Psychophysiology

WILEY

PSYCHOPHYSIOLOGY SPR

ORIGINAL ARTICLE OPEN ACCESS

Individual Differences in Statistical Learning and Semantic Adaptation: An N400 Study

Márton Munding[1,2] | Bálint Forgács[2,3] | Krisztina Sára Lukics[4,5] | Ágnes Lukács[4,5,6]

[1]Doctoral School of Psychology, ELTE Eötvös Loránd University, Budapest, Hungary | [2]Department of Cognitive Psychology, Institute of Psychology, ELTE Eötvös Loránd University, Budapest, Hungary | [3]Department of Experimental and Neurocognitive Psychology, Freie Universität Berlin, Berlin, Germany | [4]Department of Cognitive Science, Faculty of Natural Sciences, Budapest University of Technology and Economics, Budapest, Hungary | [5]MTA-BME Momentum Language Acquisition Research Group, Budapest, Hungary | [6]Cognitive Medicine Research Group, Competence Centre for Neurocybernetics of the Life Sciences Cluster of the Centre of Excellence for Interdisciplinary Research, Development and Innovation of the University of Szeged, University of Szeged, Szeged, Hungary

Correspondence: Márton Munding (munding.marton@ppk.elte.hu)

Received: 15 July 2024 | **Revised:** 6 May 2025 | **Accepted:** 24 July 2025

Funding: This work was supported by the Lendület Research Grant of the Hungarian Academy of Sciences (Momentum 96233 'Profiling learning mechanisms and learners: individual differences from impairments to excellence in statistical learning and in language acquisition', PI: Ágnes Lukács) and by Grant TKP2021-NVA-02, and by a DFG grant of the German Research Council (CI247/7-1, Bálint Forgács).

ABSTRACT

Recent empirical results have linked the N400 ERP with predictive language comprehension processes based on statistical learning (SL). However, links between SL abilities and N400 on the level of individual differences have so far been underexplored. The present study tested SL performance in 29 participants using speech segmentation and artificial grammar learning tasks, followed by EEG recordings of their N400 responses to sentences varying in cloze probability (high, intermediate, low). Mixed-effects models revealed that better online SL performance (SL-ON) was associated with larger N400 amplitudes across conditions. Additionally, working memory showed a significant main effect and interacted with SL-ON in modulating N400 amplitude, while cloze probability also had a robust, independent effect on it. These results demonstrate that individual differences in SL abilities contribute to N400 response variability, supporting the view that the semantic operations reflected by the N400 may involve some form of statistical learning as well. Our findings also raise the possibility that SL and CP tap into distinct levels of predictive mechanisms in language comprehension.

1 | Introduction

1.1 | Statistical Learning

Statistical learning (SL) is a key function of human cognition. It involves our ability to pick up on regularities of the environment around us, such as frequency distributions and transitional probabilities through observation, without overt effort (Aslin 2017; Conway 2020; Siegelman and Frost 2015), operating over all types of stimuli across modalities and domains (e.g., Conway and Christiansen 2006, 2009; Lukics and Lukács 2022). Humans are equipped with this ability from the very first days of their lives (Bulf et al. 2011; Teinonen et al. 2009), and SL remains available through the entire lifespan (see e.g., Janacsek et al. 2012; Lukács and Kemény 2015). This function allows us to keep track of environmental patterns, make anticipatory predictions based on these regularities, and adapt to changes without a demanding cognitive effort or conscious awareness of the process, all of which are critical for tasks like recognizing faces, enjoying music, or understanding language.

The role SL plays in language acquisition and processing is an especially well-studied phenomenon. As developmental research shows, infants can identify various linguistic structures at different levels of analysis, from phonemes or word boundaries to syntactic patterns, using only statistical information (Saffran et al. 1996, 1999). Already at this early age, SL allows infants to recognize transitional probabilities between elements and to effectively segment both verbal and non-verbal stimuli into coherent units. Additionally, infants' ability to identify sequential regularities of nonsensical syllables based on toy grammar rules (Gomez and Gerken 1999; Marcus et al. 1999) suggests a link between SL and syntax learning. Beyond sequential learning, SL also plays a role in non-sequential categorization, helping in phoneme category learning (Kuhl 2000) and the grasping of word meanings and concepts (Yu and Smith 2007). The availability of such abilities for learning various linguistic patterns at critical stages of language development highlights the significant role of SL in language acquisition.

Moreover, SL performance seems to be connected to certain aspects of language proficiency in adults and children alike. These include speech perception in noisy environments, lexical abilities, natural language aptitude in adults, syntactic priming in preschoolers, semantic predictions, comprehension of syntax, and reading skills in typical readers (Conway et al. 2007, 2010; Mainela-Arnold and Evans 2014; Daltrozzo et al. 2017; Kidd 2012; Kidd and Arciuli 2016; Lukacs et al. under review; Misyak and Christiansen 2012; Torkildsen et al. 2019; Arciuli and Simpson 2012; Spencer et al. 2015; see Boeve et al. 2023, for a meta-analysis). Findings also indicate a neural overlap between SL and language processing (Christiansen et al. 2012). In summary, these findings highlight the integral role of SL not only in language acquisition but also in language processing. A key question we set out to investigate in the present paper is how individual differences in SL ability may be related to on-line sentence processing.

