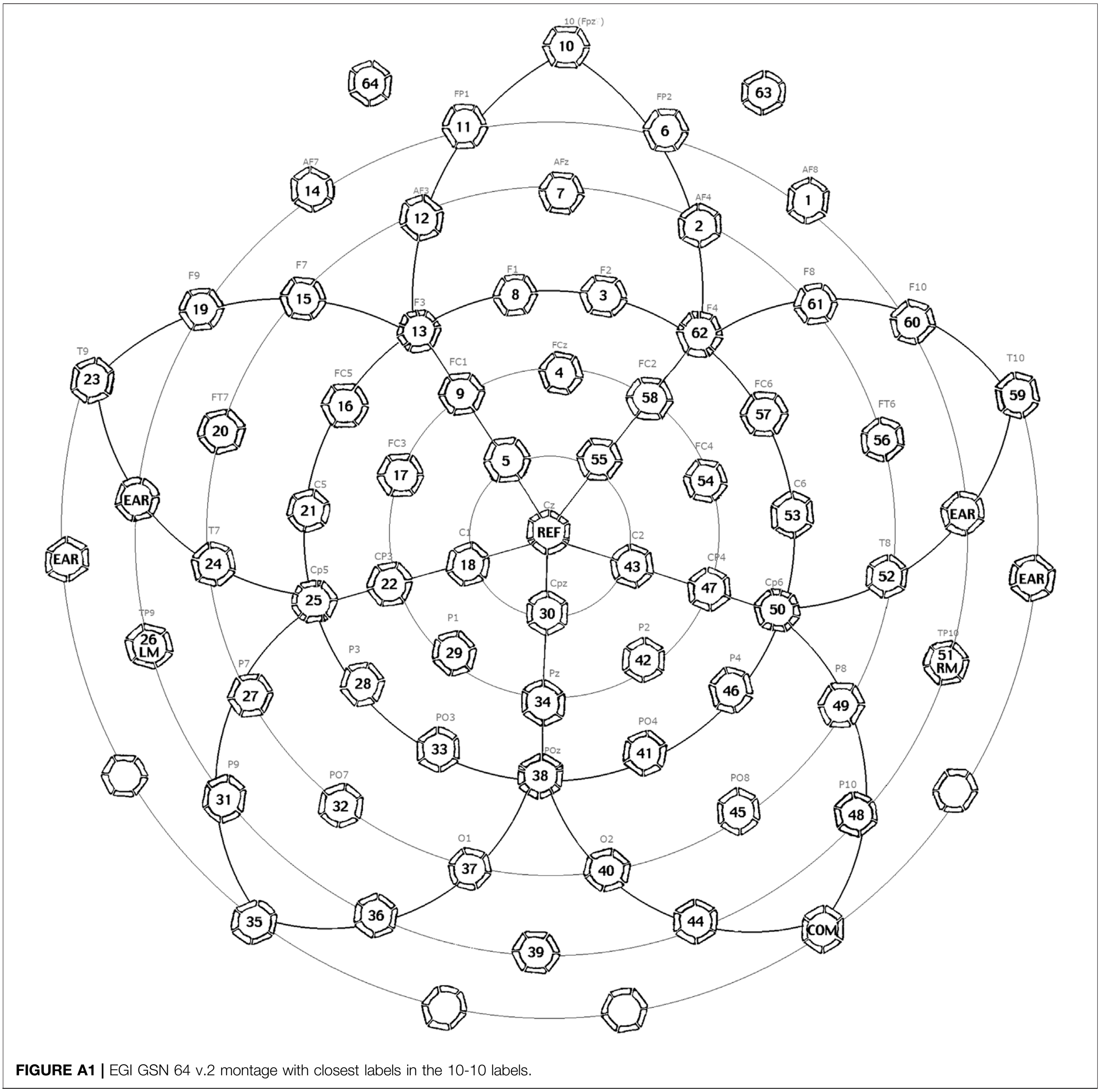


FIGURE A1 | EGI GSN 64 v.2 montage with closest labels in the 10-10 labels.

frontiers | Frontiers in Communication

CORRECTION
published: 03 May 2022
doi: 10.3389/fcomm.2022.901133

Corrigendum: Developmental Language Disorder as Syntactic Prediction Impairment

Arild Hestvik[1]*, *Baila Epstein*[2], *Richard G. Schwartz*[3] *and Valerie L. Shafer*[3]

[1] Department of Linguistics and Cognitive Science, University of Delaware, Newark, DE, United States, [2] Communication Arts, Sciences, and Disorders, Brooklyn College, Boylan Hall, Brooklyn, NY, United States, [3] PhD Program in Speech-Language-Hearing Sciences, The Graduate Center, City University of New York, New York, NY, United States

Keywords: syntax, gap-filling, prediction, event-related potentials, developmental language disorder, relative clauses, sentence processing

OPEN ACCESS

Approved by:
Frontiers Editorial Office, Frontiers Media SA, Switzerland

Correspondence:
Arild Hestvik
hestvik@udel.edu

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Communication

Received: *21 March 2022*
Accepted: *22 March 2022*
Published: *03 May 2022*

Citation:
Hestvik A, Epstein B, Schwartz RG and Shafer VL (2022) Corrigendum: Developmental Language Disorder as Syntactic Prediction Impairment. Front. Commun. 7:901133. doi: 10.3389/fcomm.2022.901133

A Corrigendum on

Developmental Language Disorder as Syntactic Prediction Impairment
by Hestvik, A., Epstein, B., Schwartz, R. G., and Shafer, V. L. (2022). Front. Commun. 6:637585. doi: 10.3389/fcomm.2021.637585

In the original article, there was a mistake in **Figure 10** as published. The order of the labels on the horizonal axis was “filled gap” then “control”. The correct order should be “control” and then “filled gap.” The corrected **Figure 10** appears below.

The authors apologize for this error and state that this does not change the scientific conclusions of the article in any way. The original article has been updated.

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

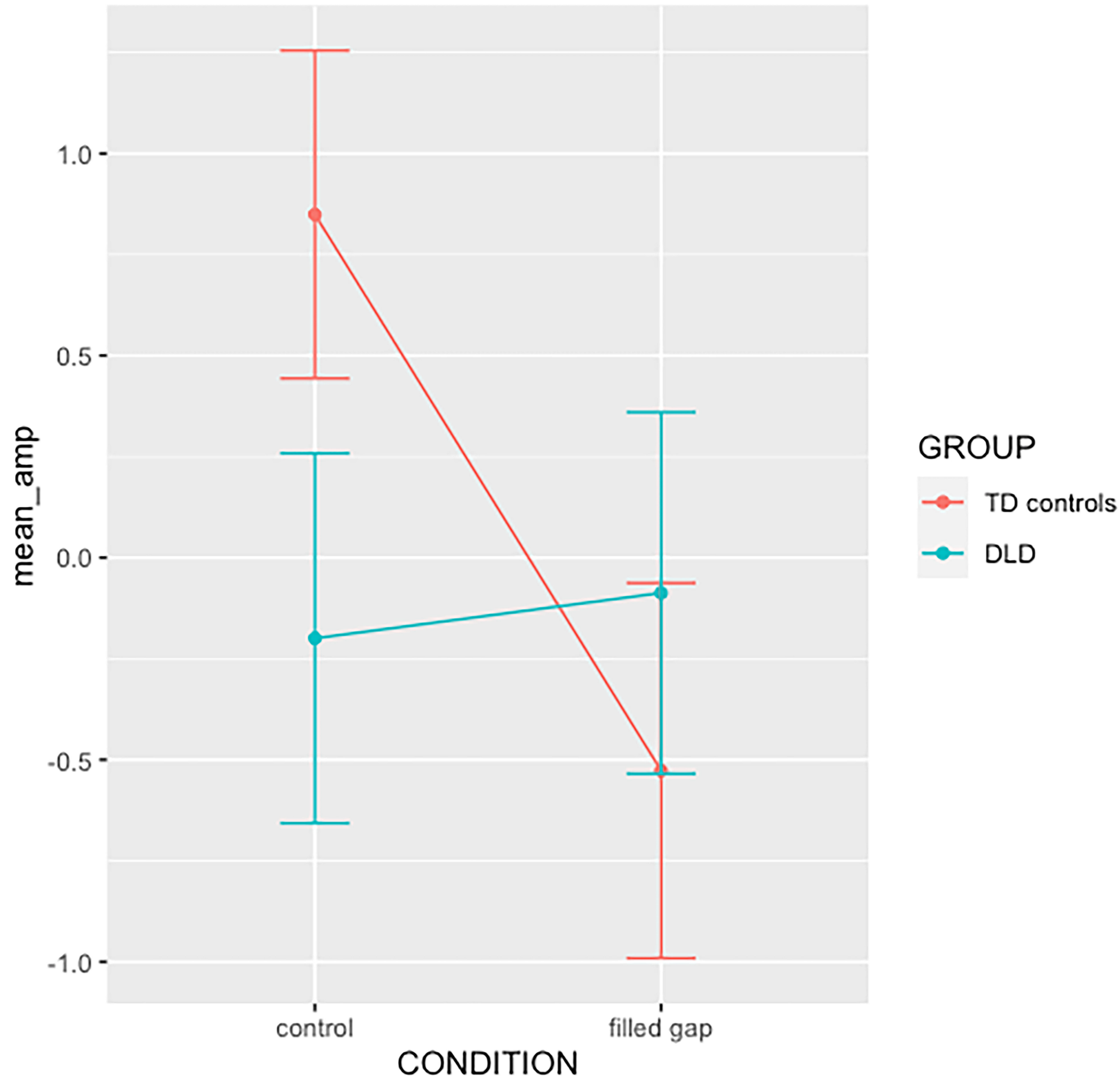


FIGURE 10 | Interaction plot for the 2x2 design GROUP x Condition interaction in TF04SF1. Error bar indicate standard error.

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 26 August 2022
DOI 10.3389/fpsyg.2022.817516

Check for updates

OPEN ACCESS

EDITED BY
Chia-Ying Lee,
Academia Sinica, Taiwan

REVIEWED BY
Chun-Hsien Hsu,
National Central University, Taiwan
Matthew Euler,
The University of Utah, United States

*CORRESPONDENCE
Ina Bornkessel-Schlesewsky
ina.bornkessel-schlesewsky@unisa.edu.au

SPECIALTY SECTION
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

RECEIVED 18 November 2021
ACCEPTED 22 July 2022
PUBLISHED 26 August 2022

CITATION
Bornkessel-Schlesewsky I, Sharrad I, Howlett CA, Alday PM, Corcoran AW, Bellan V, Wilkinson E, Kliegl R, Lewis RL, Small SL and Schlesewsky M (2022) Rapid adaptation of predictive models during language comprehension: Aperiodic EEG slope, individual alpha frequency and idea density modulate individual differences in real-time model updating. *Front. Psychol.* 13:817516. doi: 10.3389/fpsyg.2022.817516

Rapid adaptation of predictive models during language comprehension: Aperiodic EEG slope, individual alpha frequency and idea density modulate individual differences in real-time model updating

Ina Bornkessel-Schlesewsky[1]*, Isabella Sharrad[1], Caitlin A. Howlett[2], Phillip M. Alday[3], Andrew W. Corcoran[4,5], Valeria Bellan[1,2], Erica Wilkinson[2], Reinhold Kliegl[6], Richard L. Lewis[7,8], Steven L. Small[9] and Matthias Schlesewsky[1]

[1]Cognitive Neuroscience Laboratory, Australian Research Centre for Interactive and Virtual Environments, University of South Australia, Adelaide, SA, Australia, [2]Innovation, Implementation and Clinical Translation (IIMPACT) in Health, University of South Australia, Adelaide, SA, Australia, [3]Beacon Biosignals, Boston, MA, United States, [4]Cognition and Philosophy Laboratory, Monash University, Melbourne, VIC, Australia, [5]Monash Centre for Consciousness and Contemplative Studies, Monash University, Melbourne, VIC, Australia, [6]Division of Training and Movement Science, University of Potsdam, Potsdam, Germany, [7]Department of Psychology, University of Michigan, Ann Arbor, MI, United States, [8]Weinberg Institute for Cognitive Science, University of Michigan, Ann Arbor, MI, United States, [9]School of Behavioral and Brain Sciences, University of Texas at Dallas, Dallas, TX, United States

Predictive coding provides a compelling, unified theory of neural information processing, including for language. However, there is insufficient understanding of how predictive models adapt to changing contextual and environmental demands and the extent to which such adaptive processes differ between individuals. Here, we used electroencephalography (EEG) to track prediction error responses during a naturalistic language processing paradigm. In Experiment 1, 45 native speakers of English listened to a series of short passages. Via a speaker manipulation, we introduced changing intra-experimental adjective order probabilities for two-adjective noun phrases embedded within the passages and investigated whether prediction error responses adapt to reflect these intra-experimental predictive contingencies. To this end, we calculated a novel measure of speaker-based, intra-experimental surprisal ("speaker-based surprisal") as defined on a trial-by-trial basis and by clustering together adjectives with a similar meaning. N400 amplitude at the position of the critical second adjective was used as an outcome measure of prediction error. Results showed that N400 responses attuned to speaker-based surprisal over the course of the experiment, thus indicating that listeners rapidly adapt their predictive models to reflect local environmental contingencies (here: the probability of one type of adjective following another when uttered by a particular speaker). Strikingly, this occurs in spite of the wealth of prior linguistic experience that participants bring to the laboratory. Model adaptation effects were strongest for participants with a steep aperiodic (1/f) slope in resting EEG and low individual alpha frequency (IAF), with idea density (ID) showing a more complex pattern. These results were replicated in a separate sample of 40 participants in Experiment 2, which employed a highly similar design to Experiment 1. Overall, our results suggest that individuals with a steep aperiodic slope adapt their predictive models most strongly to context-specific probabilistic information. Steep aperiodic slope is thought to reflect low neural noise, which in turn may be associated with higher neural gain control and better cognitive control. Individuals with a steep aperiodic slope may thus be able to more effectively and dynamically reconfigure their prediction-related neural networks to meet current task demands. We conclude that predictive mechanisms in language are highly malleable and dynamic, reflecting both the affordances of the present environment as well as intrinsic information processing capabilities of the individual.

KEYWORDS

language comprehension, predictive coding, precision, EEG, N400, aperiodic slope, idea density, individual alpha frequency (IAF)

1. Introduction

Predictive coding (e.g., Friston, 2005, 2009) provides a compelling theory of how the human brain processes information. Within a unified account of sensation, cognition and action (e.g., Clark, 2013), it posits that the brain utilizes generative predictive models to actively infer the causes of its sensory inputs. In other words, perception involves the brain using its internal model of the world to generate predictions about expected upcoming sensory input, which are then compared to the actual incoming sensory signals. In line with the "Bayesian brain hypothesis" (e.g., Knill and Pouget, 2004; Frith, 2007; Sanborn and Chater, 2016), this is viewed as a process of (unconscious) probabilistic inference: the prior belief arising from a probabilistic generative model is combined with the sensory evidence to yield a posterior belief (the updated model). Predictions flow from higher to lower levels of a hierarchically organized cortical architecture (via feedback connections) and prediction errors are propagated up the cortical hierarchy (via feedforward connections) to engender model updates at higher levels. While predictions at "lower" levels pertain directly to specific properties of the incoming sensory information, predictions at higher levels are more abstract and can span longer timescales (Hohwy, 2013). In this highly efficient coding scheme, sensory information need only be represented to the extent that it is not predicted (Rao and Ballard, 1999). In other words, prediction errors serve as a proxy for sensory information (Feldman and Friston, 2010; Clark, 2013)[1]. This effectively amounts to signal compression as only the non-predicted parts of the signal need to be transmitted. Overall, the architecture strives to minimize prediction errors.

Crucially, the relative weighting of a prediction error (PE) vis-à-vis the top-down predictive model depends both on the noisiness of the signal (Clark, 2013) and the (un)certainty of the prediction (Feldman and Friston, 2010; Vilares and Kording, 2011). This is known as precision weighting: precision, which is defined as the inverse of variance, reflects the confidence or certainty associated with a belief or a sensory input (Friston, 2009; Feldman and Friston, 2010; Adams et al., 2013). For example, when the sensory evidence conflicts with a prior belief, the degree to which the prior will be shifted toward the sensory evidence in forming the posterior belief depends on the certainty vested in the sensory signal (for a useful illustration, see Figure 1 in Adams et al., 2013). Thus, high-precision (i.e., low uncertainty) prediction errors are associated with higher gain (Friston, 2009) and consequently have a more substantial impact on model updating.

In addition, previous work suggests that the top-down/bottom-up balance changes across the lifespan (Moran et al., 2014) and in non-neurotypical populations (e.g., schizophrenia; see Fletcher and Frith, 2009; Adams et al., 2013). Moran et al. (2014) show that older adults tend to weight model predictions more strongly than younger adults. This means that, when faced with unpredicted sensory input, older adults will attribute higher precision to prior beliefs vis-à-vis the sensory evidence and thereby show a lower rate of learning or model adaptation than younger adults. Moran and colleagues suggest that this protects against the overfitting of internal models to the input, thus resulting in

1 For an alternative proposal, see, for example, Spratling (2008).

less complex models. For positive symptoms of schizophrenia (hallucinations and delusions), by contrast, Fletcher and Frith (2009) suggest that these "are caused by an abnormality in the brain's inferencing mechanisms, such that new evidence (including sensations) is not properly integrated, leading to false prediction errors" (p.56). Using simulations, Adams et al. (2013) show that this can be understood as resulting from less precise top-down predictions, thus "rendering everything relatively surprising" (p.13), including sensations that should not be e.g., self-generated actions; see Clark (2015) for detailed discussion.

These observations suggest that different weightings of top-down (prior) and bottom-up (sensory evidence) information can be a source of individual differences in sensory processing/perceptual inference, specifically in regard to how individuals from different populations adapt their predictive models to changing environmental contingencies. With the present study, we aimed to examine whether such inter-individual differences can also be observed in young, healthy adults (i.e., within the population most typically examined in cognitive neuroscience experiments). We used language as a test domain in which to examine this hypothesis. As a means of studying model adaptation, we investigated individual differences in the extent to which language-related brain responses (the N400 event-related potential) adapt to context-specific probabilistic information ("surprisal") as determined by the experimental environment. In the following, we will first introduce prediction-related phenomena in language and how these can be couched within the predictive coding framework, before turning to a discussion of potential predictors for individual differences in predictive language processing. Finally, we introduce the present study and our hypotheses.