1.2 | Electrophysiology of Online Language Comprehension: The N400

The N400 is an electrophysiological event-related potential (ERP) related to semantic processing. It is negative in amplitude, appears centro-parietally, between 300 and 500 ms after stimulus presentation onset, and peaks around 400 ms (Kutas and Hillyard 1980; Kutas and Federmeier 2011). It is known to be sensitive to sentence context manipulations (Kutas and Federmeier 2011), and while it has been proposed to be linked to semantic integration (Berkum et al. 1999; Hagoort et al. 2009), more recent analyses suggest that it is more likely to reflect semantic memory retrieval effort (Brouwer et al. 2012; Urbach et al. 2020). The magnitude of the N400 response has been shown to be sensitive not simply to semantic violations but to the predictability of sentence endings, with less predictable sentence final words evoking more negative responses (Kutas and Federmeier 2011). Such a predictability is often operationalized as cloze probability (CP) or the likelihood of a particular word being suggested by readers to finish a particular sentence cloze (Federmeier 2007; Taylor 1953).

1.3 | N400 Influenced by Statistical Learning?

The N400 has been suggested to originate from a functionally diverse set of neural generators (Lau et al. 2008). On one hand, the N400 is sensitive to surface-level distributional properties of language, for example, neighbor frequency (Molinaro et al. 2010), association strength (Van Petten 1993), or orthographic neighborhood size (Laszlo and Federmeier 2011). This pattern suggests that the N400 is not strictly post-lexical (Deacon et al. 2000); rather, it is also influenced by a system that continuously monitors and responds to probabilistic regularities in the input. On the other hand, the N400 is also responsive to sentence-level context (e.g., Federmeier 2007; DeLong et al. 2014; Szewczyk and Federmeier 2022; Van Petten 1993), and world knowledge (Hagoort et al. 2004; Hald et al. 2007), including discourse congruence and event structure predictability as well (Kutas and Federmeier 2011; Nieuwland and Van Berkum 2006). Such effects demonstrate that the N400 reflects updates to representations at multiple levels of abstraction.

The sensitivity of the N400 to both low- and high-level linguistic factors suggests the involvement of adaptive mechanisms during language comprehension. These mechanisms can be regarded as forms of SL, since they entail adapting to changes in linguistic frequencies of the environment or within individuals' language systems. For example, the responsiveness of the N400 to the strength of semantic associations between words (Kutas and Hillyard 1984; Van Petten 1993, 2014) appears to reflect the brain's reliance on previously encountered distributional regularities.

In recent decades, there has been a considerable shift in the views regarding the role of predictive mechanisms in linguistic cognition. With mounting experimental evidence on the predictive cognitive processes of comprehenders during communication (see the review of Pickering and Gambi 2018), based on cues from both auditory and visual modalities (Hintz et al. 2024), some propose that prediction is one of the fundamental computations of language comprehension (Clark 2013). Although the extent and nature of prediction's role remain debated (e.g., Kuperberg and Jaeger 2016), it may substantially contribute to language comprehension and, therefore, statistical learning abilities may also impact the N400 response.

Evidence supporting the contribution of prediction beyond associative strength comes from the study of Lau et al. (2013), which showed that manipulating the semantic relatedness proportions (how often semantically related word pairs occurred in a given experiment block) modulated the N400 amplitude. This finding suggests that the N400 response varies as a function of global contextual statistics and predictability as well: when semantically related word pairs occurred more frequently in the task, targets elicited reduced N400 responses, indicating sensitivity to learned contextual probabilities beyond local lexical associations.

However, while co-occurrence or frequency-related effects on the N400 are well documented (Kutas et al. 2016), they have not been identified or categorized as SL. One specific model that explicitly argues for the primary role of implicit learning in the generation of the N400 response is the Sentence Gestalt Model (Rabovsky et al. 2018). According to the Sentence Gestalt Model, each word in a sentence provides clues that constrain the formation of an implicit probabilistic representation of the sentence content: an internal model that continuously adapts to the distributional statistics of the incoming linguistic input. In this framework, the N400 response is a correlate of a semantic update of the probabilistic event model of the linguistic utterance. Conversely, a larger N400 response may correspond to more effective adaptation: words that elicited a larger N400 response were found to be associated with a higher implicit memory benefit (Hodapp and Rabovsky 2021). Hodapp and Rabovsky (2021) subjected participants to sentences varying in the expectancy of the final word, while measuring their N400 responses in the first part of the study. The second part consisted of a perceptual identification task of expected versus unexpected sentence final words to measure implicit memory performance. They found increased implicit memory performance for words that were unexpected—and elicited larger N400 responses during the reading task. They interpreted this finding as an adjustment of the internal probabilistic representations, triggered by the unexpected words. The authors suggested that their model generalizes to individual differences: those with more effective adaptation (i.e., better statistical learning abilities) are expected to show larger N400 responses. In the current study, we set out to test this idea: while the N400 is subject to individual variation (Van Den Brink et al. 2012) and sensitive to differences in working memory performance (Van Petten et al. 1997), the link between statistical learning ability and the N400 response at the level of individual differences has not yet been tested directly.

In addition, Conway et al. (2010) found that individuals with stronger SL abilities are better at speech perception in noise, that is, at identifying probabilistic patterns in language, which allows them to anticipate upcoming words more effectively based on contextual cues. They argue that SL abilities are essential for developing long-term representations of the sequential structure of language, which may guide predictions during comprehension. This sensitivity to predictability raises the possibility that the N400 reflects a similar statistical learning mechanism on the semantic level. While Rabovsky et al. (2018) Sentence Gestalt Model suggests that the N400 reflects continuous updates to probabilistic sentence representations, we aim to test directly whether the predictive semantic mechanisms responsible for the N400 are associated with individual differences in SL. If individuals with stronger SL abilities form more accurate internal models of linguistic probabilities, which allows them to generate better predictions, they may be expected to show larger N400 responses when their expectations are not met.

1.4 | Testing the Role of Statistical Learning in Predictive Semantic Retrieval

In the current study, we set out to investigate how individual differences in SL may be related to the N400 response, while controlling for working memory (WM) and processing speed. Given the multifaceted nature of SL, we used two tasks to assess individual differences: (1) a speech segmentation task to assess the learning of lower-level, simple adjacent statistical relationships; and (2) an artificial grammar learning task to assess the acquisition of more complex regularities characterized by more diverse transition probabilities and wider dependency spans. In both tasks, we assessed statistical learning as reflected by reaction times during a training phase (yielding online indices) and by judgments about the structure of stimulus samples after the training phase (yielding offline indices).

We also explored whether the effect of SL on the N400 response can be explained by individual variability in working memory or auditory and visual processing speed, which have been shown to influence sentence comprehension and to facilitate predictive processing (Hintz et al. 2024). During language-vision interactions, comprehenders may need to exploit cues in both the spoken and the visual modality efficiently, to generate predictions about upcoming words (Ferreira et al. 2013; Hintz et al. 2020). WM has also been found to strongly correlate with both statistical learning and language comprehension measures (Misyak and Christiansen 2012), and it is known to affect the N400 response during sentence comprehension. In their landmark study, Van Petten et al. (1997) demonstrated that higher N400 responses are evoked in people with higher WM capacity, but only when information is presented in sentence level semantic context, not in single-word lexical context. This difference was attributed to higher availability of sentence context for high WM readers, in contrast to the less efficient contextual processing of those with lower WM capacity.

To assess the N400 response magnitude during sentence processing, we designed an EEG experiment, where participants' N400 responses were recorded while they were reading novel sentences ending in words with varying predictability. Predictability was quantified by sentence cloze probabilities, representing the likelihood of a particular target word completing a given open sentence frame (Taylor 1953). CPs were collected and calculated from questionnaire-based data. To obtain variability in the magnitude of the N400 response, participants were shown sentences from three predefined CP categories: (1) high-CP sentences (80% < CP), (2) intermediate-CP sentences (20% < CP < 55%) and (3) low-CP sentences (CP < 20%). The difference in ERP amplitudes between high-CP and low-CP sentences was expected to elicit a typical N400 effect, with the intermediate-CP sentences falling in between.

Our hypotheses regarding the relationship between SL and the N400 response were twofold. First, based on the proposed role of SL in language processing, along with insights from predictive language processing models (Kuperberg and Jaeger 2016; Pickering and Gambi 2018), we predict an association between individuals SL performance and the magnitude of their N400 response. We expect larger N400 effects (i.e., differences) to sentences varying in CP in individuals showing higher SL performance. If the N400 indicates the updating of a predictive model (cf. Rabovsky et al. 2018), better statistical learners may be able to make a better adjustment of their internal model. Second, we hypothesize that this relationship would impact the N400 response to varying levels of CP differentially. Specifically, participants with more efficient SL ability should show a more efficient information extraction and a greater learning gain. This difference between more and less efficient learners should manifest as a larger N400 difference between the low-CP and high-CP conditions than between the intermediate-CP and high-CP conditions. As online (during-learning) measures are more direct reflections of SL capabilities than offline (after-learning) measurements (e.g., Lukács et al. 2021), we expected online indices to have a larger effect on N400 response than their offline counterparts. Finally, since both language and statistical learning abilities are influenced by WM and processing speed, we expect the relationship between SL and semantic processing to be partially but not entirely mediated by individual differences in the above two cognitive capacities. The study was preregistered at https://doi.org/10.17605/OSF.IO/A6HS4.

2 | Methods and Procedure

The detailed task descriptions, experiment materials, model building procedures, and full results can be found at https://osf.io/f2r4n/. The study was approved by the Hungarian United Ethical Review Committee for Research in Psychology (EPKEB-2018/87).

2.1 | Participants

Twenty-nine students of the Budapest University of Technology and Economics took part in the study (22 females, age: 19–42 years, $M = 25.15$, $SD = 7.4$), all of whom received course credits in return for their partaking. One participant reported having dyscalculia, while another reported being late bilingual, but neither of them was excluded from the study, assuming that these circumstances are unrelated to the cognitive abilities investigated here.

2.2 | EEG Experiment

2.2.1 | Stimuli

For the EEG study, we have created a stimulus set of Hungarian sentences varying in CP on three levels, using an innovative procedure to generate highly controlled sentences ending on identical target words. Identical target words across conditions can be presented such that each participant views an identical set of target words, but only once, and each in different conditions, thereby controlling for all possible psycholinguistic features of target words across conditions. While sentences with varying CP are typically constructed with the final word missing and CP established by sentence completion tasks, we constructed sentences in sets of three, ending on the same target word (e.g., bike), with varying expected CP according to three conditions (high, intermediate, and low CP)—and then validated our sentences using typical CP questionnaires. To make sure that we elicit a typical N400 in our participants, we used strongly constraining sentence contexts for our high- and low-CP conditions as well, with a clear violation of expectation in the latter case. Strongly constraining context sentences raise expectations to a single most likely sentence completion, as in for example, "Due to the tire puncture the postman found it hard to push his ..." (i.e., bike). In contrast, weakly constraining sentences have multiple, similarly likely completions, as in for example, "The local courier's vehicle is his trusty ..." (e.g., van, car, scooter, or bike). Sentential constraint has no impact on the N400, only CP, meaning that unexpected endings elicit an N400 in both constraint types, however, expected endings of sentences with low constraint evoke an intermediate N400 response (Federmeier 2007). For this reason, we used weakly constraining sentences ending on one of the expected final words for our intermediate-CP condition.