1.1. Prediction and predictive coding in language

Language involves a plethora of predictable information sources across a range of different levels. Here, we focus mostly on the sentence level, as this is the level of interest to the current study. When words are combined into sentences, inter-word dependencies give rise to predictability in various ways. For examples, see the Supplementary materials. Note that we use predictability here rather than prediction to make clear that we are referring to the probabilistic dependencies within the structure of language rather than any putative processing mechanisms; for overviews of probabilistic modeling in psycholinguistics, see, for example, Jurafsky (2003) and Chater and Manning (2006). Experience-based, probabilistic information sources—for example that a determiner (e.g., "the") will at some point be followed by a noun (e.g., "apple") - can be used as priors within a predictive coding architecture. This type of approach has been implemented in computational models of language processing focusing on surprisal or other information-theoretic notions (e.g., Hale, 2006; Levy, 2008); for a recent review, see Hale (2016). The notion of surprisal, which reflects how unexpected a word is given the context in which it appears, is closely related to that of prediction errors in predictive coding. Given a sequence of words $w_1, w_2, \ldots, w_t$, the surprisal of word w_t is defined as the negative logarithm of the probability of that word's occurrence, given the preceding words $w_1, \ldots, w_{t-1}$:

$$surprisal(w_t) = -logP(w_t|w_1, \ldots, w_{t-1})$$

Surprisal has been linked to neurophysiological correlates of language processing, particularly the N400 event-related potential (ERP) component (Frank et al., 2015; Kuperberg, 2016). There have also been explicit attempts to link speech and language processing to predictive coding architectures (e.g., Pickering and Garrod, 2007, 2013; Skipper et al., 2007; Poeppel et al., 2008; Rauschecker and Scott, 2009; Bornkessel-Schlesewsky et al., 2015b). In addition, several studies suggest that probabilistic information regarding higher-order language-related information is used to anticipate sensory input (Dikker et al., 2010; Dikker and Pylkkänen, 2011), a finding which is closely in line with the assumptions of the predictive coding framework.

Nevertheless, prediction as a concept has remained controversial in the cognitive neuroscience of language processing, particularly with regard to the N400; see Kuperberg and Jaeger (2016) for arguments in favor, and Van Petten and Luka (2012) for arguments against. One of the arguments most often used against active prediction—i.e., prediction that goes beyond the preactivation of a word through a semantic network (or similar) and specifically the explicit prediction of a single specific word—is that there is little evidence that N400 amplitude reflects the error signal resulting from a failed prediction. Rather, N400 amplitude appears to be attenuated with increasing predictability. According to Van Petten and Luka (2012), "current data suggest only that N400 amplitudes are reduced in the presence of supportive semantic context and provide little hint that amplitudes are increased when a hypothesis/expectation/prediction is disconfirmed. From our starting premise that predictions should generate both benefits and costs (on different occasions), the apparent absence of costs is problematic" (p.180). They view this as evidence that the N400 reflects (passive) preactivation rather than (active) prediction, with prediction manifesting itself in other ERP components, most notably late positivities with a frontal scalp distribution.

We contend, however, that this pattern of results for the N400 is, in fact, fully in line with the assumptions of a predictive coding model. Recall that, in the typical implementation of this type of model, only error signals are transmitted *via* feedforward connections because predictable sensory input is "canceled out" by top-down activity encoding the relevant predictions. Thus, a reduced signal is transmitted when the input is, to some extent, predictable. By contrast, in the absence of any predictability, the complete sensory information associated with an input item, say a word, needs to be conveyed: an entirely unpredicted/unpredictable word is associated with the largest prediction error signal. When prior context leads to a certain degree of predictability (or preactivation), prediction error is reduced. In this way, we see the attenuation of prediction errors for predicted vs. unpredictable input rather than an increased error signal for a prediction violation (again, in comparison to a context without any predictability). The pattern of N400 effects thus exactly mirrors what one would expect to observe under typical implementations of a predictive coding architecture (for detailed discussion, see Bornkessel-Schlesewsky and Schlesewsky, 2019). Indeed, predictive coding neatly accounts for the well-known observation that N400 amplitude decreases for unexpected words that match the expected word in regard to certain features (e.g., semantic category, Federmeier and Kutas, 1999) or that show a certain degree of form overlap with the expected word (e.g., *via* orthographic neighborhood, Laszlo and Federmeier, 2009, 2011). In these cases, some—but not all—aspects of the incoming input are explained away by the generative predictive model, thereby resulting in an error signal that is intermediary between that for a highly predictable item and an unpredictable item that does not share any features with the most expected continuation. This suggests that the N400 is a composite response that combines error signals at different levels; cf. Bornkessel-Schlesewsky and Schlesewsky (2013), Bornkessel-Schlesewsky and Schlesewsky (2019), and Frank and Willems (2017).

Bornkessel-Schlesewsky and Schlesewsky (2019) proposed that, more specifically, the N400 reflects a precision-weighted error signal. This account builds on the extensive literature linking the mismatch negativity (MMN) to prediction error processing in the auditory domain (e.g., Friston, 2005; Garrido et al., 2009; Moran et al., 2014) and, more specifically, to precision-weighted error responses (Todd et al., 2011, 2013, 2014). By varying the temporal stability of rules underlying the structure of sound sequences, Todd and colleagues showed that prediction-error-related MMN effects respond to the perceived salience of events and that this is influenced both by rule stability and by rule primacy (i.e., which rule was learned first). Bornkessel-Schlesewsky and Schlesewsky (2019) argue that the N400 reflects similar processes but for more complex stimuli—hence its longer latency in comparison to the MMN.

The claim that N400 amplitude correlates with a precision-weighted error signal is supported by several observations. Firstly, N400 effects vary across languages depending on the informativity of a particular feature (e.g., animacy) for sentence-level interpretation in that language (Bornkessel-Schlesewsky and Schlesewsky, 2019, 2020). This provides a natural link to precision weighting: recall that precision is defined as the inverse of variance and variance in the form-to-meaning mapping is clearly reduced for features that are highly informative (cf. work in the context of the Competition Model, e.g., Bates et al., 1982, 2001; MacWhinney et al., 1984). Secondly, N400 amplitude shows a further property that is expected in the context of a precision-weighted error signal account, namely a modulation by attention. As described in detail by Feldman and Friston (2010), selective attention increases the precision associated with an upcoming sensory stimulus. This can lead to an amplification of the prediction error signal. At a microcircuit level, prediction error amplification is thought to be implemented *via* an increased gain of error-encoding units (most likely pyramidal cells in higher cortical layers; cf. Bastos et al., 2012). Similarly, though acknowledging the vastly different level of measurement at play here, N400 amplitude for incongruent (unpredictable) vs. congruent (more predictable) words within a sentence is increased when the attentional focus on a word is increased *via* information structural (focus) and prosodic (accent) information (Wang et al., 2011).

1.2. Precision-weighting as a source of inter-individual differences in predictive coding and possible predictors for individual differences in language

We have already sketched out above how precision weighting of prediction errors not only serves to dynamically adapt a predictive coding architecture to the estimated uncertainties of prior expectations and sensory stimuli, but also how such an architecture provides a natural locus for inter-individual differences (e.g., in aging or, in a different manner, in schizophrenia) and that these are measurable using the MMN ERP component. On the basis of the claims by Bornkessel-Schlesewsky and Schlesewsky (2019) about the functional similarity of the MMN and N400, we would also hypothesize the presence of such differences in N400 effects during language processing. Moreover, given that precision weighting of priors and sensory information may plausibly differ between individuals, we will examine whether such differences manifest themselves even in a population typically considered to be relatively homogeneous, namely young healthy adults. In the following, we will introduce the three main measures that we used as predictors of individual differences in the current study: Idea Density, Individual Alpha Frequency and Aperiodic (1/f) Activity.

1.2.1. Idea density

Idea Density (ID; also known as Propositional Density or P-Density: Kintsch and Keenan, 1973) measures the number of ideas expressed relative to the total number of words used, as derived from written or oral text samples. Ideas are operationalised as predicates: for example, verbs, adjectives and negations are all counted as ideas. ID is thought to reflect the efficiency of linguistic information encoding (Cheung and Kemper, 1992; Kemper et al., 2001b; Iacono et al., 2009; Engelman et al., 2010; Farias et al., 2012) and longitudinal evidence shows that ID measures collected from young adults predict cognitive performance in older adulthood (Snowdon et al., 1996). As discussed by Kemper et al. (2001b), ID is not correlated with high school English or maths grades nor with level of educational attainment (see also Ferguson et al., 2014; Spencer et al., 2015). Kemper and colleagues suggest that "low P-Density in young adulthood may reflect suboptimal neurocognitive development, which, in turn, may increase susceptibility to age-related decline due to Alzheimer's or other diseases" (Kemper et al., 2001a, p.602). ID is relatively stable across the adult lifespan but declines in older adulthood (for results from a large-scale study involving texts from over 19,000 respondents, see Ferguson et al., 2014).

Given the link between ID and efficiency of linguistic information encoding, we hypothesized that ID may provide a proxy for the quality of an individual's language model—our rationale being that efficient encoding requires high-quality linguistic representations. If this is indeed the case, high-ID individuals will have a higher precision language model than low-ID individuals and may thus weight model-based predictions more strongly than unexpected input information in the case of a prediction error. This could entail that high-ID individuals adapt their predictive language models more slowly to local contextual affordances than low-ID individuals, in a similar manner to the slower model updating by older adults reported by Moran et al. (2014).

1.2.2. Individual alpha frequency

Evidence is accruing that perception and cognition are discrete rather than continuous (VanRullen, 2016). We perceive the world by discretely sampling sensory input. In the brain, sampling corresponds to oscillations: fluctuations between states of high and low neuronal receptivity, which are coordinated between neurons and neural assemblies to optimize communication between them (Buzsáki and Draguhn, 2004; Fries, 2005). Importantly, the speed of oscillatory activity differs between individuals. In particular, the peak frequency of the dominant alpha rhythm of the human EEG (∼8–13 Hz) varies between approximately 9 and 11.5 Hz in young adults (Klimesch, 1999). This variation in *individual alpha frequency* (IAF) is a trait-like characteristic (Grandy et al., 2013b), which shows high heritability (Posthuma et al., 2001; Smit et al., 2006) and test-retest reliability (Gasser et al., 1985; Kondacs and Szabó, 1999). IAF variability has ramifications not only for the alpha band, but also for the adjacent theta (∼4–7 Hz) and beta (∼15–30 Hz) rhythms. Consequently, IAF determines an individual's sensory sampling rate and this has consequences for the resolution with which sensory input is analyzed and represented. Samaha and Postle (2015) recently reported a compelling demonstration of this relation for the visual modality. They presented participants with two visual flashes in rapid succession and manipulated the inter-stimulus interval (ISI) between them. At very short ISIs, the two visual stimuli fuse into a single percept. Crucially, inter-individual variability in the two-flash-fusion-threshold was correlated with IAF; for a related demonstration of IAF being *causally* related to the length of the temporal window within which multimodal stimuli are integrated with one another, see Cecere et al. (2015).

In addition to correlating with the resolution of sensory sampling, IAF is associated with a range of higher cognitive abilities. High-IAF individuals process information more quickly (Surwillo, 1961, 1963), and perform better on memory tasks (Klimesch, 1999) and general intelligence measures (*g*) (Grandy et al., 2013a). For a different result see Ociepka et al. (2022), who found a relationship between IAF and processing speed but not between IAF and general intelligence. IAF decreases with age from young adulthood onwards (Köpruner et al., 1984; Klimesch, 1999), thus accompanying the well-known decline of many cognitive abilities in older adulthood (e.g., Hedden and Gabrieli, 2004; Salthouse, 2011). Previous work also indicates that language processing and language learning strategies differ between high- and low-IAF individuals (Bornkessel et al., 2004; Bornkessel-Schlesewsky et al., 2015a; Kurthen et al., 2020; Nalaye et al., 2022).

On account of its link to the rate of sensory sampling, we hypothesized that IAF may serve as a proxy for the general quality (i.e., resolution, signal-to-noise ratio) of the sensory input, which, in turn, influences more complex aspects of information processing. If true, this would mean that incoming sensory information is associated with a higher precision for high-IAF individuals in comparison to low-IAF individuals. In the case of a prediction error, high-IAF individuals may thus weight unexpected input information more strongly vis-à-vis model predictions than low-IAF individuals. Consequently, high-IAF individuals may adapt their predictive language models more quickly to local contextual affordances than low-IAF individuals.

1.2.3. Aperiodic (1/f) activity

Complementing the examination of individual differences in oscillatory neural activity (e.g., *via* IAF), a growing body of literature has begun to investigate the possible role of individual differences in non-oscillatory (aperiodic) brain activity. Aperiodic activity follows a $P \sim 1/f^{\beta}$ power law (He, 2014), where P corresponds to power, f to frequency and β is the so-called "power-law exponent." This overall relationship of lower frequencies in the human EEG being associated with higher amplitudes (power) than higher frequencies has long been recognized. Only more recently, however, has it become clear that the power law exponent parameter—which governs the steepness of the power decrease with increasing frequency—changes dynamically depending on a variety of factors including age and task, as well as an individual's cognitive state (e.g., He, 2014; Voytek et al., 2015; Donoghue et al., 2020). In addition to potentially being clinically relevant (He, 2014), this variability may also reveal individual differences in cognitive processing in healthy individuals. For example, Ouyang et al. (2020) reported that, when both aperiodic (1/f) slope and alpha activity were taken into account, aperiodic slope rather than alpha activity predicted individual differences in processing speed for an object recognition task. These authors thus suggest that previous observations of an association between alpha activity and processing speed may have been due to a confound between oscillatory and aperiodic activity in earlier analyses (cf. also Donoghue et al., 2020). In the domain of language processing, Dave et al. (2018) recently observed a modulation of prediction-related N400 effects by 1/f slope such that a steeper slope predicted more pronounced N400 effects. Further, Cross et al. (2022) found that the learning of certain types of grammatical rules in an artificial language is likewise predicted by inter-individual variability in 1/f slope.

Regarding potential mechanisms underlying the effects of aperiodic slope on cognitive processing, one prominent approach posits that steepness of the aperiodic slope reflects the degree of neural noise (Voytek et al., 2015). Specifically, highly synchronous neural spiking (equated with "lower neural noise") is thought to correlate with a steeper 1/f slope, while more asynchronous or aberrant firing (equated with "higher neural noise") is associated with a flatter slope (Buzsáki et al., 2012; Voytek and Knight, 2015). This notion of neural noise may, in turn, be associated with the balance between excitatory and inhibitory activity within neural networks (e.g., Gao et al., 2017). As Voytek et al. (2015) show, aging is associated with a flattening of the 1/f slope and this physiological change may underlie effects of cognitive aging such as a slowing of processing speed.

It is important to acknowledge that, in the context of aperiodic activity estimates obtained from scalp EEG, any inferences drawn about individual differences in neural noise are indirect and must be viewed with a certain degree of caution. Nevertheless, we believe that the existing literature supports an association between scalp-recorded aperiodic slope estimates and neural noise, albeit indirectly. Freeman and Zhai (2009) successfully simulated 1/f slopes obtained from intracranial EEG *via* a computational model of mutual excitation among pyramidal cells. They concluded that "variation in the observed slope is attributed to variation in the level of the background activity that is homeostatically regulated by the refractory periods of the excitatory neurons" (Freeman and Zhai, 2009, p.97). Voytek et al. (2015) in turn demonstrated that 1/f slope and age show a similar relationship in intracranial and scalp EEG measures, thus supporting the association between scalp-recorded 1/f slope and neural noise.

In the context of the current study, we will examine the proposal by Dave et al. (2018) that more synchronous neural networks—as reflected in a steeper aperiodic slope—are associated with stronger predictive processing. If this proposal holds, we should observe a stronger reliance on top-down predictive models for individuals with a steeper 1/f slope and, consequently, a potentially slower adaptation of internal predictive models to current contextual affordances than for individuals with a shallower 1/f slope.

1.3. The present study

The present study examined how ID, IAF and aperiodic activity are related to prediction error signals in language processing. In Experiment 1, participants listened to 150 short passages (approximately 5 sentences in length) while their EEG was recorded. An example passage is given below:

> *Example of the passages presented to participants in the current study:*
> Florence was enjoying her long-awaited holiday in Singapore with her close friends. One of the activities she was most looking forward to was visiting the zoo, where she had the opportunity to ride a **huge gray elephant**. Although standing in the **warm humid air** was dreadful, being waved to through the enclosure by the zookeeper brought a smile to her face.

The critical passages (60%, i.e., 90 of 150) each contained 2 two-adjective noun phrases (marked in bold in the example above), which could either have an expected (canonical) or unexpected (non-canonical) order (e.g., canonical: "the huge gray elephant"; non-canonical: "the gray huge elephant"; for seminal work on ERP correlates of adjective order variations, see Kemmerer et al., 2007). With this manipulation, we intended to elicit prediction error responses due to the unexpectedness of the non-canonical adjective orders. In addition, we varied the probability of encountering non-canonical adjective orders by means of a speaker manipulation. Specifically, passages were recorded by two male speakers with varying probabilities of canonical orders. Thus, for the "canonical" speaker, approximately 70% of the critical 180 two-adjective noun phrases were presented to participants in canonical order, while for the "non-canonical" speaker, only approximately 30% of adjectives were canonically ordered.

Building on the proposal that N400 amplitude reflects precision-weighted prediction error signals (Bornkessel-Schlesewsky and Schlesewsky, 2019), our primary outcome variable was the amplitude of the N400 event-related potential at the position of the critical second adjective within the two-adjective noun phrases embedded in our passages.