As a first step, we generated 360 novel sentences, in sets of three ending on the same word, in accordance with our three conditions. Next, we assessed their CP using sentence completion questionnaires; based on the outcomes, we rephrased sentence frames that did not yield the expected likelihoods for sentence final words, and then ran another round of sentence completion questionnaires. Eventually, we selected 67 sentence triplets ending on the same noun that fulfilled all our requirements (201 sentences in total). An additional 44 pairs of sentences were also selected where the same sentence ending target word appeared in two conditions out of the three (88 additional sentences), and created 11 more sentences to have 300 sentences in total, in three conditions:

1. High-CP: Strongly constraining sentences that end on the most expected, highest CP final word ($n = 108$). The minimum CP value was 80%, and all sentence final target words were the most likely completions (e.g., "One of our most common seasonings is the iodized sea *salt*.")
2. Int-CP: Intermediate CP of weakly constraining sentences, ending on one of their expected endings ($n = 89$). Neither of the endings were highly likely, yet all of them were plausible and somewhat expected: the target word's CP was required to be between 20% and 55%. In 45 sentences the sentence ending word was the most likely, in 40 sentences the second, and in 4 sentences the third most expected ending (e.g., "For a good cake batter, you will need a little *salt*.")
3. Low-CP: Strongly constraining sentences that end on an unexpected final word ($n = 103$). These sentence final words were also plausible completions of the sentence frames, and their CP were required to be as low as possible, below 20%. (e.g., "The streets in January were covered in white *salt*.")

Each stimulus item was a simple declarative sentence made up of six units. Words accompanied by articles, prefixes, or suffixes were taken as a single unit. Each sentence ended on a target word presented in itself. The target word was always an uninflected noun, most of the time preceded by an adjective unit, as part of an *adjective + noun* expression. Each target word was presented only once for each participant, and care was taken that no target word featured in any other sentences.

CPs of sentences were collected through online sentence completion questionnaires (Taylor 1953), where participants were asked to give the most probable ending for open-ended sentences (e.g., 'The cheap ball-point pen quickly ran out of blue ________'). A total of 95 students of Budapest University of Technology and Economics took part in the sentence completion study (82 female, age: 19–46, $m = 22.7$, $SD = 4.6$). They received course credits for their participation. Every participant received a unique pseudo-randomized questionnaire, ensuring that participants saw only one sentence with the one particular intended target ending. We obtained a minimum of 29 answers for each sentence. Based on these answers, we calculated CPs in the following way: $CP = (\text{instances of a given answer word})/(\text{number of all answers}) \times 100$. Next, based on the results, we rewrote sentences with unsatisfactory CP ratings and ran a second sentence completion study, with a different group of participants

(116 students, 92 females, age: 19–59, $m = 25.7$, $SD = 8.9$) In the final stage, we excluded sentences that did not meet our criteria ($n = 60$), selected 201 sentences ending on the same target word in all three conditions and 88 sentences ending on the same target in two conditions, and generated 11 more sentences with unique target words (4 High-CP, 1 Int-CP, 6 Low-CP), to have 300 sentences in total. Sentences were then assigned into six different pseudo-randomized, 120-sentence-long lists, so that each list contained a closely matched number of sentences from each category, while each sentence ending target word appeared in every list only once. For the full list of sentences and further details on the stimulus production and selection, see Data S1 (https://osf.io/4nm8v).

2.2.2 | EEG Experimental Procedures

In the EEG experiment, participants read sentences presented on a computer screen in a dimly lit room, with one word/unit at a time, while their ERPs were recorded. In each trial a single sentence was presented. Trials started with a fixation cross, displayed for 500 ms at the center of the screen, with sentence units (words together with pronouns, prefixes and suffixes) presented centrally one at a time, for 350 ms with 150 ms blank screen in between, using yellow (Arial 32 point) font over a gray background. Every tenth sentence trial was followed by a yes-or-no question about the contents of the preceding sentence, to ensure sustained attention. For instance, the sentence "The scarf was knitted for the grandson by his loving grandma" was (sentences were presented in Hungarian, for a full list, see Data S1: https://osf.io/4nm8v) followed by the question "Did the grandma knit a scarf?". Participants were asked to respond to these questions by pressing the designated key on a keyboard.

2.2.3 | EEG Recording and Data Analysis

EEG recording was conducted with BrainVision Recorder (Brain Products GmbH., Munich, Germany) over 32 Ag/AgCl electrodes (64 channels EasyCap, EasyCap GmbH, Herrsching-Breitbrunn, Germany); 1000 Hz sampling rate; impedance $< 5\,k\Omega$. Electrodes were placed according to the extended 10/20 international electrode system (Chatrian et al. 1985), with 4 additional electrodes used for the registration of the eye movements. EEG analysis was performed using a custom-written Matlab script (Mathworks, Natick Massachusetts, USA), based on EEGLAB functions (Delorme and Makeig 2004). The original dedicated reference channels of the Brain Products GmbH system, the average of left and right mastoid electrodes (M1/M2) were used as reference. A detailed description of recording and preprocessing, together with the analysis scripts used, is available in Data S1 (https://osf.io/4nm8v).

Data preprocessing and artifact rejection was conducted in four phases. First, the continuous EEG data were visually inspected to remove channels which contained non-stereotyped artifacts that would affect the quality of ICA (Independent Component Analysis) decomposition (Mognon et al. 2011), applied in the subsequent step. EEG was band-pass filtered offline (0.1–40 Hz) using zero-phase shift forward and a reverse-IIR Butterworth-filter. A 48–52 Hz Parks-McClellan stop-band Notch-filter in ERPLAB (Lopez-Calderon and Luck 2014) was used to remove electric-interference from the 50 Hz line. Second, the data were subjected to ICA decomposition using the EEGLAB function "cudaica" (Raimondo et al. 2012) to extract independent components based on the segmented EEG records. Third, based on the extracted independent components, to identify and remove stereotyped artifact-related components (e.g., eye-blinks and eye-movements) from the EEG, we applied ADJUST (Mognon et al. 2011). In case of segmented data, ADJUST conducts an analysis based on consecutive 900 ms windows for the purpose of computing the temporal features (Mognon et al. 2011). Fourth, noisy channels were rejected and interpolated using a spherical method. After all artifact-removal, stimulus-locked epochs were derived by segmenting the EEG data into 900 ms windows, starting from 100 ms pre-stimulus. Segments were baseline-corrected over a 100 ms pre-stimulus window and then averaged for each participant in each conditions, to obtain the ERP waveforms. Mixed-effects modeling (Baayen 2008; Baayen et al. 2008) analysis of the ERP data was carried out over the mean microvolts for each trial of each participant averaged in the N400 time window (300–500 ms) over a centro-parietal region-of-interest (ROI: Cz, C3, C4, CP5, CP1, CP2, CP6, P3, Pz, P4), following Kutas and Federmeier (2011).

2.3 | Statistical Learning Measures

2.3.1 | Word Segmentation Task

To assess participants' SL ability, we first used a modified version of the study of Saffran et al. (1996). The stimuli consisted of twelve consonant-vowel syllables generated from Hungarian phonemes (*cé*/tseː/, *csa*/tʃɑ/, *ha*/hɑ/, *gá*/gaː/, *gye*/ɟɛ/, *jü*/jɥ/, *lo*/lo/, *pe*/pɛ/, *ri*/ri/, *sö*/ʃø/, *tu*/tu/, *vi*/vi/) and combined into four trisyllabic words (*csagyejü*/tʃɑɟɛjɥ/, *cévigá*/tseːvigaː/, *lohari*/lohɑri/, *söpetu*/ʃøpɛtu/). Syllables were 270 ms long with 30 ms ISIs between them and had a uniform pitch of 215 Hz.

2.3.1.1 | Online Training Phase. During an online training phase, we assessed statistical learning by looking at how effectively participants can process upcoming elements in structured speech streams. They were presented with speech streams consisting of the four words with no pauses or other cues for word boundaries between words. Thus, in the streams, syllables within words followed each other with a transition probability of 1 ($TP = 1.00$), while transition probability on word boundaries was 0.33 ($TP = 0.33$). The training phase consisted of three training blocks (blocks TRN1 to TRN3, 15×4 words in a pseudorandom order), a fourth random block (RND4) where the twelve syllables were presented without being combined into words (15×12 syllables in a pseudorandom order), and finally, a recovery block (REC5, 15×4 words in a pseudorandom order). A target element was assigned to each participant (the last syllable of a word, counterbalanced between participants) which participants had to detect by pressing a button throughout the five blocks of the training.

2.3.1.2 | Offline Two-Alternative Forced Choice Task (2AFC Task). In the 2AFC task, we used trigrams and bigrams which were more or less structured based on the training material. Structured trigrams were words (with

TP = 1.00 between all syllables based on the training streams), structured bigrams were bigrams from words (with TP = 1.00 between the two syllables), partly structured trigrams were part-words (spanning a word boundary in the stream, with a TP = 1.00 and a TP = 0.33), and partly structured bigrams were bigrams from part-words (spanning a word boundary in the stream, with TP = 0.33 between the two syllables). Unstructured trigrams were non-words obtained by reordering the syllables (with TP = 0.09 between all syllables), and unstructured bigrams were bigrams from non-words (with TP = 0.09 between the two syllables). In each trial, participants were offered a more and a less structured trigram or bigram (e.g., a trigram word versus a trigram part-word, or a bigram part-word versus a bigram non-word) and had to make a familiarity decision based on the training material ("Which sequence was more similar to the previously presented speech stream?"). Altogether, the task consisted of 36 trials.

2.3.1.3 | Production Task. In the production task, we used the four words from the training phase. In each trial, a syllable was missing in one of the three positions from a word, and participants had to choose the missing element in a 3AFC task. There were 12 trials altogether.

2.3.2 | Artificial Grammar Learning (AGL) Task

As another assessment method for statistical learning, we used a modified version of the AGL task of Knowlton and Squire (1996). The grammar used in the present task is illustrated in Figure 1. The vocabulary of the grammar consisted of five consonant-vowel-consonant nonwords made up from Hungarian phonemes (*péf*/peːf/, *rud*/rud/, *gán*/gaːn/, *ket*/kɛt/, *zot*/zot/). They had a uniform pitch of 125 Hz, the duration of words was 400 ms and they were presented with an ISI of 50 ms within sentences and 750 ms between sentences. Altogether, 34 grammatical strings were generated following the rules of the grammar, yielding 3 three-word sentences, 6 four-word sentences, 5 five-word sentences, 7 six-word sentences and 13 seven-word sentences. 8 sentences were selected for test strings only used in the offline tasks, and the remaining 26 strings were used in the online training phase.

2.3.2.1 | Online Training Phase. The online training phase consisted of five training blocks: Blocks TRN1 to TRN3 and Block REC5 contained the same training sequences, while block RND4 contained 26 random sequences which were generated by arranging the words of the grammar in a pseudo-random order. Sequences within all blocks were presented in a randomized order. A target element (*péf*/peːf/) was assigned to all participants, which they had to detect by a button press throughout the five blocks of the training.

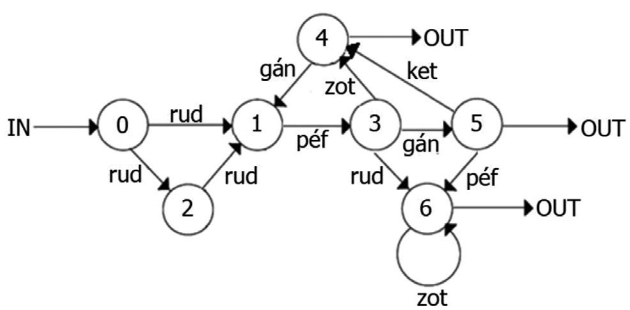

FIGURE 1 | The artificial grammar used in the study (after Knowlton and Squire 1996).