Through our experimental design, we aimed to examine inter-individual differences in the processing of prediction errors elicited by the non-canonical adjective orders. We used IAF, ID and aperiodic activity (1/f) as our primary predictors of individual differences as outlined above but also collected an additional battery of cognitive and linguistic tests (see the Methods section for further details). Furthermore, we included the speaker manipulation as an additional manipulation of prediction precision. Here, our rationale was that the high number of non-canonical adjective orders produced by the non-canonical speaker would call for adaptation of participants' existing language model, according to which a non-canonical order of two adjectives should be unexpected (cf. the notion of "active listening" put forward by Friston et al., 2021). Participants who adapt more quickly to the contingencies of the current input— i.e., more readily adapt their established predictive model in the face of prediction errors—should thus be expected to show N400 responses aligned with the experimental environment rather than their global language experience. As described above, we tentatively hypothesized that this readiness to adapt might be more pronounced in high-IAF and low-ID individuals on account of the high precision of the sensory input or low precision of the predictive language model, respectively. Individuals with a steep 1/f slope were expected to show a similar pattern to individuals with a high ID (i.e., slower model adaptation) on account of the link that has been postulated between lower neural noise (associated with a steeper 1/f slope) and stronger predictive processes (Dave et al., 2018). In spite of these hypotheses, this was an exploratory study given the complexity of the domain under examination and the fact that this research question has not yet been examined to date—neither in the area of language nor with respect to other cognitive domains.

Given the novelty of the research question, we also report a follow-up experiment with a similar experimental design (Experiment 2), in which we examined whether the results of Experiment 1 could be replicated.

2. Experiment 1

2.1. Methods

2.1.1. Participants

Forty-five young adults (31 female; mean age: 22.9 years, sd: 3.9, range: 18–33) participated in Experiment 1. Participants were right-handed as assessed by the Edinburgh handedness inventory (Oldfield, 1971), native speakers of English who had not learnt another language prior to starting school. They reported having no diagnosis of neurological or psychiatric conditions, normal hearing and normal or corrected-to-normal vision. The experimental protocol was approved by the University of South Australia's Human Research Ethics Committee (protocol number 36348).

2.1.2. Materials

The critical materials for this experiment were 90 short passages (approximately 5 sentences in length), each of which contained two critical two-adjective noun phrases (NPs; e.g., "a huge gray elephant"). Critical NPs occurred at different positions in each passage so that their occurrence would not be predictable. The order of the prenominal adjectives was manipulated such that, in some cases, they adhered to the expected sequence of "value > size > dimension > various physical properties > color" (Kemmerer et al., 2007, p.240). We will refer to adjective orders adhering to this sequencing as canonical (C) in what follows and to those that do not as non-canonical (N). Passages were recorded by two male speakers of Australian English with the probability of adjectives in the critical NPs occurring in a canonical or a non-canonical order manipulated across speakers. Thus, when listening to the passages, participants were exposed to one speaker (henceforth: the canonical speaker) who produced more canonical than non-canonical orders (C:N ratio of 69%:31%) and another speaker (henceforth: the non-canonical speaker) who produced more non-canonical orders (C:N ratio of 31%:69%). To counterbalance the assignment of speakers to passages, we constructed two versions of the critical materials. Thus, canonicity of speaker varied both within subjects and within items, but the (non-canonical vs. canonical) speaker assignment was fixed throughout the course of each session. The distribution of canonical and non-canonical orders across speakers, versions and passages is shown in Table 1.

Each participant listened to the critical passages from one of the two versions interspersed with 60 filler passages in a pseudo-randomized order. The filler passages included a separate experimental manipulation involving passive sentences and relative clauses and did not contain any two-adjective noun phrases. Thus, every participant was presented with 150 passages in total.

To ensure that participants were listening attentively, they were presented with yes-no comprehension questions after approximately 1/3 of passages. An example comprehension question for the passage example above is: "Did the zookeeper wave at Florence?" (correct answer = yes).

2.1.3. Language models

The principal aim of the present study was to examine how individuals differ in the adaptation of their predictive models to the current environment during language processing. To this end, we focused on the processing of the second adjective (ADJ2) in the critical 2-adjective NPs embedded in the passages. We used bigram-based surprisal to quantify predictability of ADJ2

TABLE 1 Counterbalancing of canonical and non-canonical adjective orders across versions.

Version	Speaker	Passages	Orders within passage
1	Canonical	1–45	1–7: NC 8–14: CN 15–21: NN 22–45: CC
1	Non-canonical	46–90	46-52: CN 53–59: NC 60-66: CC 67–90: NN
2	Canonical	46–90	46–52: NC 53–59: CN 60–66: NN 67–90: CC
2	Non-canonical	1–45	1–7: CN 8–14: NC 15–21: CC 22–45: NN

Within each version, the canonical speaker produced 62 canonical and 28 non-canonical adjective orders (CN ratio = 31:69%), while the non-canonical speaker produced 62 non-canonical and 28 canonical adjective orders (CN ratio = 69:31%). Note that passages were presented in a pseudo-randomized order and interspersed with filler passages; i.e., the passage numbers shown here do not reflect the order of presentation. Abbreviations for orders within passages: C = canonical; N = non-canonical, with the first letter referring to the first two-adjective NP within the passage and the second letter referring to the second two-adjective NP.

in the context of the preceding adjective. To allow us to estimate predictability at the level of adjective classes, we first established adjective clusters for our materials. This was accomplished using the following procedure, which was implemented in R (R Core Team, 2021) using the `tidyverse` (Wickham et al., 2019) and `tidymodels` (Kuhn and Wickham, 2020) collections of packages as well as the packages `tidytext` (Silge and Robinson, 2016) and `widyr` (Robinson, 2021). For package version numbers, please see the analysis scripts provided with the raw data (see Data Availability Statement).

Procedure for determining adjective clusters and calculating cluster-based surprisal:

1. We used pre-derived word vectors from van Paridon and Thompson (2021) to determine similarities between adjectives. Word vectors, also known as word embeddings, provide a numerical representation of word meaning. They are created by machine learning models, which learn lexical relationships from word co-occurrences in large text corpora. For a recent example of how word vectors may serve as useful representations of word meaning when investigating human language processing, see Pereira et al. (2018). Here, we used Van Paridon and Thompson's top 1 million vectors from a combined Wikipedia and Open Subtitles corpus.
2. To reduce dimensionality, we performed a principal components analysis (PCA), thus reducing the 300 vectors from van Paridon and Thompson (2021) to 5 principal components (PCs).
3. Six adjective clusters were identified on the basis of the PCs using k-means clustering. The value of k=6 was selected *via* visual inspection of the total within-cluster sum of squares. Three of the six clusters are visualized in Figure 1 and a full list is provided in the Supplementary materials for Experiment 1.
4. Cluster-based unigram and bigram frequencies were computed as cluster-based sums of unigram and bigram counts from the Open Subtitles corpus for English (751 million words) as made available by van Paridon and Thompson (2021). From these, surprisal values for adjective 2 (ADJ2) in the context of adjective 1 (ADJ1) were calculated as:

$$surprisal(ADJ2) = -log(\frac{ClusterBigramFrequency(ADJ1ADJ2)}{ClusterUnigramFrequency(ADJ1)}$$

Here, *ClusterBigramFrequency(ADJ1ADJ2)* refers to the frequency with which two-adjective bigrams comprising a first adjective belonging to the cluster of ADJ1 and a second adjective belonging to the cluster of ADJ2 occur in the Open Subtitles corpus. *ClusterUnigramFrequency(ADJ1)* refers to the frequency with which adjectives belonging to the cluster of ADJ1 occur in Open Subtitle corpus. In the remainder of the paper, we will refer to these corpus-based surprisal values as **global surprisal**.

In a second step, we computed *incremental surprisal* for ADJ2 within the experimental context to be able to track how listeners' expectations change as a function of being exposed to the experimental environment. To track surprisal incrementally over the course of the experiment, we calculated the NP-by-NP cumulative intra-experimental frequencies for the ADJ1-ADJ2 bigram cluster and the ADJ1 unigram cluster and then computed surprisal as described above. This was done separately for each speaker, thus allowing us to examine to what extent participants' expectations adapted to the distributional properties of each of the two speakers within the experiment. We henceforth refer to this speaker-based measure of intra-experimental surprisal as **speaker-based surprisal**. Using speaker-based surprisal, we aimed to examine how participants' N400 responses—as an assumed proxy for precision-weighted prediction error signals— were modulated by the exposure to adjective order variations throughout the course of the experiment and by each speaker.

Corpus-based word (unigram) frequencies for ADJ2 were included in all analyzes as a control variable. These were taken from the same unigram corpus as used for global surprisal

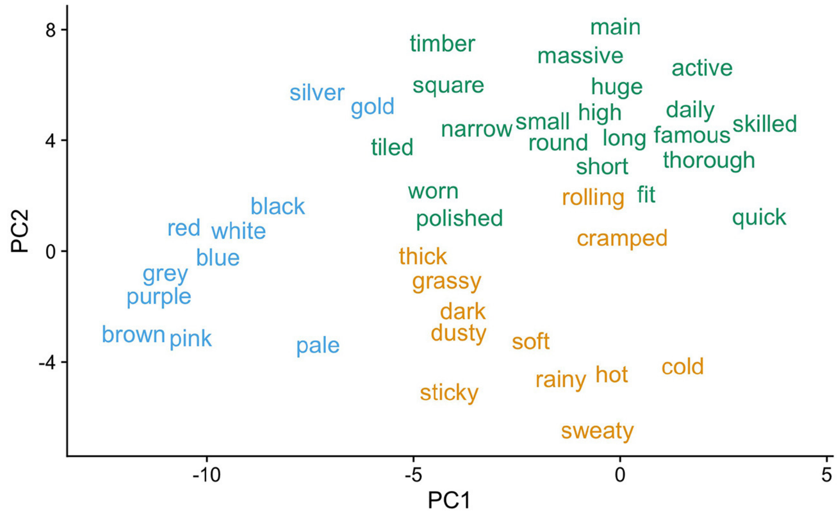


FIGURE 1
Three adjective clusters produced by the current clustering procedure for Experiment 1. Clusters are visualized with regard to their variability on principal components PC1 and PC2. Note, for example, how the clustering procedure distinguishes color adjectives from other adjective types.

calculation above and log-transformed prior to inclusion in the analysis.

2.1.4. Behavioral individual differences measures

2.1.4.1. Idea density (ID)

Participants provided a written text sample in response to the prompt "Describe your favorite game." This corresponds to the Essay Composition task of the Wechsler Individual Achievement Test—Australian and New Zealand Standardized, Third Edition (WIAT-III A&NZ; Pearson Clinical). From this text, we calculated ID using the automated Computerized Propositional Idea Density Rater (CPIDR; Brown et al., 2008).

2.1.4.2. Cognitive tests

Participants completed an additional battery of cognitive tests. These included:

- The two-subtest version of the Wechsler Abbreviated Scale of Intelligence—Second Edition (WASI-II; Pearson Clinical), comprising Vocabulary and Matrix reasoning tasks
- Three additional language-related subtests from the WIAT-III, namely Oral Word Fluency, Sentence Repetition and Sentence Composition
- A reading-span task (Daneman and Carpenter, 1980)

In accordance with our hypotheses, we focus on ID and the resting state EEG-based individual differences metrics (1/f slope and Individual Alpha Frequency; see below) as our primary measures of individual differences for the purposes of the present study.

2.1.5. Procedure

Participants completed two in-lab testing sessions: (1) a behavioral session comprising the cognitive tests/text sample production, and (2) an EEG session comprising the collection of resting-state EEG recordings as well as the main language comprehension task. Sessions were either completed on the same day, separated by a break (approximately 30 min), or on 2 days (with the second session completed within 7 days of the first session).

2.1.5.1. Behavioral session

In the behavioral session, after the consent process, participants completed a questionnaire to provide demographic, language and well-being details. They subsequently completed the cognitive tests as described above. The behavioral session took maximally 1.5 h to complete.

2.1.5.2. EEG session

In the EEG session, participants were fitted with an EEG cap and underwent a 2-min eyes-open and 2-min eyes-closed resting state EEG recording prior to commencing the main task. For the main task, each trial commenced with the 500 ms presentation of a fixation asterisk in the center of a computer screen, after which the auditory presentation of a passage commenced *via* loudspeakers. After the auditory passage was complete, the fixation asterisk remained on screen for another 500 ms. Subsequently, participants were presented with a comprehension question in approximately 1/3 of all trials, to which they responded with "yes" or "no" using two buttons on a game controller. The assignment of "yes" and "no" responses to the left and right controller buttons was counterbalanced across participants and the maximal response time was set at 4,000 ms. For trials without a comprehension question, participants were asked to "Press the YES key to proceed." Following the participant's response or after the allocated response time had elapsed, the next trial commenced after an inter-trial interval of 1,500 ms. Participants were asked to avoid any movements or blinks during the presentation of the fixation asterisk if possible.

Note that, as the intermittent comprehension questions only served to ensure that participants listened attentively, comprehension data was not analyzed in the present paper. Log files for the comprehension task are, however, provided with the raw data for the experiment (see Data Availability statement).

The 150 passages were presented in 5 blocks, between which participants took short self-paced breaks. Prior to commencing the main task, participants completed a short practice session. After the main task, the resting state recordings were repeated. Overall, the EEG session took approximately 3 h including electrode preparation and participant clean-up.

2.1.6. EEG recording and preprocessing

The EEG was recorded from 64 electrodes mounted inside an elastic cap (Quik-CapEEG) using a Neuroscan Synamps2 amplifier (Compumedics Neuroscan, Abbotsford, VIC, Australia). The electrooculogram (EOG) was recorded via electrodes placed at the outer canthi of both eyes as well as above and below the left eye. The EEG recording was sampled at 1,000 Hz and referenced to the right mastoid.

Data preprocessing was undertaken using MNE Python version 0.23.0 (Gramfort et al., 2013, 2014). EEG data were re-referenced to an average reference and downsampled to 500 Hz prior to further processing. EOG-artifacts were corrected using an ICA-based correction procedure, with independent components (ICs) found to correlate most strongly with EOG events (via the `create_eog_epochs` function in MNE) excluded. Raw data were filtered using a 0.1—30 Hz bandpass filter to exclude slow signal drifts and high frequency noise. Epochs were extracted in a time window from –200 to 1,000 ms relative to critical word (ADJ2) onset and mean single-trial amplitudes were extracted for the prestimulus (–200 to 0 ms) and N400 (300–500 ms) time windows using the `retrieve` function from the philistine Python package (Alday, 2018).

2.1.6.1. Resting-state EEG-based individual differences measures: Individual alpha frequency (IAF) and aperiodic (1/f) activity

IAF and aperiodic slope estimates were calculated from participants' eyes-closed resting-state recordings.

To calculate IAF, we used a Python-based implementation (Alday, 2018) of the procedure described in Corcoran et al. (2018) and drawing on electrodes P1, Pz, P2, PO3, POz, PO4, O1, Oz and O2. We estimated both peak alpha frequency (PAF) and center of gravity (COG) measures (cf. Corcoran et al., 2018, for discussion) and calculated the mean of pre and post estimates by participant for each measure. For participants who did not have estimable IAF values for one of the two recording sessions, their IAF estimate from the other session was used as their overall IAF metric. This was the case for 3 participants in Experiment 1.

Aperiodic (1/f) intercept and slope estimates were calculated in Python using the YASA toolbox (Vallat and Walker, 2021). YASA implements the irregular-resampling auto-spectral analysis (IRASA) method for separating oscillatory and aperiodic activity (Wen and Liu, 2016). As for IAF, by-participant intercept and slope estimates were computed as means of pre and post resting-state recordings from electrodes F7, F5, F3, F1, Fz, F2, F4, F6, F8, FT7, FC5, FC3, FC1, FCz, FC2, FC4, FC6, FT8, T7, C5, C3, C1, Cz, C2, C4, C6, T8, TP7, CP5, CP3, CP1, CPz, CP2, CP4, CP6, TP8, P7, P5, P3, P1, Pz, P2, P4, P6, P8, PO7, PO5, PO3, POz, PO4, PO6, PO8, O1, Oz, and O2.

2.1.7. Data analysis

The data analysis was undertaken using R (R Core Team, 2021) and Julia (Bezanson et al., 2017). We used R for data pre- and post-processing. For data import and manipulation, we used the `tidyverse` collection of packages (Wickham et al., 2019) as well as the `vroom` package (Hester and Wickham, 2021). Figures were created using `ggplot2` (Wickham, 2016; Wickham et al., 2021) as well as the packages `cowplot` (Wilke, 2021) and `patchwork` (Pedersen, 2020). All figures with color employ the Okabe Ito color palette from `colorblindr` (McWhite and Wilke, 2021). Other packages used include `corrr` (Kuhn et al., 2020), `kableExtra` (Zhu, 2021) and `here` (Müller, 2020). For package version numbers, please see the analysis scripts provided with the raw data (see Data Availability Statement). For R, see the html outputs in the src/subdirectory; for Julia see the Manifest.toml file.

EEG data were analyzed using linear mixed effects models (LMMs) with the `MixedModels.jl` package in Julia (Bates et al., 2021). We used the JellyMe4 package (Alday, 2021) to move model objects from Julia to R for visualization purposes.

For the ERP data, we examined single-trial N400 amplitude as our outcome variable of interest. To this end, we analyzed mean EEG voltage 300–500 ms post onset of the critical second adjective (ADJ2) in a centro-parietal region of interest (C3, C1, Cz, C2, C4, P3, P1, Pz, P2, P4, CP3, CP1, CPz, CP2, CP4).

2.1.7.1. Linear mixed modeling approach

We adopted a parsimonious LMM selection approach (Bates et al., 2015; Matuschek et al., 2017), which seeks to identify LMMs that are supported by the data and not overparameterized. Model selection was undertaken without consideration of fixed-effects estimates (i.e., without consideration of which fixed effects reached significance).

Fixed effects initially included log-transformed unigram frequency, speaker-based surprisal, adjective order canonicity, epoch (as a proxy for how long participants had been exposed to the experimental stimuli), mean prestimulus amplitude and their interactions. Prestimulus amplitude (–200 to 0 ms) was included as a predictor in the model as an alternative to traditional EEG baselining (see Alday, 2019). The categorical factor canonicity was encoded using sum contrasts (cf. Schad et al., 2020; Brehm and Alday, 2022); thus, model intercepts represent the grand mean. All continuous predictors were z-transformed prior to being included in the models.

Although not of interest within the scope of the current paper, we modeled the main effect of prestimulus amplitude with a second-order and the main effect of speaker-based surprisal with a third-order polynomial trend. The inclusion of these higher-order trends was supported by the data, significantly improved model fit, and guarded against the interpretation of spurious interactions of their linear trends with other fixed effects (Matuschek and Kliegl, 2018). Non-significant higher-order interactions involving fixed effects were removed from the model when they were not part of the theoretical expectations and this did not lead to a significant reduction in goodness of model fit as assessed *via* likelihood-ratio tests (LRTs).