2.3.2.2 | Offline Two-Alternative Forced Choice Task (2AFC Task). In the 2AFC task, we used sentences, trigrams and bigrams which were (1) well-formed versus ill-formed, or (2) more frequent versus less frequent in the training material. In each 2AFC trial, participants were offered a more and a less structured sentence, trigram or bigram and were asked to make a grammaticality decision based on the training material ("Which sequence was more similar to the previously presented language?"). In the sentence trials, 12 trials contained transition violations (involving bigrams that did not occur in the training materials) in the ungrammatical sentence. These constituted transition probability (TP) based trials. In 4 sentence trials, no transition violations were present, so participants had to rely on the rules of the grammar to make their decisions; these were classified as rule based trials. In the bigrams and trigram conditions, 12 trials could be resolved based on transition violations, (TP based trials), while 10 trials required reliance on bigram and trigram frequency in the training material (frequency based trials). Altogether, there were 38 trials.

2.3.2.3 | Production Task. In this phase, participants saw incomplete sentences on the computer screen, and they had to choose the missing element in a 3AFC task. The 24 test trials contained eight grammatical trigrams and the eight test sentences. Each test sentence occurred in two trials, while the trigrams were presented only once. Trials varied in the position and the identity of the missing item within the sequence.

2.3.3 | Statistical Learning Indices

Statistical learning indices were calculated in two steps. As the first step, we calculated indices from the online and offline tasks in both the speech segmentation and AGL tasks. From the online target detection tasks, we calculated *SEGM* and *AGL median RT difference*: [((Block RND4 median RT—Block TRN3 median RT) + (Block RND4 median RT—Block REC5 median RT))/2]. The following offline indices were calculated from the 2AFC tasks: 2AFC accuracy rates in the speech segmentation task in trigram and bigram trials, yielding *SEGM 2AFC bigram* and *SEGM 2AFC trigram*, 2AFC accuracy in the AGL sentence rule- and TP-based trials, yielding *AGL 2AFC sentence rule* and *AGL 2AFC sentence TP*, 2AFC phrase accuracy in the AGL frequency- and TP-based bigram and trigram trials, yielding *AGL 2AFC phrase frequency* and *AGL 2AFC phrase TP*, while from the offline production tasks the mean accuracy rate of 3AFC trials was calculated for both the speech segmentation and AGL tasks, yielding *SEGM production* and *AGL production*. As a second step, based on the results of a factor analysis in another study (Lukics et al. 2021), we combined the statistical learning measures into two scores (the *AGL median RT difference* index was omitted because of its low level of reliability, and the two *AGL 2AFC phrase* indices were omitted because they did not converge with other measures of statistical learning, see Data S1 for further description: https://osf.io/4nm8v):

1. Offline statistical learning (SL-OFF): mean of *SEGM 2AFC bigram, SEGM 2AFC trigram, AGL 2AFC sentence rule, AGL 2AFC sentence TP, SEGM production*, and *AGL production*.
2. Online statistical learning (SL-ON): *SEGM median RT difference*.

2.4 | Tasks Assessing General Cognitive Capacities

2.4.1 | N-Back Task

In the n-back task, different letters were presented individually in the middle of the screen, and participants were instructed to press the spacebar for items identical to the one presented one, two or three steps earlier (in the 1-back, 2-back, and 3-back conditions, respectively), and button 'A' for any other stimulus. The task consisted of four (one 1-back, two 2-back, one 3-back) blocks, and each block consisted of 60 trials: 10 targets, 10 lures and 40 foils. Participants were instructed to give a response for each trial. Blocks started with the presentation of a fixation cross for 1000 ms, which was followed by the presentation of the individual trials for 500 ms, with an ISI of 1500 ms.

2.4.2 | Processing Speed

2.4.2.1 | Reaction Time (RT). We used a reaction time task that measured the speed of stimulus processing in the acoustic and visual modalities. In the first, visual block participants were instructed to look at a blank screen and focus on the appearance of an image on the screen, and to press the spacebar as fast as they could each time when the image appeared. The visual block was followed by an acoustic block, in which participants had to focus on a tone presented via headphones. For both modalities, 32 trials were presented. ISIs varied randomly between 1223 and 4988 ms for the visual modality, and 1041 and 4883 ms for the auditory modality. Trials were presented in a randomized order.

2.4.2.2 | Decision Speed. We used a same/different decision task to further assess visual and acoustic processing speed. In the visual condition, we used a pair of visual gratings and participants had to decide whether the orientation of two gratings were the same or different. In the auditory condition, we used pure tones (73 tones between 150 and 870 Hz), and participants had to decide whether the tones were identical or different. There were 146 trials altogether in this task, and participants had to complete as many trials as they could in 60 s. Trials were presented in a randomized order. In both decision speed tasks, participants below 0.75 accuracy were excluded from the analysis.

2.4.3 | Indices of Working Memory and Processing Speed

Participants' WM capacity was estimated by a composite index from the n-back task, calculated as the mean of the d-primes from the 2-back 3-back blocks (as we did not expect the scores in the 1-back block to exhibit variability resulting from individual differences of WM in the case of adult participants). Processing speed indices were also calculated: based on the results of the earlier factor analysis, we calculated the visual processing index as the mean of visual RT (median RT of trials) and visual decision speed RT scores (median RT of correct trials). Auditory processing speed was calculated as the median RT of trials, while auditory processing decision speed was calculated as the median RT of correct trials.