The random-effect (RE) structure was selected in two steps, again using LRTs to check improvement in goodness of fit and random-effects PCA (rePCA) to guard against overparameterization during model selection. The results of the first step led to a RE structure with variance components for grand means, prestimulus amplitude and prestimulus amplitude (2nd order) by subject, item and channel. In a second step, we added by-subject and by-item variance components for effects of canonicity, epoch, unigram frequency and speaker-based surprisal to the RE structure. Correlation parameters were not significant for the by-subject and by-item variance components and constrained to zero.

Using the speaker-based surprisal LMM (as described above) as a reference, we added, in turn, fixed-effect covariates for individual differences in (1) 1/f slope, (2) IAF (peak alpha frequency), and (3) ID to the model to check the extent to which they moderate/modulate adaptation to speaker-based surprisal. In each of these three additional LMMs, adding the respective individual differences covariate as a by-item variance component significantly improved the goodness of model fit.

The model selection procedure is transparently documented in Julia scripts in the Open Science Framework repository for this paper (see Data Availability Statement).

2.1.7.2. Reporting and visualization of results

As our primary research question was how listeners adapt their predictive models to the experimental context, we focus on interactions of speaker-based surprisal and epoch in the interpretation of our results. Thus, for each LMM, we focus on the highest order interaction(s) including these predictors and the current individual-differences predictor of interest where relevant. These are reported, visualized and interpreted in the main text. Model summaries are included in the Supplementary materials, with only significant effects reported in the model summary tables to increase readability. For full model summaries including all effects, see the repository for the paper. For the visualization of effects, we used the `broom.mixed` package (Bolker and Robinson, 2021) to extract fitted values and the `remef` package (Hohenstein and Kliegl, 2021) to extract partial effects. By visualizing partial effects, we focus on the effects of interest while adjusting for additional model parameters that are not of primary interest here where appropriate.

2.2. Results

2.2.1. Individual differences measures

Distributions of the (z-transformed) individual differences measures are shown in Supplementary Figure S1.

2.2.2. EEG data

2.2.2.1. Sanity check analysis

In a first step, we ran a "sanity check" analysis to determine whether the current data showed expected modulations of N400 amplitude by unigram frequency and global (corpus-based) surprisal defined at the level of adjective clusters (see section on language models above). For this, we followed the general modeling strategy outlined in the Data analysis section above, but including global surprisal rather than speaker-based surprisal.

The sanity check analysis confirmed the expected effects of word frequency and surprisal on N400 amplitude. At the position of the critical second adjective, N400 amplitudes were higher for words with a lower frequency of occurrence and for words with higher corpus-based surprisal values. These effects are visualized in Figure 2 (see Supplementary Table S1 for the model summary). As is apparent from the model summary, there was a significant interaction of Unigram Frequency x Global Surprisal x Prestimulus amplitude (Estimate = 0.0497, Std. Error = 0.0203, $z = 2.45$, $p = 0.01$). However, as we were only interested in general trends for word frequency and global surprisal for the purposes of our sanity check, we visualize the partial effects of these two predictors adjusted for the other predictors.

2.2.2.2. N400 amplitude attunes to speaker-based surprisal over the course of the experiment

The speaker-based surprisal model (see Supplementary Table S2 for the model summary) revealed an interaction of Speaker-based Surprisal x Epoch x Canonicity x Prestimulus Amplitude (Estimate $= -0.0693$, Std. Error $= 0.0173$, $z = -4.01$, $p < 0.0001$). Figure 3 visualizes the partial effect of Speaker-based Surprisal x Epoch x Canonicity, adjusted for Prestimulus Amplitude. As is apparent from the figure, the effect of speaker-based surprisal becomes stronger over the course of the experiment, i.e., the longer participants are exposed to the peculiarities of each speaker, the stronger the effect of speaker-based surprisal on N400 amplitude. This supports our assumption that listeners attune their internal predictive models to the current context. Strikingly, the effect of speaker-based surprisal overrides the effect of adjective order canonicity by the end of the experiment [cf. Alday et al. (2017) for the finding that language-related EEG responses adapt to the local context within a story].

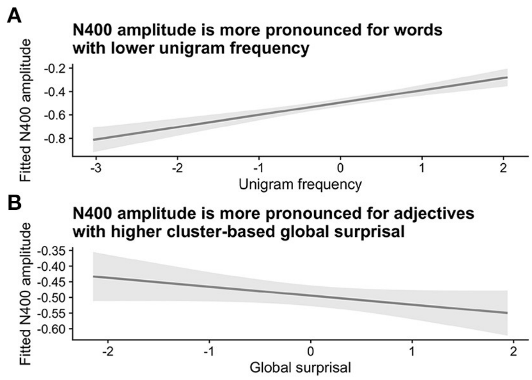


FIGURE 2
Sanity check analysis for Experiment 1. Panel A shows the relationship between N400 amplitude and (log-transformed) unigram frequency, while Panel B shows the relationship between N400 amplitude and global (corpus-based) surprisal, as defined using bigrams at the adjective cluster level. Both unigram frequency and surprisal values were z-transformed. Shaded areas indicate 95% confidence intervals.

2.2.2.3. Inter-individual differences in predictive model adaptation

Having determined that effects of speaker-based surprisal (z-transformed) on N400 amplitude became stronger over the course of the experiment, we next sought to examine how individuals differed with regard to this adaptation process and which of our metrics best predicted these assumed individual differences. To this end, we in turn added each of our individual differences metrics of interest—individual alpha frequency (IAF), aperiodic (1/f) slope and idea density (ID)—to the speaker-based surprisal model without individual differences. As revealed by likelihood ratio tests and goodness-of-fit metrics, all of these models showed an improved fit to the data over the base model without an individual differences predictor. Table 2 provides an overview of the goodness-of-fit metrics, demonstrating that all models including individual differences

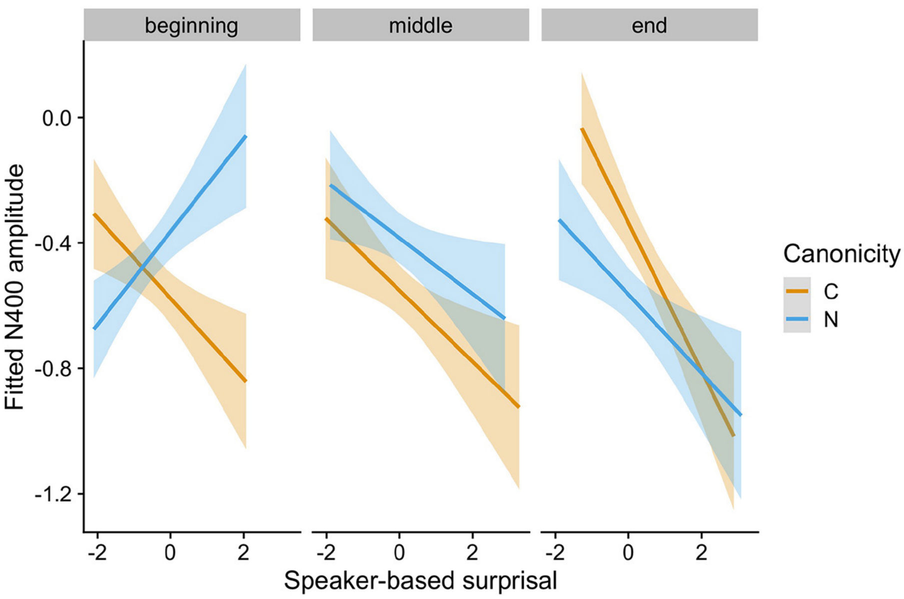


FIGURE 3
Changes in the relationship between speaker-based surprisal (z-transformed) and N400 amplitude over the course of Experiment 1 for canonical (C) and non-canonical (N) adjective orders. The figure visualizes partial effects as calculated using the *remef* package, adjusted for prestimulus amplitude. Note that position in the experiment (operationalised *via* epoch in the statistical model) is trichotomised into beginning, middle and end for visualization purposes only; epoch was included in the model as a continuous predictor.

TABLE 2 Model comparison for the models including speaker-based surprisal in Experiment 1.

	Model	dof	Deviance	AIC	AICc	BIC
1	No Ind. differences	54	513,348	513,456	513,456	513,965
2	Slope	89	512,745	512,923	512,924	513,761
3	IAF	83	513,232	513,398	513,398	514,179
4	ID	84	512,520	512,688	512,688	513,479

The table shows goodness-of-fit metrics for the best-fitting model without individual differences in comparison to the models including aperiodic (1/f) slope, IAF, and ID, respectively. Note that the differing degrees of freedom for the individual-differences models are a result of the pruning of non-significant higher-order interactions from some models; see Data Analysis section for details.

covariates outperform the model without individual differences in terms of AIC. With the exception of the IAF model, this also holds for BIC.

In line with our primary research question, for the interpretation of the individual differences results, we focus on the top-level interaction(s) involving Speaker-based Surprisal, Epoch and the individual differences predictor of interest (cf. the discussion of our LMM modeling approach in the Data Analysis section).

For the model including aperiodic slope, the top-level interaction was Prestimulus Amplitude x Speaker-based Surprisal x Epoch x Canonicity x Frequency x Slope (Estimate = 0.1531, Std. Error = 0.0780, $z = 1.96$, $p < 0.05$). For the model including IAF, it was Speaker-based Surprisal x Epoch x Canonicity x Frequency x IAF (Estimate = −0.0505, Std. Error = 0.0172, $z = -2.94$, $p < 0.01$). The ID model showed an interaction of Prestimulus Amplitude x Speaker-based Surprisal x Epoch x Canonicity x ID (Estimate = 0.0821, Std. Error = 0.0163, $z = 5.03$, $p < 0.0001$). In view of the complexity of these models and the fact that our primary interest for the purposes of the present paper lies in examining how adaptation to speaker-based surprisal is modulated by these individual differences metrics, we visualize partial effects of Speaker-based Surprisal x Epoch x Individual Differences Covariate of Interest for each model in turn in the following, adjusting for any additional moderating effects. For model summaries, see Supplementary Tables S3–S5.

Figure 4 visualizes how the intra-experimental adaptation to speaker-based surprisal is modulated by aperiodic slope. It demonstrates that, though N400 responses had attuned to speaker-based surprisal for all participants by the end of the experiments (mirroring the effects observed in Figure 3), individuals with a steep aperiodic slope adapt most rapidly to intra-experimental contingencies (cf. the pattern of N400 responses in the middle portion of the experiment).

Figures 5, 6 show the adaptation to speaker-based surprisal as moderated by IAF and ID, respectively. For IAF, it is apparent that adaptation is quickest for individuals with a low IAF. At a first glance, the pattern is similar for ID, i.e., low-ID individuals show a more rapid adaptation to speaker-based surprisal. However, it is notable that individuals with a high ID show the most pronounced change in the *pattern* of speaker-based surprisal N400 effects over the course of the experiment, demonstrating a slight “anti surprisal" effect at the beginning of the experiment but adapting to show the expected attunement to speaker-based surprisal by the end.

2.3. Discussion

Experiment 1 examined N400 ERP responses to investigate how, during naturalistic language processing, individuals update their internal predictive models to reflect current contextual or environmental information. While listening to short passages recorded by two speakers of Australian English, participants showed an adaptation to experiment- and speaker-specific adjective order patterns with increasing exposure to these patterns over the course of the experiment. By the end of the experiment, N400 responses at the position of the critical second adjective (ADJ2) in two-adjective noun phrases embedded in the passages had attuned to speaker-based surprisal. In other words: N400 amplitude reflected the (information-theoretic) surprisal for encountering an adjective of type ADJ2 following an adjective of the type encountered at the ADJ1 position, given the speaker reading the passage. Adjective type was defined using a word-vector-based clustering procedure and speaker-based surprisal was defined incrementally *via* the participant’s prior exposure to two-adjective noun phrases for a particular speaker at each point over the course of the experiment. N400 attunement to speaker-based surprisal led to an alignment of N400 amplitudes for canonical and non-canonical adjective orders by the end of the experiment. It is important to keep in mind, however, that these measures (i.e., adjective clusters and surprisal) were correlations rather than experimental manipulations.

In addition, we observed inter-individual differences in regard to the strength of N400-attunement to speaker-based surprisal. All three individual differences predictors examined—aperiodic (1/f) slope, Individual Alpha Frequency (IAF) and Idea Density (ID)—led to improvement of mixed model fit over the best model not including individual differences predictors. Individuals with a steep aperiodic slope, which is thought to reflect low neural noise, showed the most pronounced and earliest attunement to speaker-based surprisal. A similar pattern was observed for individuals with a low IAF. For ID, the pattern was somewhat more mixed: while low-ID individuals appeared to show an earlier attunement to speaker-based surprisal, high-ID individuals showed a more substantial change of speaker-surprisal-related response from the beginning to the end of the experiment. These findings were examined further in Experiment 2.

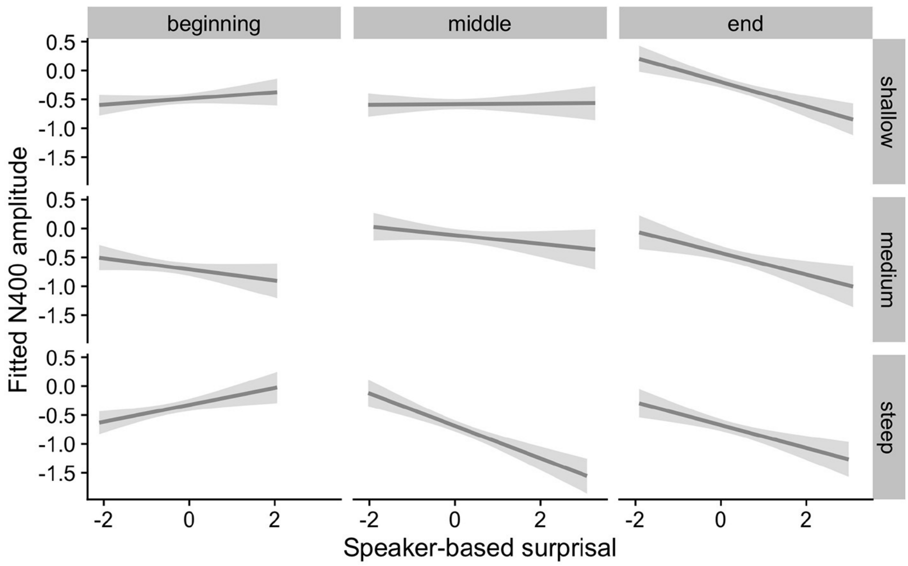


FIGURE 4
Effect of aperiodic (1/f) slope on changes in the relationship between speaker-based surprisal (z-transformed) and N400 amplitude over the course of Experiment 1. The figure visualizes partial effects as calculated using the *remef* package, adjusted for prestimulus amplitude, canonicity and frequency. Note that position in the experiment (operationalised *via* epoch in the statistical model) is trichotomised (into beginning, middle, end) for visualization purposes only; epoch was included in the model as a continuous predictor. The same holds for 1/f slope, which is trichotomised into steep, medium and shallow for visualization purposes but was entered into the statistical model as a continuous predictor. Shaded areas indicate 95% confidence intervals.

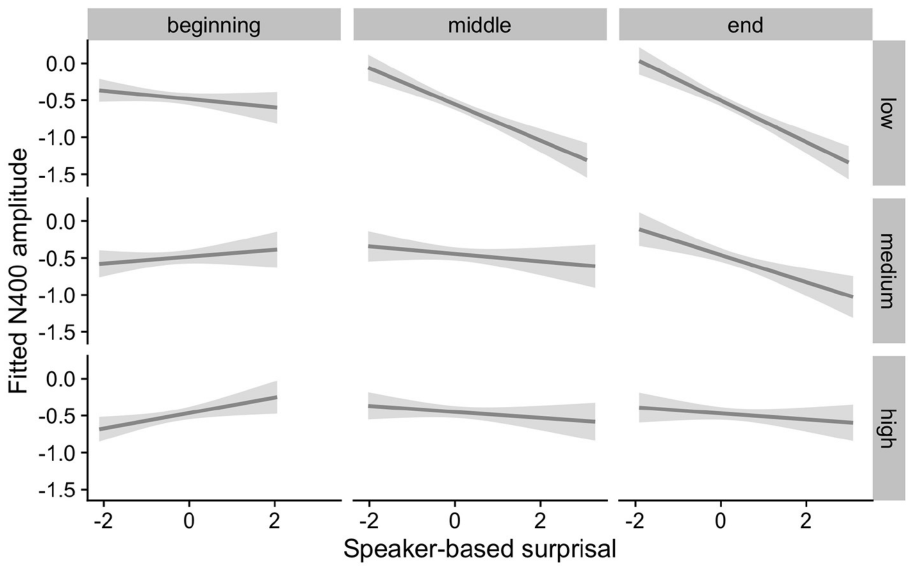


FIGURE 5
Effect of individual alpha frequency (IAF) on changes in the relationship between speaker-based surprisal (z-transformed) and N400 amplitude over the course of Experiment 1. The figure visualizes partial effects as calculated using the *remef* package, adjusted for canonicity and frequency. Note that position in the experiment (operationalised *via* epoch in the statistical model) is trichotomised into beginning, middle and end for visualization purposes only; epoch was included in the model as a continuous predictor. The same holds for IAF, which is trichotomised into low, medium and high for visualization purposes but was entered into the statistical model as a continuous predictor. Shaded areas indicate 95% confidence intervals.