2.5 | Statistical Analyses

Statistical tests were conducted over single trials using linear mixed-effects modeling (Baayen 2008; Baayen et al. 2008), in R (Core Team R 2017) and using the lme4 package (Bates 2016) with trial-level N400 amplitudes in the 300–500 ms time window averaged over the centro-parietal ROI as the dependent variable, with CP-Category (with three levels, high, intermediate, and low), SL indices (online and offline separately, to avoid collinearity, as these measures correlate) as main, and working memory and processing speed (auditory and visual) as control fixed-effect predictors. Since there is no clear agreement for keeping predictors maximal (Barr et al. 2013) or parsimonious (Bates et al. 2015) (i.e., building models bottom up to avoid overfitting vs. top-down to avoid underfitting), we started with a basic model that included CP-Category and SL as fixed effect predictors and introduced a maximal random effect predictor structure (of slopes and intercepts) initially that was reduced stepwise if the model failed to converge. Eventually, no random slopes and intercepts were possible to include in either of the models.

Outlier data points > 3 SD away from each individual's mean ERPs were removed prior to statistical analyses to conform with mixed-effects modeling presumptions. Model building for each research question followed a multistep process with introducing predictors of N400 amplitudes stepwise, in a gradual fashion. As a preparatory step, we first introduced the order of trials (*trial*) as a fixed effect and simple random effects of *participant ID* and *item ID* (i.e., identifiers of target nouns), against a model of random effects only, to check whether responses were modulated by fatigue. It was included in subsequent models only if it significantly improved them, as a fixed effect.

The baseline model was set up to predict N400 responses by CP-Category (High-CP, Int-CP, Low-CP) as a fixed effect, and *participant ID* and *item ID* (i.e., identifier of target nouns) as random effects [i.e., baseline_N400 = lmer(N400 ~ CP-Category + (1|ParticipantID) + (1|ItemID))]. Next, SL scores for online and offline SL performance were introduced to two separate models. As a last step, indices of cognitive abilities (i.e., measures of processing speed and WM) were also introduced as fixed effect predictors. Where appropriate, the measure of the extent of the model fitting the data, AIC, served as the basis to compare models and selected the better fitting one. We present here below only the statistical outcomes of the final models, and only those models where covariates were found to have a significant effect. Model residual plots did not exhibit visible deviations from normality

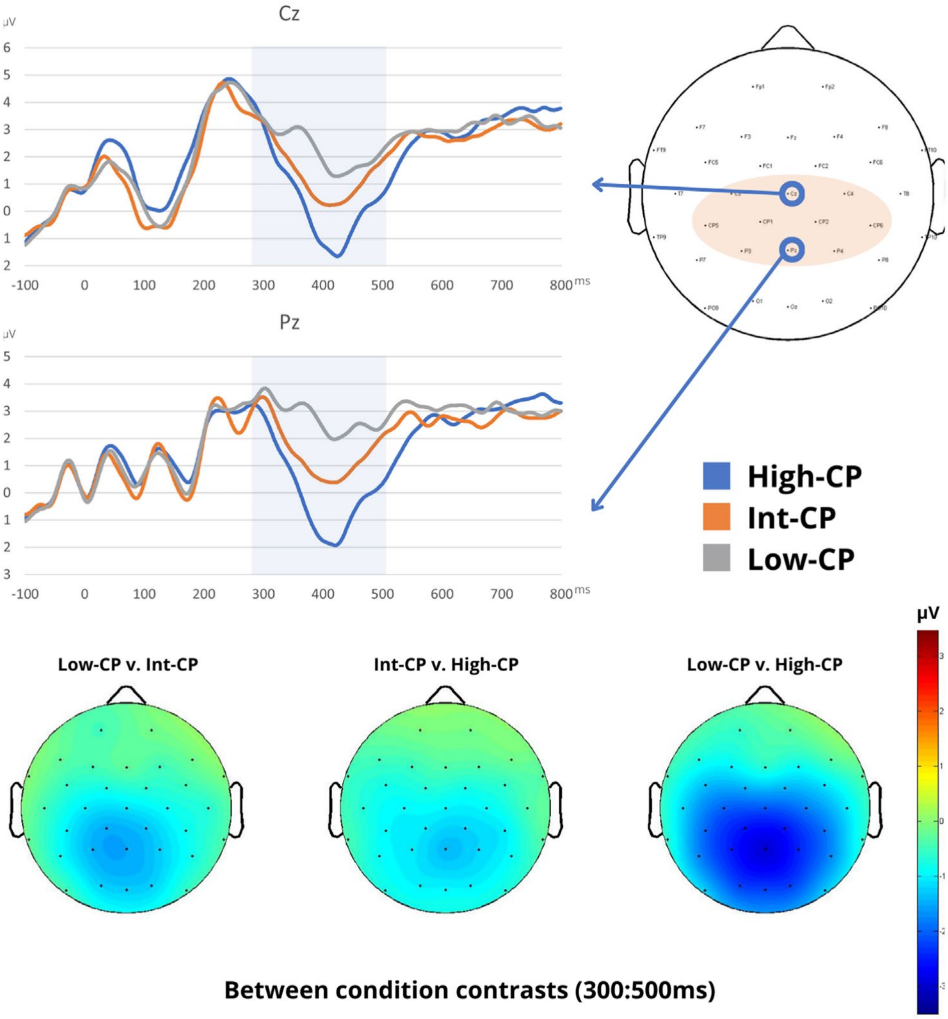

FIGURE 2 | The evoked N400 ERP responses across the three CP-conditions over two prominent exemplar electrodes. Scalp topographical maps show contrasts between CP-conditions in the N400 time-window.

and homoscedasticity. To calculate the reported p-values, the Kenward–Roger approximation was used from the mixed() function (Singmann and Kellen 2017). For better readability, we report in the main text only F-values and p-values (but not beta coefficients and their estimated error).