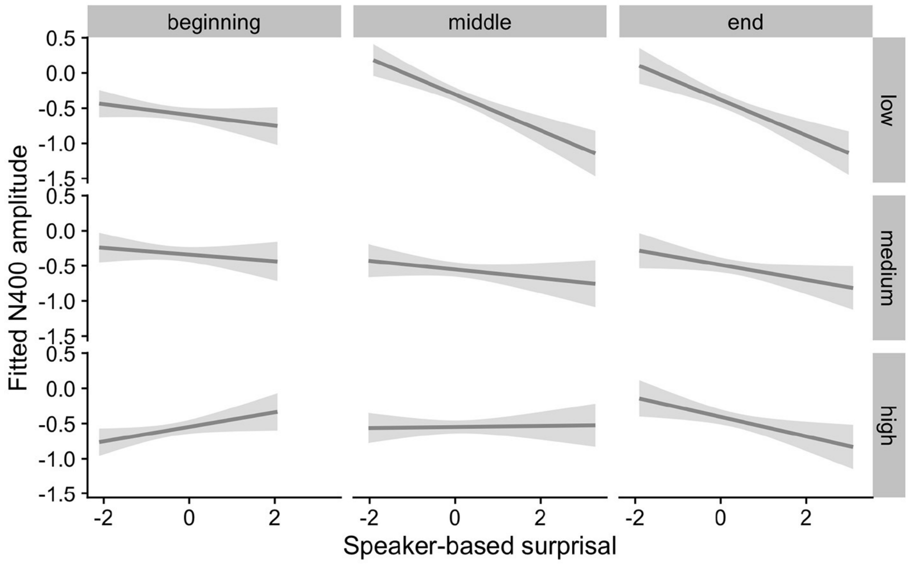


FIGURE 6
Effect of idea density (ID) on changes in the relationship between speaker-based surprisal (z-transformed) and N400 amplitude over the course of Experiment 1. The figure visualizes partial effects as calculated using the *remef* package, adjusted for prestimulus amplitude and canonicity. Note that position in the experiment (operationalised *via* epoch in the statistical model) is trichotomised into thirds (beginning, middle, end) for visualization purposes only; epoch was included in the model as a continuous predictor. The same holds for ID, which is trichotomised into low, medium and high for visualization purposes but was entered into the statistical model as a continuous predictor. Shaded areas indicate 95% confidence intervals.

3. Experiment 2

3.1. Methods

In view of the exploratory nature of the current study and the novel results of Experiment 1, we ran a second Experiment to determine whether these results could be replicated. Experiment 2 employed a very similar design to Experiment 1 with a new sample of young adults as participants.

3.1.1. Participants

Forty young adults (mean age: 23.8 years, sd: 6.3, range: 18–39) participated in Experiment 2, with 30 identifying as female, 9 identifying as male and 1 identifying as other. Inclusion and exclusion criteria were as for Experiment 1 and the experiment was approved under the same protocol by the University of South Australia's Human Research Ethics Committee. None of the participants for Experiment 2 had taken part in Experiment 1.

3.1.2. Materials

Participants again listened to 150 short passages in Experiment 2, which were adapted from those used in Experiment 1. In contrast to Experiment 1, in which only 90 of the 150 passages contained two critical two-adjective NPs, in Experiment 2, all 150 passages contained two critical NPs. This change was incorporated in order to increase the number of critical items per participant and thus improve our ability to track changes in N400 activity across the course of the experiment. In addition, we made minor modifications to some of the critical NPs from Experiment 1. As for Experiment 1, the full experimental materials are available on the study repository (see Data Availability statement).

The passages were again recorded by two male speakers of Australian English, one of which had already been one of the speakers for Experiment 1. As for Experiment 1, one of the speakers (the "canonical speaker") had a higher probability of producing canonical vs. non-canonical two-adjective orders (approximately 70%:30%), while the other (the "non-canonical speaker") had a lower probability of producing canonical vs. non-canonical orders (approximately 30%:70%). The assignment of speaker to the canonical or non-canonical role was counterbalanced across participants. In order to further accentuate the speaker-specific adjective order characteristics, presentation of the two speakers was alternated in a block-based manner in this experiment. The experiment commenced with one block of the canonical speaker, followed by two blocks of the non-canonical speaker and two further blocks of the canonical speaker.

Comprehension questions were again presented after approximately 1/3 of all passages.

3.1.3. Language models

Adjective clusters and speaker-based surprisal were calculated following the same procedure as for Experiment 1. The adjective clusters for Experiment 2 are listed in the Supplementary materials.

3.1.4. Behavioral individual differences measures

3.1.4.1. Idea density

Participants were given 10 min to produce a written text sample of approximately 300 words in response to the prompt "Describe an unexpected event in your life." ID was calculated as in Experiment 1.

3.1.4.2. Cognitive tests

Participants completed an additional battery of cognitive tests. These included:

- The four-subtest version of the Wechsler Abbreviated Scale of Intelligence—Second Edition (WASI-II; Pearson Clinical), comprising Block design, Vocabulary, Matrix reasoning and Similarities tasks
- Three subtests from the Test of Adolescent and Adult Language-Fourth Edition (TOAL-4), namely Word opposites, Derivations and Spoken analogies
- Semantic and phonological verbal fluency tasks
- A computer-based hearing test to measure pure-tone hearing thresholds (pure-tone audiometry)

As for Experiment 1, we focus on ID and the resting state EEG-based individual differences metrics (1/f slope and Individual Alpha Frequency, IAF; see below) as our primary measures of individual differences.

3.1.5. Procedure

The two in-lab testing sessions (behavioral and EEG) for Experiment 2 were comparable to those in Experiment 1. The procedure for the EEG testing session was also identical to that for Experiment 1 with two exceptions. Firstly, participants completed a short (approximately 3.5 min) passive auditory oddball paradigm prior to the main language processing task. This task was included as part of a larger lifespan study and will not be considered here. Secondly, a subset of participants completed two (rather than one) eyes-closed resting state EEG recording sessions both before and after the experiment: one in which they were instructed to relax and one in which they were asked to try to keep their mind blank. For the purposes of calculating resting-state individual difference metrics (IAF and 1/f slope), we used the eyes-closed recordings with the "relax" instructions, as these were comparable to the eyes-closed resting-state recordings with only a single session.

3.1.6. EEG recording and preprocessing

The EEG was recorded from 64 electrodes mounted inside an elastic cap (actiCAP) using a Brain Products actiCHamp amplifier (Brain Products GmbH, Gilching, Germany). The electrooculogram (EOG) was recorded *via* electrodes placed at the outer canthi of both eyes as well as above and below the left eye. The EEG recording was sampled at 500 Hz and referenced to FCz.

Data preprocessing was undertaken as for Experiment 1 with the exception that, as a first step in the preprocessing procedure for Experiment 2, the data were converted to the brain imaging data structure for electroencephalography (EEG-BIDS; Pernet et al., 2019) using the MNE-BIDS Python package (Appelhoff et al., 2019).

3.1.6.1. Resting-state EEG-based individual differences measures: Individual alpha frequency and aperiodic (1/f) activity

IAF and aperiodic slope estimates were calculated as for Experiment 1. Due to slightly differing electrode configurations, there were minor differences in the electrodes used for the IAF and aperiodic activity analyzes in this experiment. The electrodes used for IAF (peak alpha frequency) estimation were: P1, Pz, P2, PO3, POz, PO4, O1, O2. The electrodes used for aperiodic slope estimation were: F7, F3, Fz, F4, F8, FC5, FC1, FC2, FC6, T7, C3, Cz, C4, T8, CP5, CP1, CP2 CP6, P7, P3, Pz, P4, P8, PO9, O1, O2, PO10, AF7, AF8, F5, F1, F2, F6, FT7, FC3, FC4, FT8, C5, C1, C2, C6, TP7, CP3, CPz, CP4, TP8, P5, P1, P2, P6, PO7, PO3, POz, PO4, PO8.

3.1.7. Data analysis

The data analysis was undertaken as for Experiment 1.

As our primary research question for Experiment 2 was whether it is possible to replicate the inter-individual difference effects observed in Experiment 1, we focus on the mixed model analyses examining 1/f slope, IAF and ID and how these modulate the effect of speaker-based surprisal across the course of the experiment.

3.2. Results

3.2.1. Individual differences measures

Distributions of the (z-transformed) individual differences measures are shown in Supplementary Figure S2.

3.2.2. EEG data

For the model including aperiodic slope, the top-level interactions involving Speaker-based Surprisal, Epoch and Slope were Prestimulus Amplitude x Speaker-based Surprisal x Epoch x Canonicity x Slope (Estimate = –0.0421, Std. Error = 0.0173, z = –2.44, $p < 0.02$) and Frequency x Speaker-based Surprisal x Epoch x Canonicity x Slope (Estimate = –0.0472, Std. Error = 0.0210, z = –2.24, $p <$ 0.03).

For the model including IAF, the top-level interaction was Prestimulus Amplitude x Frequency x Speaker-based Surprisal x Epoch x IAF (Estimate = 0.0415, Std. Error = 0.0148, z = 2.80, $p < 0.01$); for the ID model, it was Prestimulus Amplitude x Frequency x Speaker-based Surprisal x Epoch x Canonicity x ID (Estimate = –0.0534, Std. Error = 0.0181, z = –2.95, $p < 0.01$). Model summaries are presented in Supplementary Tables S6–S8.

The effects of interest are visualized in Figures 7–9. As for Experiment 1, we visualize partial effects of Speaker-based Surprisal x Epoch x Individual Differences Covariate of Interest for each model in turn in the following, adjusting for any additional moderating effects.

Overall, the results of Experiment 2 replicate the effects observed in Experiment 1. Individuals with a steep 1/f slope or a low IAF show more pronounced adaptation to speaker-based, intra-experimental probabilistic information over the course of the experiment in comparison to their counterparts with a shallow 1/f slope or a high IAF. By contrast, the pattern for ID is less clear.

3.3. Combined analysis of Experiments 1 and 2

Finally, we conducted a combined analysis of Experiments 1 and 2 in order to examine whether the inter-individual differences of interest would also be observable with a more substantial sample size (n=85). To this end, we again computed the three individual-differences models involving 1/f slope, IAF and ID using the same modeling approach as before. The only exception was the addition of a main effect of Experiment in the fixed effects in order to capture any intrinsic differences in EEG activity between the two experiments (e.g., due to the use of different amplifiers).

For the combined model including aperiodic slope, the top-level interactions involving Speaker-based Surprisal, Epoch and Slope were Prestimulus Amplitude x Frequency x Speaker-based Surprisal x Epoch x Slope (Estimate = 0.0352, Std. Error = 0.0113, z = 3.11, $p < 0.01$) and Prestimulus Amplitude x Speaker-based Surprisal x Epoch x Canonicity x Slope (Estimate = –0.0613, Std. Error = 0.0110, z = -5.58, $p < 0.0001$).

For the model including IAF, the top-level interactions of interest were Prestimulus Amplitude x Frequency x Speaker-based Surprisal x Epoch x IAF (Estimate = 0.0424, Std. Error = 0.0109, z = 3.88, $p < 0.001$) and Frequency x Speaker-based Surprisal x Epoch x Canonicity x IAF (Estimate = –0.0432, Std. Error = 0.0118, z = –3.65, $p < 0.001$). For the ID model, it was Prestimulus Amplitude x Frequency x Speaker-based Surprisal x Epoch x Canonicity x ID (Estimate = –0.0239, Std. Error = 0.0120, z = -1.99, $p < 0.05$). Model summaries are presented in Supplementary Tables S9–S11.

The effects of interest are visualized in Figures 10–12. As for the analysis of Experiments 1 and 2, we visualize partial effects of Speaker-based Surprisal x Epoch x Individual Differences Covariate of Interest for each model in turn in the following, adjusting for any additional moderating effects.

3.4. Discussion

The results of Experiment 2 and the combined analysis of Experiments 1 and 2 broadly support the findings of Experiment 1. The findings for 1/f slope and IAF are highly compatible across all analyses: participants with a steep 1/f slope and those with a low IAF show a more substantial model adaptation to intra-experimental probabilistic information than those with a shallow 1/f slope or a high IAF. The findings for ID are not as clear for the individual analyses of Experiments 1 and 2; however, the combined analysis shows an emerging trend for increased model adaptation over the course of the experiment by individuals with a low ID.

4. General discussion

We have reported two ERP studies designed to investigate inter-individual differences in internal model updating during naturalistic language processing. By means of a novel measure of speaker-based surprisal for adjective orders, we examined the degree to which N400 responses track context-specific probabilistic information tied to the experimental environment. This measure, "speaker-based surprisal", reflects the predictability of adjective type for the second adjective in a two-adjective sequence given the type of the first adjective for a particular speaker. Adjective type was determined in a data-driven manner using a cluster-based analysis of semantic (word-vector-based) similarity between adjectives, and speaker-based probabilities were manipulated by having one speaker utter a higher percentage of expected orders and a second speaker utter a higher percentage of unexpected orders.

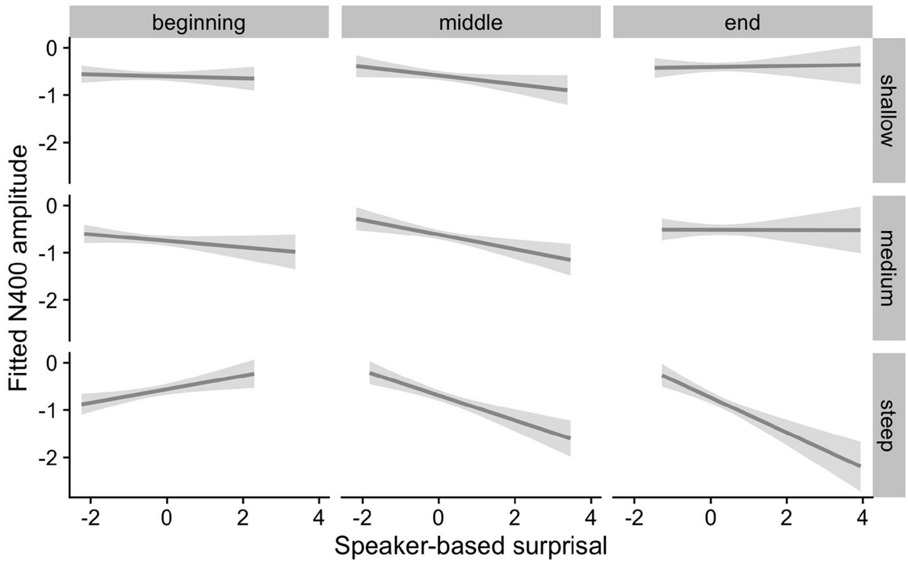


FIGURE 7
Effects of aperiodic (1/f) Slope on changes in the relationship between speaker-based surprisal (z-transformed) and N400 amplitude over the course of Experiment 2. The figure visualizes partial effects as calculated using the *remef* package, adjusted for Prestimulus Amplitude and Canonicity. Note that position in the experiment (operationalised *via* epoch in the statistical model) is trichotomised into beginning, middle and end for visualization purposes only; epoch was included in the model as a continuous predictor. The same holds for the individual differences variables, which are trichotomised for visualization purposes but were entered into the statistical models as a continuous predictors. Shaded areas indicate 95% confidence intervals.

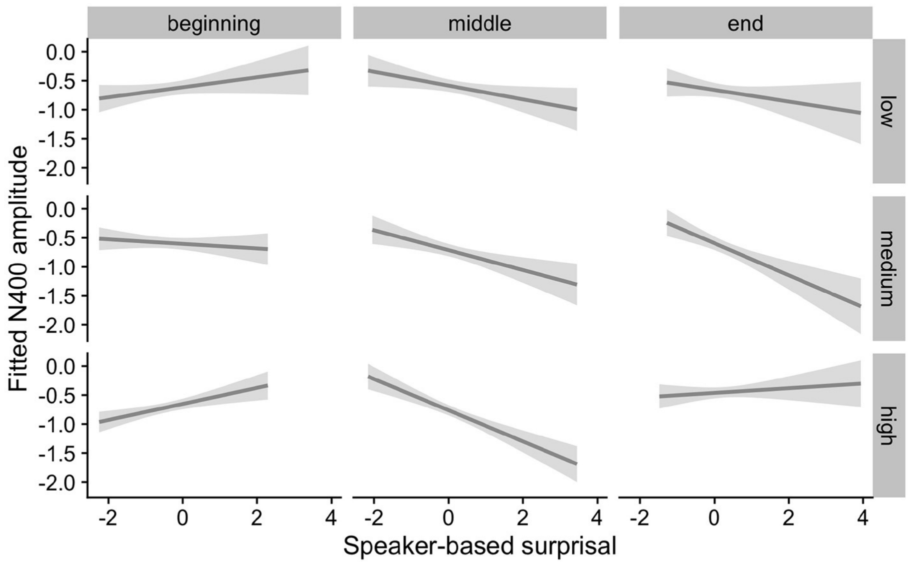


FIGURE 8
Effects of IAF on changes in the relationship between speaker-based surprisal (z-transformed) and N400 amplitude over the course of Experiment 2. The figure visualizes partial effects as calculated using the *remef* package, adjusted for Prestimulus Amplitude and Frequency. Note that position in the experiment (operationalised *via* epoch in the statistical model) is trichotomised into beginning, middle and end for visualization purposes only; epoch was included in the model as a continuous predictor. The same holds for the individual differences variables, which are trichotomised for visualization purposes but were entered into the statistical models as a continuous predictors. Shaded areas indicate 95% confidence intervals.

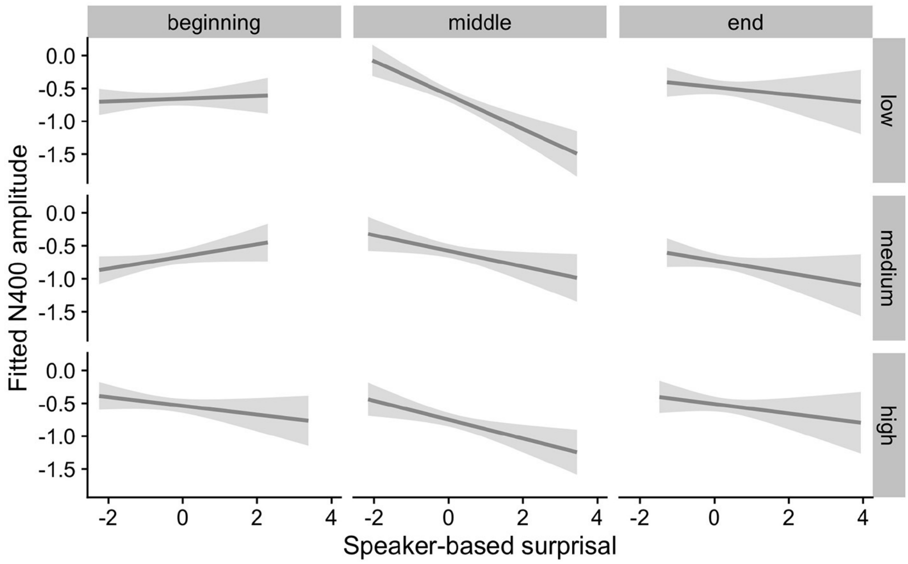


FIGURE 9
Effects of ID on changes in the relationship between speaker-based surprisal (z-transformed) and N400 amplitude over the course of Experiment 2. The figure visualizes partial effects as calculated using the *remef* package, adjusted for Prestimulus Amplitude, Frequency and Canonicity. Note that position in the experiment (operationalised *via* epoch in the statistical model) is trichotomised into beginning, middle and end for visualization purposes only; epoch was included in the model as a continuous predictor. The same holds for the individual differences variables, which are trichotomised for visualization purposes but were entered into the statistical models as a continuous predictors. Shaded areas indicate 95% confidence intervals.

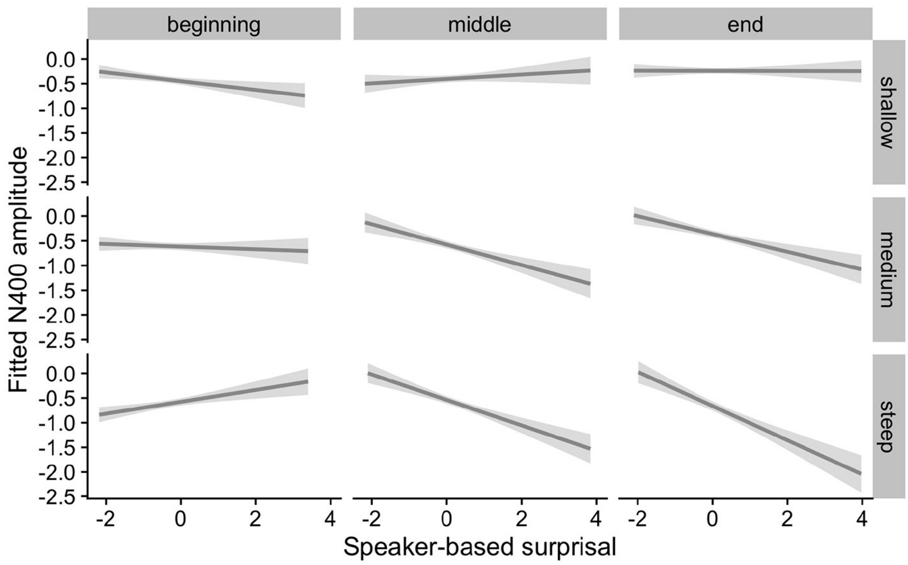


FIGURE 10
Effects of aperiodic (1/f) Slope on changes in the relationship between speaker-based surprisal (z-transformed) and N400 amplitude over the course of the experiment in the combined analysis of Experiments 1 and 2 ($n = 85$). The figure visualizes partial effects as calculated using the *remef* package, adjusted for Prestimulus Amplitude and Canonicity. Note that position in the experiment (operationalised *via* epoch in the statistical model) is trichotomised into beginning, middle and end for visualization purposes only; epoch was included in the model as a continuous predictor. The same holds for the individual differences variables, which are trichotomised for visualization purposes but were entered into the statistical models as a continuous predictors. Shaded areas indicate 95% confidence intervals.

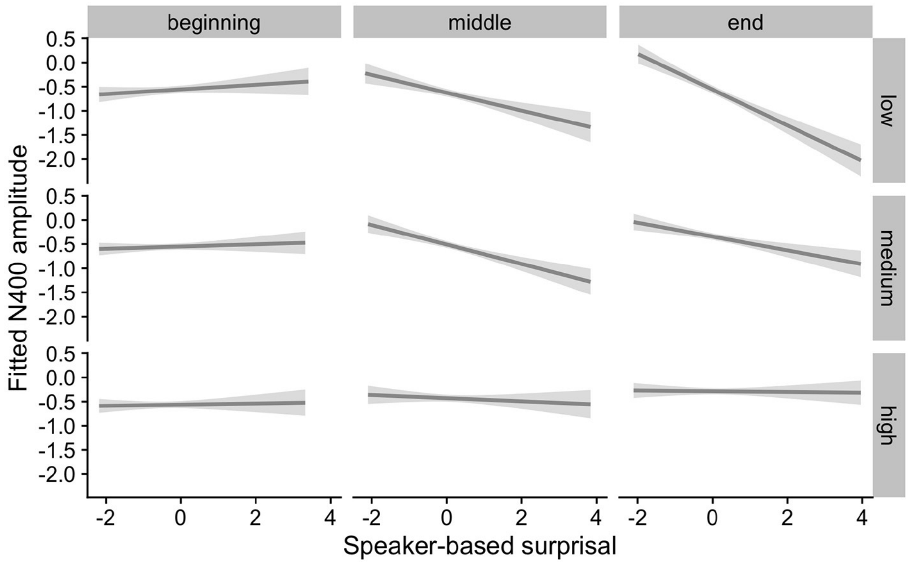


FIGURE 11
Effects of IAF on changes in the relationship between speaker-based surprisal (z-transformed) and N400 amplitude over the course of the experiment in the combined analysis of Experiments 1 and 2 ($n = 85$). The figure visualizes partial effects as calculated using the *remef* package, adjusted for Prestimulus Amplitude and Frequency. Note that position in the experiment (operationalised *via* epoch in the statistical model) is trichotomised into beginning, middle and end for visualization purposes only; epoch was included in the model as a continuous predictor. The same holds for the individual differences variables, which are trichotomised for visualization purposes but were entered into the statistical models as a continuous predictors. Shaded areas indicate 95% confidence intervals.

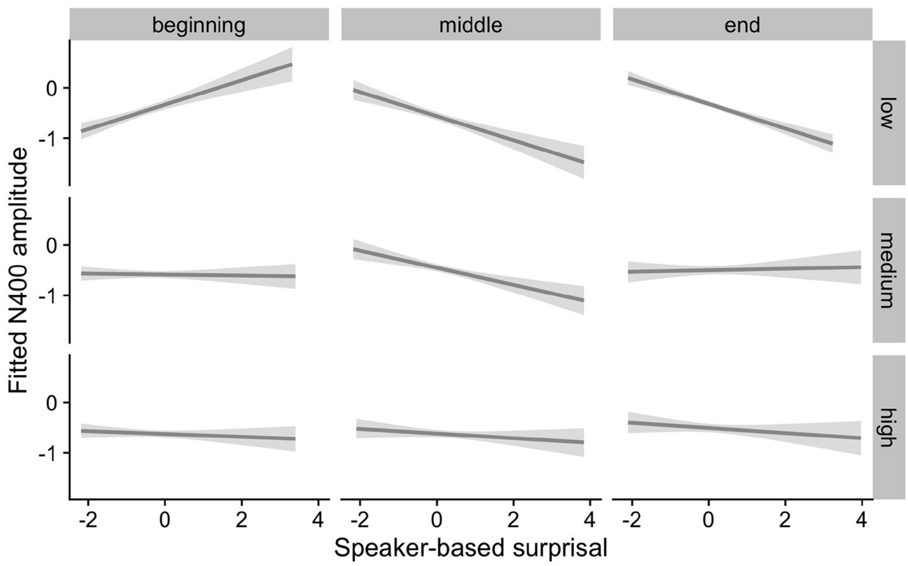


FIGURE 12
Effects of ID on changes in the relationship between speaker-based surprisal (z-transformed) and N400 amplitude over the course of the experiment in the combined analysis of Experiments 1 and 2 ($n = 85$). The figure visualizes partial effects as calculated using the *remef* package, adjusted for Prestimulus Amplitude, Frequency and Canonicity. Note that position in the experiment (operationalised *via* epoch in the statistical model) is trichotomised into beginning, middle and end for visualization purposes only; epoch was included in the model as a continuous predictor. The same holds for the individual differences variables, which are trichotomised for visualization purposes but were entered into the statistical models as a continuous predictors. Shaded areas indicate 95% confidence intervals.

4.1. Individuals incrementally adapt their predictive language models to reflect current contextual information

The current findings present compelling evidence to suggest that individuals incrementally adapt their predictive language models to reflect current contextual information. In spite of only being exposed to new, intra-experimental adjective order regularities for a relatively short period of time, participants' N400 responses had attuned to this new information by the end of the experimental session. Strikingly, this rapid attunement occurred in spite of the wealth of linguistic experience that participants bring to the laboratory from their lifelong exposure to their native language. The importance of intra-experimental information vis-à-vis prior linguistic experience is further underscored by the observation that intra-experimental surprisal effects were aligned for canonical and non-canonical adjective orders by the end of the experiment. This suggests that experiment-specific adjective order probabilities eventually took on a higher weighting in shaping individuals' predictive models than their prior language experience.

Further attesting to the extremely fine-grained nature of the model adaptation process is the observation that N400 amplitude increasingly reflected intra-experimental adjective order surprisal, as calculated incrementally (i.e., on a trial-by-trial basis) for the experimental materials to which a participant had been exposed at each point in the experiment. Moreover, the adaptation took speaker-specific information into account ("speaker-based surprisal"). Previous studies have already demonstrated an adaptation of language comprehension processes to intra-experimental probabilities (Fine et al., 2013), including speaker-specific information (e.g., Kroczek and Gunter, 2017, 2021; Brothers et al., 2019). However, the present study is, to best of our knowledge, the first to demonstrate a gradual attunement to incremental, trial-by-trial fluctuations of intra-experimental, speaker-based surprisal over the course of an experiment.

When intra-experimental probabilities do not align with prior probabilities acquired through experience outside the laboratory, the precision of an individual's global language model is reduced. Model adaptation must thus take place to accommodate speaker-based, intra-experimental contingencies. These are increasingly incorporated into the listener's internal predictive model with increasing exposure to the experimental materials. The attunement of N400 amplitudes to speaker-based surprisal over the course of the experiment thus provides converging support for the proposal that N400 effects reflect precision-weighted prediction error signals (Bornkessel-Schlesewsky and Schlesewsky, 2019). As hypothesized by Bornkessel-Schlesewsky and Schlesewsky (2019), N400 effects thereby functionally mirror MMN effects as observed in auditory oddball paradigms designed to modulate predictive model precision (Todd et al., 2011, 2013, 2014). In these studies, the identity of standard and deviant tones within an auditory oddball paradigm was periodically changed, thus requiring an adaptation of the predictive model. Todd and colleagues observed increased MMN amplitudes within tone sequences that were presented for longer periods of time, i.e., when predictive models had sufficient time to stabilize and increase in precision. However, they also found a primacy effect such that MMN effects were larger for deviations from the tone that was initially established as the standard (Todd et al., 2011). This is indicative of an advantage for the first predictive model to be established and thus attests to the integration of new information with prior knowledge during the course of predictive model adaptation. We suggest that our results show a similar pattern: the observation of speaker-based surprisal effects at the level of adjective clusters demonstrates that intra-experimental contingencies were integrated with prior linguistic knowledge, since the clusters were derived using corpus-based word vectors. Participants were thus clearly still drawing on their prior knowledge of which adjectives tend to behave similarly, while at the same time adjusting their expectations based on the occurrence of these adjectives within the experiment.

4.2. Individual differences in predictive model adaptation

The fine-grained predictive model adaptation observed in the current study differed between individuals. In this regard, we had hypothesized that individuals with steeper 1/f slopes and individuals with higher ID would show a similar adaptation pattern on account of their strong predictive language models, and that this pattern would contrast with that observed for individuals with a higher IAF. Our results provided some converging support for these assumptions but also yielded some previously unexpected insights. Firstly, for 1/f slope and IAF, the directionality of the effects was the opposite of what we had expected: our results suggest a more pronounced adaptation for individuals with a steeper 1/f slope vs. less pronounced adaptation for individuals with a higher IAF. Secondly, the results for ID were less clear in the individual analyses of Experiments 1 and 2, but the combined analysis of both experiments revealed a trend for lower-ID individuals to show more rapid model adaptation, in line with our original hypothesis.

In the following, we discuss 1/f slope, IAF and ID in turn.

4.2.1. Individuals with a steeper aperiodic (1/f) slope show more pronounced effects of model adaptation than those with a shallower aperiodic slope

Participants with a steeper aperiodic (1/f) slope showed a more substantial N400 attunement to speaker-based surprisal over the course of the experiment than their counterparts with a shallower aperiodic slope. This result supports and extends the findings by Dave et al. (2018) that individuals with a steep 1/f slope showed more pronounced prediction-related N400 effects than individuals with a shallow 1/f slope. Dave and colleagues proposed that individuals with low neural noise, as reflected in a steeper 1/f slope, show enhanced prediction (i.e., their study showed a relationship between 1/f slope and N400 effects marking successful vs. unsuccessful lexical prediction). While we had originally hypothesized that this might correlate with a reduced degree of adaptation to intra-experimental contigencies, our findings suggest that, to the contrary, enhanced prediction may in fact be related to an individual's ability to flexibly adapt their neural predictive coding infrastructure to current environmental and task conditions.[2]

This assumption can be linked to the notion that steeper 1/f slopes are indicative of lower levels of neural noise. It is proposed that steeper 1/f slopes in both intracranial and scalp EEG reflect more synchronous neural firing and concomitantly lower rates of aberrant firing or random background activity (for a review of the physiological mechanisms and modeling work that supports this claim, see Voytek and Knight, 2015). The higher signal-to-noise ratio associated with this more synchronous activity can be viewed as reflecting lower neural noise (Hong and Rebec, 2012)[3]. An increase of random neural background activity in aging (increased neural noise) goes hand in hand with increased variability and slowing of neural and behavioral responses to external stimuli (Hong and Rebec, 2012) as well as with a flattening of 1/f slope (Voytek et al., 2015). For example, Tran et al. (2020) observed that increased resting-state neural noise, as reflected in a flatter 1/f slope, in older adults correlated with increased variability of stimulus-related neurophysiological responses (peak alpha inter-trial coherence, ITC) in a visual discrimination task. In relation to predictive coding, lower neural noise possibly allows for a more dynamic and efficient adaptation of task- and context-related neural networks in accordance with current task demands, thus facilitating accurate and context-appropriate predictions.

Pertermann et al. (2019b) recently suggested that there is a relationship between neural noise as indexed by 1/f and neural gain control *via* the noradrenergic system. Release of noradrenaline from the brainstem locus coeruleus leads to increased excitatory and decreased inhibitory responses to a stimulus of interest, thus resulting in stronger stimulus discriminability and a more binary response function (i.e., stronger neural gain, Aston-Jones and Cohen, 2005). In their study, Pertermann et al. (2019b) observed a correlation between 1/f slope and pupil dilation—an index of noradrenergic system activation—in a go/no-go task and specifically for no-go trials requiring response inhibition.

The potential link between lower neural noise and higher neural gain suggests that individuals with a steeper aperiodic slope may be more effective in discriminating between relevant and irrelevant information for the flexible adaptation of their predictive models to the current context. This aligns with an active inference perspective on attention, according to which attention is preferentially allocated toward sensory evidence with a high precision (Parr and Friston, 2017). By optimizing the allocation of attention toward salient/task-relevant information, this could lead to a more rapid establishment of higher-precision models by individuals with a steeper 1/f slope—or, perhaps more precisely, models in which precision is appropriately weighted in light of prior evidence.

4.2.2. Stronger model adaptation for individuals with lower individual alpha frequency

Turning now to IAF, it initially appears somewhat counterintuitive that individuals with a higher IAF show less predictive model adaptation than individuals with a lower IAF. After all, higher IAF correlates with faster processing cycles (Cecere et al., 2015; Samaha and Postle, 2015) and previous findings suggest that older adults with a high IAF show a higher propensity to reanalyze ambiguous ("garden path") sentences when it becomes apparent that the reading initially adopted was incorrect (Kurthen et al., 2020). On the basis of these previous observations, we had thus hypothesized that high-IAF individuals would show a higher propensity for predictive model adaptation than low-IAF individuals. Upon closer consideration, however, the present study differed from the above-cited studies in several important respects. Firstly, in the study by Kurthen et al. (2020), reanalysis did not require an adaptation of the predictive model but rather the correction of a previous processing decision within the bounds of the current model's strategy space. By contrast, the adaptive demands of the present study required participants to learn new, intra-experimental probabilities associated with each speaker and adapt their predictive models to these new contingencies. Secondly, the time frames relevant for these

2 It is worth noting in this context that Dave et al. (2018) examined on-task 1/f activity during their sentence comprehension tasks, while we examined resting-state 1/f activity in the present study. Some recent research suggests that 1/f slope can be linked to global states of consciousness and arousal (Lendner et al., 2020), which could affect predictive model updating through improved attentional regulation and, hence, increased sensitivity to both prediction errors and contextual states. Future research will need to further examine the relationship between resting and on-task 1/f.

3 A complementary perspective on the physiological underpinnings of the 1/f slope is that it indexes the balance between excitatory and inhibitory activity: a flatter slope correlates with more stochastic excitatory firing, which is consistent with reduced inhibitory firing in aging (Gao et al., 2017).

adaptive learning processes were substantially longer than the perceptual windows of interest in the studies by Cecere et al. (2015) and Samaha and Postle (2015), as participants were required to learn two-adjective sequencing regularities over the course of an experimental session. Previous work on the localization of targets moving in space revealed an advantage for individuals with a lower IAF (Howard et al., 2017), with the authors suggesting that this result could be due to the longer timescales involved in the task (movement was between 2 and 4 s in length) in comparison to the transient stimuli used, for example, by Samaha and Postle (2015). In the language domain, Nalaye et al. (2022) recently found that lower-IAF individuals outperformed their higher-IAF counterparts when learning a modified miniature language based on Mandarin Chinese. Akin to the study by Howard et al. (2017), this paradigm involved learning regularities on timescales of multiple seconds. In the present study, lower-IAF individuals may have likewise been better able to adapt their predictive models to the intra-experimental probabilities that unfolded over multiple seconds (intra-stimulus) and minutes (inter-stimulus). However, this explanation remains tentative at present and requires more systematic examination in future research.

4.2.3. A more complex relationship between model adaptation and idea density

As ID measures the efficiency of linguistic encoding (Cheung and Kemper, 1992; Kemper et al., 2001b; Iacono et al., 2009; Engelman et al., 2010; Farias et al., 2012), we examined it as a proxy for the quality of an individual's language model. We thus hypothesized that individuals with lower ID and, hence, a lower quality language model, would show a faster adaptation to new linguistic information. While the results of Experiments 1 and 2 both showed a less clear pattern for ID in comparison to 1/f slope and IAF, the combined analysis of the two experiments does provide some converging evidence for the hypothesis that lower-ID individuals adapted their language models more substantially to the intra-experimental contingencies presented to them.

Low ID in young adulthood is a risk factor for cognitive decline and dementia in old age (Snowdon et al., 1996; Kemper et al., 2001a) and has been suggested to reflect "suboptimal neurocognitive development" (Kemper et al., 2001a, p.602). The notion that lower-ID individuals show a more flexible adaptation of their internal predictive models to the current environment may thus, at a first glance, appear somewhat counterintuitive. Note, however, that faster adaptation in the present study should not necessarily be considered a superior processing strategy. After all, high adaptability means that individuals adjusted expectations accrued through a lifetime of language experience to speaker-specific patterns encountered within a brief experimental session. This could, at least under certain circumstances, lead to the type of "overfitting" of internal predictive models that may be problematic for cognitive performance in older adulthood (Moran et al., 2014).

To better examine the utility of a rapid adaptation strategy, future research could consider model adaptation in different reward contexts, i.e., comparing circumstances where high model malleability is useful to those where it is detrimental to optimal performance. This could yield further insights on calibrated model adaptation, in which the strong prior evidence provided by a high-quality language model is weighed against the increasing quantity of incoming evidence which contradicts the prior model. In addition, the role of domain specificity requires further consideration: of our three individual differences measures of interest, only ID was directly related to the domain under consideration (language), while the other two can be considered to reflect more general characteristics of neural information processing. Future research will need to examine the role of such purported domain-specific vs. domain-general influences in more detail.

Such considerations also reflect a limitation of the current study, namely that possible interactions between individual differences measures were not considered. These are, in our view, outside of the scope of what is already a highly complex pattern of results in a new domain of investigation. However, if our interpretation of the present findings is correct, future studies should be able to further illuminate the mechanics of individuals' model adaptation by taking into account the interplay of the various individual differences metrics examined here.

4.3. Implications for predictive coding in language and beyond

Our results demonstrate that predictive processing during language comprehension adapts flexibly to current contextual and environmental demands, involving both intrinsic linguistic properties (adjective type) as well as communicative aspects (identity of the speaker). They thus extend previous work linking N400 responses to surprisal (e.g., Frank et al., 2015; Frank and Willems, 2017) by demonstrating that corpus-based surprisal may need to be complemented by surprisal metrics that are more closely aligned to the experimental context. To further understand the implications of our findings for predictive coding in language, future research should examine the persistence of predictive model adaptations. It appears unlikely that a single session of exposure to new grammatical or communicative regularities would lead to a permanent adaptation of linguistic models. The application of adapted models to future situations could, however, be governed by cognitive control processes such as those proposed in hierarchical models of cognitive control (e.g., Koechlin and Summerfield, 2007). Here, contextual or episodic information provides control cues to override prepotent stimulus-response mappings and instantiate new mappings for the duration of the appropriate context's or episode's presence. Within the context of the present study, speaker identity could have functioned as one such control cue—in addition to the broader contextual cue of undertaking a language processing task in a laboratory. Participants with a steeper 1/f slope and lower neural noise may be more adept at using such control cues to flexibly switch between alternate predictive models (cf. the association between 1/f neural noise and cognitive control in non-neurotypical populations such as children with ADHD; Pertermann et al., 2019a; Robertson et al., 2019; Ostlund et al., 2021).

A more comprehensive understanding of language processing in contextually rich, naturalistic settings could thus be facilitated by a closer examination of the interplay between predictive coding and cognitive control. Alternatively, cognitive control mechanisms could even be couched within a predictive coding architecture, as proposed by the Hierarchical Error Representation (HER) framework. The HER, which is able to account for a wide range of cognitive control-related findings including hierarchical aspects of cognitive control, posits that "a major function of prefrontal cortex is learning to predict likely prediction errors" (Alexander and Brown, 2018, p.2).

Such an approach could have far-reaching implications for language, including in helping to link linguistic phenomena across different timescales: from processing mechanisms at the scale of tens or hundreds of milliseconds to language change. We have previously suggested that precision-weighted prediction error signals could provide an "early warning signal" for impending language change (Bornkessel-Schlesewsky et al., 2020). Specifically, based on findings from Icelandic, we proposed that reduced N400 effects to a construction that is incompatible with the current prescriptive grammar signal lower predictive precision and, hence, a possible propensity for change. The present findings provide converging support for the very early stages of this proposed process by showing how a loss of precision for a prior linguistic model can lead to rapid model adaptation in accordance with current environmental contingencies. They further suggest that the temporal trajectories for model adaptation differ between individuals, with early adopters being characterized by lower neural noise (steeper aperiodic slope), lower Individual Alpha Frequency and, possibly, lower Idea Density.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://osf.io/32amz/.

Ethics statement

The studies involving human participants were reviewed and approved by University of South Australia's Human Research Ethics Committee. The participants provided their written informed consent to participate in this study.

Author contributions

IB-S, IS, CH, and EW prepared the experiments. IS, CH, and EW collected the data. IB-S and RK performed the data analysis. IB-S wrote the first draft of the manuscript. All authors contributed to conception and design of the study, manuscript revision, and approved the submitted version.

Funding

The research reported here was funded by an Australian Research Council Future Fellowship awarded to IB-S (FT160100437). AC acknowledges the support of the Three Springs Foundation.

Acknowledgments

The authors would like to thank John Ellett for help with preparing the experimental materials, Alin Grecu, Tim Harrison, and Casey Tonkin for recording the experimental stimuli and Nicole Vass for help with data collection.

Conflict of interest

Author PA was employed by Beacon Biosignals, Boston.

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.817516/full#supplementary-material

References

Adams, R. A., Stephan, K. E., Brown, H. R., Frith, C. D., and Friston, K. J. (2013). The computational anatomy of psychosis. *Front. Psychiatry* 4, 47. doi: 10.3389/fpsyt.2013.00047

Alday, P. (2021). *Palday/JellyMe4.jl: V0.2.6 (v0.2.6)*. Zenodo. doi: 10.5281/zenodo.5621582

Alday, P. M. (2018). *Philistine (v0.1) [Source code]*. Available online at: https://github.com/palday/philistine/

Alday, P. M. (2019). How much baseline correction do we need in ERP research? Extended GLM model can replace baseline correction while lifting its limits. *Psychophysiology* 56, 13451. doi: 10.1111/psyp.13451

Alday, P. M., Schlesewsky, M., and Bornkessel-Schlesewsky, I. (2017). Electrophysiology reveals the neural dynamics of naturalistic auditory language processing: event-related potentials reflect continuous model updates. *eNeuro* 4, ENEURO.0311-16.2017. doi: 10.1523/ENEURO.0311-16.2017

Alexander, W. H., and Brown, J. W. (2018). Frontal cortex function as derived from hierarchical predictive coding. *Sci. Rep*. 8, 3843. doi: 10.1038/s41598-018-21407-9

Appelhoff, S., Sanderson, M., Brooks, T., van Vliet, M., Quentin, R., Holdgraf, C., et al. (2019). MNE-BIDS: Organizing electrophysiological data into the BIDS format and facilitating their analysis. *J. Open Source Softw*. 4, 1896. doi: 10.21105/joss.01896

Aston-Jones, G., and Cohen, J. D. (2005). Adaptive gain and the role of the locus coeruleus-norepinephrine system in optimal performance. *J. Comp. Neurol*. 493, 99–110. doi: 10.1002/cne.20723

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., and Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron* 76, 695–711. doi: 10.1016/j.neuron.2012.10.038

Bates, D., Alday, P., Kleinschmidt, D., José Bayoán Santiago Calderón, P., Zhan, L., Noack, A., et al. (2021). *JuliaStats/MixedModels.jl: V4.4.0*. Zenodo. doi: 10.5281/zenodo.5542701

Bates, D., Kliegl, R., Vasishth, S., and Baayen, H. (2015). Parsimonious mixed models. *arXiv:1506.04967 [stat]*. doi: 10.48550/arXiv.1506.04967

Bates, E., Devescovi, A., and Wulfeck, B. (2001). Psycholinguistics: a cross-language perspective. *Annu. Rev. Psychol*. 52, 369–396. doi: 10.1146/annurev.psych.52.1.369

Bates, E., McNew, S., MacWhinney, B., Devescovi, A., and Smith, S. (1982). Functional constraints on sentence processing: a cross-linguistic study. *Cognition* 11, 245–299. doi: 10.1016/0010-0277(82)90017-8

Bezanson, J., Edelman, A., Karpinski, S., and Shah, V. B. (2017). Julia: a fresh approach to numerical computing. *SIAM Rev*. 59, 65–98. doi: 10.1137/141000671

Bolker, B., and Robinson, D. (2021). *broom.mixed: Tidying Methods for Mixed Models*. R package version 1019 0.2.7.

Bornkessel, I. D., Fiebach, C. J., Friederici, A. D., and Schlesewsky, M. (2004). "Capacity" reconsidered: interindividual differences in language comprehension and individual alpha frequency. *Exp. Psychol*. 51, 279–289. doi: 10.1027/1618-3169.51.4.279

Bornkessel-Schlesewsky, I., Philipp, M., Alday, P. M., Kretzschmar, F., Grewe, T., Gumpert, M., et al. (2015a). Age-related changes in predictive capacity versus internal model adaptability: electrophysiological evidence that individual differences outweigh effects of age. *Front. Aging Neurosci*. 7, 217. doi: 10.3389/fnagi.2015.00217

Bornkessel-Schlesewsky, I., Roehm, D., Mailhammer, R., and Schlesewsky, M. (2020). Language processing as a precursor to language change: evidence from icelandic. *Front. Psychol*. 10, 3013. doi: 10.3389/fpsyg.2019.03013

Bornkessel-Schlesewsky, I., and Schlesewsky, M. (2013). Reconciling time, space and function: a new dorsal-ventral stream model of sentence comprehension. *Brain Lang*. 125, 60–76. doi: 10.1016/j.bandl.2013.01.010

Bornkessel-Schlesewsky, I., and Schlesewsky, M. (2019). Towards a neurobiologically plausible model of language-related, negative event-related potentials. *Front. Psychol*. 10, 298. doi: 10.3389/fpsyg.2019.00298

Bornkessel-Schlesewsky, I., and Schlesewsky, M. (2020). "Cross-linguistic neuroscience of language," in *The Cognitive Neurosciences, 6th Edn*, eds M. S. Gazzaniga, G. R. Mangun, and D. Poeppel (Cambridge, MA: MIT Press), 841–848.

Bornkessel-Schlesewsky, I., Schlesewsky, M., Small, S. L., and Rauschecker, J. P. (2015b). Neurobiological roots of language in primate audition: common computational properties. *Trends Cogn Sci*. 19, 142–150. doi: 10.1016/j.tics.2014.12.008

Brehm, L., and Alday, P. M. (2022). Contrast coding choices in a decade of mixed models. *J. Mem. Lang*. 125, 104334. doi: 10.1016/j.jml.2022.104334

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

Brown, C., Snodgrass, T., Kemper, S. J., Herman, R., and a Covington, M. (2008). Automatic measurement of propositional idea density from part-of-speech tagging. *Behav. Res. Methods* 40, 540–545. doi: 10.3758/BRM.40.2.540

Buzsáki, G., Anastassiou, C. A., and Koch, C. (2012). The origin of extracellular fields and currents – EEG, ECoG, LFP and spikes. *Nat. Rev. Neurosci*. 13, 407–420. doi: 10.1038/nrn3241

Buzsáki, G., and Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science* 304, 1926–1929. doi: 10.1126/science.1099745

Cecere, R., Rees, G., and Romei, V. (2015). Individual differences in alpha frequency drive crossmodal illusory perception. *Curr. Biol*. 25, 231–235. doi: 10.1016/j.cub.2014.11.034

Chater, N., and Manning, C. D. (2006). Probabilistic models of language processing and acquisition. *Trends Cogn. Sci*. 10, 335–344. doi: 10.1016/j.tics.2006.05.006

Cheung, H., and Kemper, S. (1992). Competing complexity metrics and adults' production of complex sentences. *Appl. Psycholinguist*. 13, 53. doi: 10.1017/S0142716400005427

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci*. 36, 181–204. doi: 10.1017/S0140525X12000477

Clark, A. (2015). *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. New York, NY: Oxford University Press.

Corcoran, A. W., Alday, P. M., Schlesewsky, M., and Bornkessel-Schlesewsky, I. (2018). Toward a reliable, automated method of individual alpha frequency (IAF) quantification. *Psychophysiology* 55, e13064. doi: 10.1111/psyp.13064

Cross, Z. R., Corcoran, A. W., Schlesewsky, M., Kohler, M. J., and Bornkessel-Schlesewsky, I. (2022). Oscillatory and aperiodic neural activity jointly predict language learning. *J. Cogn. Neurosci*. 24, 1–20. doi: 10.1162/jocn_a_01878

Daneman, M., and Carpenter, P. A. (1980). Individual differences in working memory and reading. *J. Verbal Learn. Verbal Behav*. 19, 450–466. doi: 10.1016/S0022-5371(80)90312-6

Dave, S., Brothers, T., and Swaab, T. (2018). 1/f neural noise and electrophysiological indices of contextual prediction in aging. *Brain Res*. 1691, 34–43. doi: 10.1016/j.brainres.2018.04.007

Dikker, S., and Pylkkänen, L. (2011). Before the N400: Effects of lexical-semantic violations in visual cortex. *Brain Lang*. 118, 23–28. doi: 10.1016/j.bandl.2011.02.006

Dikker, S., Rabagliati, H., Farmer, T. A., and Pylkkänen, L. (2010). Early occipital sensitivity to syntactic category is based on form typicality. *Psychol. Sci*. 21, 629–634. doi: 10.1177/0956797610367751

Donoghue, T., Haller, M., Peterson, E. J., Varma, P., Sebastian, P., Gao, R., et al. (2020). Parameterizing neural power spectra into periodic and aperiodic components. *Nat. Neurosci*. 23, 1655–1665. doi: 10.1038/s41593-020-00744-x

Engelman, M., Agree, E. M. A. Meoni, L., and Klag, M. J. (2010). Propositional density and cognitive function in later life: Findings from the precursors study. *J. Gerontol. B Psychol. Sci. Soc. Sci*. 65, 706–711. doi: 10.1093/geronb/gbq064

Farias, S. T., Chand, V., Bonnici, L., Baynes, K., Harvey, D., Mungas, D., et al. (2012). Idea density measured in late life predicts subsequent cognitive trajectories: implications for the measurement of cognitive reserve. *J. Gerontol. B Psychol. Sci. Soc. Sci*. 67, 677–686. doi: 10.1093/geronb/gbr162

Federmeier, K. D., and Kutas, M. (1999). A rose by any other name: long-term memory structure and sentence processing. *J. Mem. Lang*. 41, 469–495. doi: 10.1006/jmla.1999.2660

Feldman, H., and Friston, K. J. (2010). Attention, uncertainty, and free-energy. *Front. Hum. Neurosci*. 4, 215. doi: 10.3389/fnhum.2010.00215

Ferguson, A., Spencer, E., Craig, H., and Colyvas, K. (2014). Propositional idea density in women's written language over the lifespan: computerized analysis. *Cortex* 55, 107–121. doi: 10.1016/j.cortex.2013.05.012

Fine, A. B., Jaeger, T. F., Farmer, T. A., and Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PLoS ONE* 8, e77661. doi: 10.1371/journal.pone.0077661

Fletcher, P. C., and Frith, C. D. (2009). Perceiving is believing: A Bayesian approach to explaining the positive symptoms of schizophrenia. *Nat. Rev. Neurosci.* 10, 48–58. doi: 10.1038/nrn2536

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

Frank, S. L., and Willems, R. M. (2017). Word predictability and semantic similarity show distinct patterns of brain activity during language comprehension. *Lang. Cogn. Neurosci.* 32, 1192–1203. doi: 10.1080/23273798.2017.1323109

Freeman, W. J., and Zhai, J. (2009). Simulated power spectral density (PSD) of background electrocorticogram (ECoG). *Cogn. Neurodyn.* 3, 97–103. doi: 10.1007/s11571-008-9064-y

Fries, P. (2005). A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends Cogn. Sci.* 9, 474–480. doi: 10.1016/j.tics.2005.08.011

Friston, K. J. (2005). A theory of cortical responses. *Philos. Trans. R. Soc. B Biol. Sci.* 360, 815–836. doi: 10.1098/rstb.2005.1622

Friston, K. J. (2009). The free-energy principle: a rough guide to the brain? *Trends Cogn. Sci.* 13, 293–301.

Friston, K. J., Sajid, N., Quiroga-Martinez, D. R., Parr, T., Price, C. J., and Holmes, E. (2021). Active listening. *Hear. Res.* 399, 107998. doi: 10.1016/j.heares.2020.107998

Frith, C. (2007). *Making up the Mind: How the Brain Creates Our Mental World*. Oxford: Blackwell.

Gao, R., Peterson, E. J., and Voytek, B. (2017). Inferring synaptic excitation/inhibition balance from field potentials. *Neuroimage* 158, 70–78. doi: 10.1016/j.neuroimage.2017.06.078

Garrido, M. I., Kilner, J. M., Stephan, K. E., and Friston, K. J. (2009). The mismatch negativity: a review of underlying mechanisms. *Clin. Neurophysiol.* 120, 453–463. doi: 10.1016/j.clinph.2008.11.029

Gasser, T., Bächer, P., and Steinberg, H. (1985). Test-retest reliability of spectral parameters of the EEG. *Electroencephalogr. Clin. Neurophysiol.* 60, 312–319. doi: 10.1016/0013-4694(85)90005-7

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., et al. (2013). MEG and EEG data analysis with MNE-Python. *Front. Neurosci.* 7, 267. doi: 10.3389/fnins.2013.00267

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., et al. (2014). MNE software for processing MEG and EEG data. *Neuroimage* 86, 446–460. doi: 10.1016/j.neuroimage.2013.10.027

Grandy, T. H., Werkle-Bergner, M., Chicherio, C., Lövdén, M., Schmiedek, F., and Lindenberger, U. (2013a). Individual alpha peak frequency is related to latent factors of general cognitive abilities. *Neuroimage* 79, 10–18. doi: 10.1016/j.neuroimage.2013.04.059

Grandy, T. H., Werkle-Bergner, M., Chicherio, C., and Schmiedek, F. (2013b). Peak individual alpha frequency qualifies as a stable neurophysiological trait marker in healthy younger and older adults. *Psychophysiology* 50, 570–582. doi: 10.1111/psyp.12043

Hale, J. (2006). Uncertainty about the rest of the sentence. *Cogn. Sci.* 30, 643–672. doi: 10.1207/s15516709cog0000_64

Hale, J. (2016). Information-theoretical complexity metrics: Information-theoretical complexity metrics. *Lang. Linguist. Compass* 10, 397–412. doi: 10.1111/lnc3.12196

He, B. J. (2014). Scale-free brain activity: past, present, and future. *Trends Cogn. Sci.* 18, 480–487. doi: 10.1016/j.tics.2014.04.003

Hedden, T., and Gabrieli, J. D. E. (2004). Insights into the ageing mind: a view from cognitive neuroscience. *Nat. Rev. Neurosci.* 5, 87–96. doi: 10.1038/nrn1323

Hester, J., and Wickham, H. (2021). *vroom: Read and Write Rectangular Text Data Quickly*. R package version 1.5.3.

Hohenstein, S., and Kliegl, R. (2021). *remef: Remove Partial Effects*. R package version 1.0.7.

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

Hong, S. L., and Rebec, G. V. (2012). A new perspective on behavioral inconsistency and neural noise in aging: Compensatory speeding of neural communication. *Front. Aging Neurosci.* 4, 27. doi: 10.3389/fnagi.2012.00027

Howard, C. J., Arnold, C. P., and Belmonte, M. K. (2017). Slower resting alpha frequency is associated with superior localisation of moving targets. *Brain Cogn.* 117, 97–107. doi: 10.1016/j.bandc.2017.06.008

Iacono, D., Markesbery, W. R., Gross, M., Pletnikova, O., Rudow, G., Zandi, P., et al. (2009). The nun study. Clinically silent AD, neuronal hypertrophy, and linguistic skills in early life. *Neurology* 73, 665–673. doi: 10.1212/WNL.0b013e3181b01077

Jurafsky, D. (2003). "Probabilistic modeling in psycholinguistics: linguistic comprehension and production," in *Probabilistic Linguistics, volume 21*, eds R. Bod, J. Hay, and S. Jannedy (Cambridge, MA; London: MIT Press), 39–95.

Kemmerer, D., Weber-Fox, C., Price, K., Zdanczyk, C., and Way, H. (2007). Big brown dog or brown big dog? An electrophysiological study of semantic constraints on prenominal adjective order. *Brain Lang.* 100, 238–256. doi: 10.1016/j.bandl.2005.12.002

Kemper, S., Greiner, L. H., Marquis, J. G., Prenovost, K., and Mitzner, T. L. (2001a). Language decline across the life span: Findings from the nun study. *Psychol. Aging* 16, 227–239. doi: 10.1037/0882-7974.16.2.227

Kemper, S., Thompson, M., and Marquis, J. G. (2001b). Longitudinal change in language production: effects of aging and dementia on grammatical complexity and propositional content. *Psychol. Aging* 16, 600–614. doi: 10.1037/0882-7974.16.4.600

Kintsch, W., and Keenan, J. (1973). Reading rate and retention as a function of the number of propositions in the base structure of sentences. *Cogn. Psychol.* 5, 257–274. doi: 10.1016/0010-0285(73)90036-4

Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Res. Rev.* 29, 169–195. doi: 10.1016/S0165-0173(98)00056-3

Knill, D. C., and Pouget, A. (2004). The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends Neurosci.* 27, 712–719. doi: 10.1016/j.tins.2004.10.007

Koechlin, E., and Summerfield, C. (2007). An information theoretical approach to prefrontal executive function. *Trends Cogn. Sci.* 11, 229–235. doi: 10.1016/j.tics.2007.04.005

Kondacs, A., and Szab,ó, M. (1999). Long-term intra-individual variability of the background EEG in normals. *Clin. Neurophysiol.* 110, 1708–1716. doi: 10.1016/S1388-2457(99)00122-4

Köpruner, V., Pfurtscheller, G., and Auer, L. (1984). "Quantitative EEG in normals and in patients with cerebral ischemia," in *Brain Ischemia: Quantitative EEG and Imaging Techniques, volume 62 of Progress in Brain Research*, eds G. Pfurtscheller, E. Jonkman, and F. Lopes da Silva (Amsterdam: Elsevier), 29–50.

Kroczek, L. O., and Gunter, T. C. (2021). The time course of speaker-specific language processing. *Cortex* 141, 311–321. doi: 10.1016/j.cortex.2021.04.017

Kroczek, L. O. H., and Gunter, T. C. (2017). Communicative predictions can overrule linguistic priors. *Sci. Rep.* 7, 17581. doi: 10.1038/s41598-017-17907-9

Kuhn, M., Jackson, S., and Cimentada, J. (2020). *corrr: Correlations in R*. R package version 0.4.3

Kuhn, M., and Wickham, H. (2020). Tidymodels: A Collection Of Packages For Modeling And Machine Learning Using Tidyverse Principles. Query: Provide the city and publisher name for "Kuhn and Wickham, 2020."

Kuperberg, G. R. (2016). Separate streams or probabilistic inference? What the N400 can tell us about the comprehension of events. *Lang. Cogn. Neurosci.* 31, 602–616. doi: 10.1080/23273798.2015.1130233

Kuperberg, G. R., and Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Lang. Cogn. Neurosci.* 31, 32–59. doi: 10.1080/23273798.2015.1102299

Kurthen, I., Meyer, M., Schlesewsky, M., and Bornkessel-Schlesewsky, I. (2020). Individual differences in peripheral hearing and cognition reveal sentence processing differences in healthy older adults. *Front. Neurosci.* 14, 573513. doi: 10.3389/fnins.2020.573513

Laszlo, S., and Federmeier, K. D. (2009). A beautiful day in the neighborhood: an event-related potential study of lexical relationships and prediction in context. *J. Mem. Lang.* 61, 326–338. doi: 10.1016/j.jml.2009.06.004

Laszlo, S., and Federmeier, K. D. (2011). The N400 as a snapshot of interactive processing: evidence from regression analyses of orthographic neighbor and lexical associate effects. *Psychophysiology* 48, 176–186. doi: 10.1111/j.1469-8986.2010.01058.x

Lendner, J. D., Helfrich, R. F., Mander, B. A., Romundstad, L., Lin, J. J., Walker, M. P., et al. (2020). An electrophysiological marker of arousal level in humans. *Elife* 9, e55092. doi: 10.7554/eLife.55092

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

MacWhinney, B., Bates, E., and Kliegl, R. (1984). Cue validity and sentence interpretation in English, German, and Italian. *J. Verbal Learn. Verbal Behav.* 23, 127–150. doi: 10.1016/S0022-5371(84)90093-8

Matuschek, H., and Kliegl, R. (2018). On the ambiguity of interaction and nonlinear main effects in a regime of dependent covariates. *Behav. Res. Methods* 50, 1882–1894. doi: 10.3758/s13428-017-0956-9

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., and Bates, D. (2017). Balancing Type I error and power in linear mixed models. *J. Mem. Lang.* 94, 305–315. doi: 10.1016/j.jml.2017.01.001

McWhite, C. D., and Wilke, C. O. (2021). *colorblindr: Simulate colorblindness in R figures*. R package version 0.1.0.

Moran, R. J., Symmonds, M., Dolan, R. J., and Friston, K. J. (2014). The brain ages optimally to model its environment: evidence from sensory learning over the adult lifespan. *PLoS Comput. Biol.* 10, e1003422. doi: 10.1371/journal.pcbi.1003422

Müller, K. (2020). *here: A Simpler Way to Find Your Files*. R package version 1.0.1.

Nalaye, H., Cross, Z. R., Schlesewsky, M., and Bornkessel-Schlesewsky, I. (2022). Electrophysiological indices of individual differences in adult language learning. *bioRxiv*. doi: 10.1101/2022.06.07.495229

Ociepka, M., Kałamała, P., and Chuderski, A. (2022). High individual alpha frequency brains run fast, but it does not make them smart. *Intelligence* 92, 101644. doi: 10.1016/j.intell.2022.101644

Oldfield, R. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia* 9, 97–113. doi: 10.1016/0028-3932(71)90067-4

Ostlund, B. D., Alperin, B. R., Drew, T., and Karalunas, S. L. (2021). Behavioral and cognitive correlates of the aperiodic (1/f-like) exponent of the EEG power spectrum in adolescents with and without ADHD. *Dev. Cogn. Neurosci.* 48, 100931. doi: 10.1016/j.dcn.2021.100931

Ouyang, G., Hildebrandt, A., Schmitz, F., and Herrmann, C. S. (2020). Decomposing alpha and 1/f brain activities reveals their differential associations with cognitive processing speed. *Neuroimage* 205, 116304. doi: 10.1016/j.neuroimage.2019.116304

Parr, T., and Friston, K. J. (2017). Working memory, attention, and salience in active inference. *Sci. Rep.* 7, 14678. doi: 10.1038/s41598-017-15249-0

Pedersen, T. L. (2020). *patchwork: The Composer of Plots*. R package version 1.1.1.

Pereira, F., Lou, B., Pritchett, B., Ritter, S., Gershman, S. J., Kanwisher, N., et al. (2018). Toward a universal decoder of linguistic meaning from brain activation. *Nat. Commun.* 9, 963. doi: 10.1038/s41467-018-03068-4

Pernet, C. R., Appelhoff, S., Gorgolewski, K. J., Flandin, G., Phillips, C., Delorme, A., et al. (2019). EEG-BIDS, an extension to the brain imaging data structure for electroencephalography. *Scientific Data* 6, 103. doi: 10.1038/s41597-019-0104-8

Pertermann, M., Bluschke, A., Roessner, V., and Beste, C. (2019a). The modulation of neural noise underlies the effectiveness of methylphenidate treatment in attention-deficit/hyperactivity disorder. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* 4, 743–750. doi: 10.1016/j.bpsc.2019.03.011

Pertermann, M., Mückschel, M., Adelhöfer, N., Ziemssen, T., and Beste, C. (2019b). On the interrelation of 1/ *f* neural noise and norepinephrine system activity during motor response inhibition. *J. Neurophysiol.* 121, 1633–1643. doi: 10.1152/jn.00701.2018

Pickering, M. J., and Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends Cogn. Sci.* 11, 105–110. doi: 10.1016/j.tics.2006.12.002

Pickering, M. J., and Garrod, S. (2013). An integrated theory of language production and comprehension. *Behav. Brain Sci.* 36, 329–347. doi: 10.1017/S0140525X12001495

Poeppel, D., Idsardi, W. J., and van Wassenhove, V. (2008). Speech perception at the interface of neurobiology and linguistics. *Philos. Trans. R. Soc. B Biol. Sci.* 363, 1071–1086. doi: 10.1098/rstb.2007.2160

Posthuma, D., Neale, M. C., Boomsma, D. I., and De Geus, E. J. C. (2001). Are smarter brains running faster? Heritability of alpha peak frequency, IQ, and their interrelation. *Behav. Genet.* 31, 567–579. doi: 10.1023/A:10133454 11774

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Rao, R. P., and Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* 2, 79–87. doi: 10.1038/4580

Rauschecker, J. P., and Scott, S. K. (2009). Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nat. Neurosci.* 12, 718–724. doi: 10.1038/nn.2331

Robertson, M. M., Furlong, S., Voytek, B., Donoghue, T., Boettiger, C. A., and Sheridan, M. A. (2019). EEG power spectral slope differs by ADHD status and stimulant medication exposure in early childhood. *J. Neurophysiol.* 122, 2427–2437. doi: 10.1152/jn.00388.2019

Robinson, D. (2021). widyr: Widen, Process, then Re-Tidy Data. R package version 0.1.4

Salthouse, T. A. (2011). Neuroanatomical substrates of age-related cognitive decline. *Psychol. Bull.* 137, 753–784. doi: 10.1037/a0023262

Samaha, J., and Postle, B. R. (2015). The speed of alpha-band oscillations predicts the temporal resolution of visual perception. *Curr. Biol.* 25, 2985–2990. doi: 10.1016/j.cub.2015.10.007

Sanborn, A. N., and Chater, N. (2016). bayesian brains without probabilities. *Trends Cogn. Sci.* 20, 883–893. doi: 10.1016/j.tics.2016.10.003

Schad, D. J., Vasishth, S., Hohenstein, S., and Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: a tutorial. *J. Mem. Lang.* 110, 104038. doi: 10.1016/j.jml.2019.104038

Silge, J. and Robinson, D. (2016). tidytext: Text mining and analysis using tidy data principles in R. J. Open Source Softw. 1/, 37. doi: 10.21105/joss.00037

Skipper, J. I., van Wassenhove, V., Nusbaum, H. C., and Small, S. L. (2007). Hearing lips and seeing voices: how cortical areas supporting speech production mediate audiovisual speech perception. *Cereb*. Cortex 17, 2387–2399. doi: 10.1093/cercor/bhl147

Smit, C. M., Wright, M. J., Hansell, N. K., Geffen, G. M., and Martin, N. G. (2006). Genetic variation of individual alpha frequency (IAF) and alpha power in a large adolescent twin sample. *Int. J. Psychophysiol.* 61, 235–243. doi: 10.1016/j.ijpsycho.2005.10.004

Snowdon, D. A., Kemper, S. J., Mortimer, J. A., Greiner, L. H., Wekstein, D. R., and Markesbery, W. R. (1996). Linguistic ability in early life and cognitive function and Alzheimer's disease in late life. Findings from the Nun Study. *JAMA* 275, 528–532. doi: 10.1001/jama.1996.03530310034029

Spencer, E., Ferguson, A., Craig, H., Colyvas, K., Hankey, G. J., and Flicker, L. (2015). Propositional idea density in older men's written language: findings from the HIMS study using computerised analysis. *Clin. Linguist. Phonet.* 29, 85–101. doi: 10.3109/02699206.2014.956263

Spratling, M. (2008). Predictive coding as a model of biased competition in visual attention. *Vis. Res.* 48, 1391–1408. doi: 10.1016/j.visres.2008.03.009

Surwillo, W. W. (1961). Frequency of the alpha rhythm, reaction time and age. *Nature* 191, 823–824. doi: 10.1038/191823a0

Surwillo, W. W. (1963). The relation of simple response time to brain-wave frequency and the effects of age. *Electroencephalogr. Clin. Neurophysiol.* 15, 105–114. doi: 10.1016/0013-4694(63)90043-9

Todd, J., Heathcote, A., Mullens, D., Whitson, L. R., Provost, A., and Winkler, I. (2014). What controls gain in gain control? Mismatch negativity (MMN), priors and system biases. *Brain Topography* 27, 578–589. doi: 10.1007/s10548-013-0 344-4

Todd, J., Provost, A., and Cooper, G. (2011). Lasting first impressions: a conservative bias in automatic filters of the acoustic environment. *Neuropsychologia* 49, 3399–3405. doi: 10.1016/j.neuropsychologia.2011. 08.016

Todd, J., Provost, A., Whitson, L. R., Cooper, G., and Heathcote, A. (2013). Not so primitive: context-sensitive meta-learning about unattended sound sequences. *J. Neurophysiol.* 109, 99–105. doi: 10.1152/jn.00581.2012

Tran, T. T., Rolle, C. E., Gazzaley, A., and Voytek, B. (2020). Linked sources of neural noise contribute to age-related cognitive decline. *J. Cogn. Neurosci.* 32, 1813–1822. doi: 10.1162/jocn_a_01584

Vallat, R., and Walker, M. P. (2021). A universal, open-source, high-performance tool for automated sleep staging. *Preprint BioRxiv*. doi: 10.1101/2021.05.28.446165

van Paridon, J., and Thompson, B. (2021). Subs2vec: Word embeddings from subtitles in 55 languages. *Behav. Res. Methods* 53, 629–655. doi: 10.3758/s13428-020-01406-3

Van Petten, C., and Luka, B. (2012). Prediction during language comprehension: benefits, costs, and ERP components. *Int. J. Psychophysiol.* 83, 176–190. doi: 10.1016/j.ijpsycho.2011.09.015

VanRullen, R. (2016). Perceptual cycles. *Trends Cogn. Sci.* 20, 723–735. doi: 10.1016/j.tics.2016.07.006

Vilares, I., and Kording, K. (2011). Bayesian models: The structure of the world, uncertainty, behavior, and the brain. *Ann. N. Y. Acad. Sci.* 1224, 22–39. doi: 10.1111/j.1749-6632.2011.05965.x

Voytek, B., and Knight, R. T. (2015). Dynamic network communication as a unifying neural basis for cognition, development, aging, and disease. *Biol. Psychiatry* 77, 1089–1097. doi: 10.1016/j.biopsych.2015.04.016

Voytek, B., Kramer, M. A., Case, J., Lepage, K. Q., Tempesta, Z. R., Knight, R. T., et al. (2015). Age-related changes in 1/f neural electrophysiological noise. *J. Neurosci.* 35, 13257–13265. doi: 10.1523/JNEUROSCI.2332-14.2015

Wang, L., Bastiaansen, M., Yang, Y., and Hagoort, P. (2011). The influence of information structure on the depth of semantic processing: how focus and pitch accent determine the size of the N400 effect. *Neuropsychologia* 49, 813–820. doi: 10.1016/j.neuropsychologia.2010.12.035

Wen, H., and Liu, Z. (2016). Separating fractal and oscillatory components in the power spectrum of neurophysiological signal. *Brain Topogr.* 29, 13–26. doi: 10.1007/s10548-015-0448-0

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer. https://ggplot2.tidyverse.org

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., et al. (2019). Welcome to the tidyverse. *J. Open Source Softw.* 4, 1686. doi: 10.21105/joss.01686

Wickham, H., Chang, W., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C., et al. (2021). *ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics [Manual]*. Available online at: https://CRAN.R-project.org/package=ggplot2

Wilke, C. O. (2021). *cowplot: Streamlined Plot Theme and Plot Annotations for ggplot2 [Manual]*. Available online at: https://wilkelab.org/cowplot/

Zhu, H. (2021). *kableExtra: Construct Complex Table with kable and Pipe Syntax [Manual]*. Available online at: https://CRAN.R-project.org/package=kableExtra