3 | Results

3.1 | The Effect of Sentence Ending Predictability on the N400

In the baseline model, to confirm the success of our N400 manipulation, we tested the effect of CP-category on N400 amplitudes [N400 ~ CP-category + (1|participant) + (1|trial)]. The analysis revealed that the CP-category had a significant effect ($F[2, 1077] = 20.9$, $p < 0.001$); to explain it, we conducted pairwise comparisons. We found a significant difference between High-CP versus Low-CP, $F(1, 1474) = 41.2$, $p < 0.001$, a typical N400 effect; and also between High-CP versus Int-CP: $F(1, 425) = 8.11$, $p = 0.005$; and Low-CP versus Int-CP: $F(1, 549) = 12.3$, $p < 0.001$ as well, demonstrating a graded N400 response (Figure 2).

In an exploratory analysis, we ran our main tests with target word CP-values included as a continuous variable instead of the categorical CP-variable, using the Z-transformed CP scores of sentence ending target words, which yielded the same pattern of results: independent main effects, with no interaction between the two. (For the details, see Data S1: https://osf.io/4nm8v).

3.2 | Statistical Learning and Sentence Ending Predictability Effects on the N400

Next, we analyzed the effect of CP-categories and online statistical learning (SL-ON) on N400 amplitudes (N400 ~ CP-category × SL-ON + (1|participant) + (1|trial)). The analysis revealed a main effect of both CP-category, $F(2, 1076.4) = 20.93$, $p < 0.001$, and SL-ON, $F(1, 27.0) = 5.43$, $p = 0.027$; however, the interaction between the two terms was not significant, $F(2, 3383.2) = 0.09$, $p = 0.911$. Follow-up pairwise comparisons of CP-category revealed the same graded N400 response pattern (High-CP vs. Low-CP: $p < 0.001$; High-CP vs. Int-CP: $p = 0.005$; and Low-CP vs. Int-CP: $p < 0.001$). To further examine the independent main effects, we tested for correlation between SL-ON and the typical N400 effect (calculating the difference between

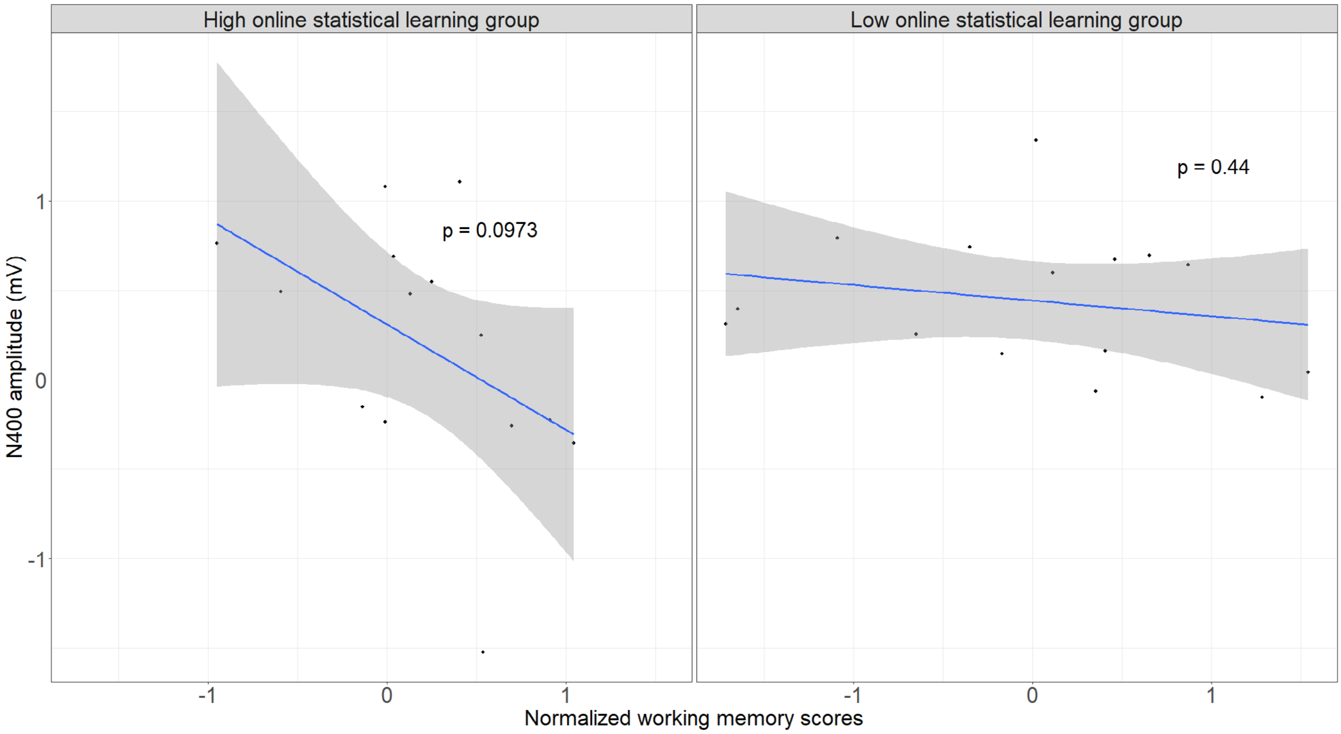

FIGURE 3 | The interaction of SL-ON and WM on N400 responses. In the low SL-ON group, WM performance had no effect on N400 responses, while the analysis revealed a marginally significant effect of WM on N400 responses in the high SL-ON group.

ERPs to Low-CP vs. High-CP sentences for each individual) and found no significant correlation ($r = 0.036$, $p = 0.85$) between the two.

We ran the same analysis with SL-OFF as above, which revealed again a significant main effect of CP-category ($F(2, 1111) = 20.98$, $p < 0.001$), but no main effect of SL-OFF, $F(1, 29) = 0.04$, $p = 0.853$, and no interaction between the predictors either, $F(2, 3660) = 0.25$, $p = 0.780$. Pairwise comparisons of CP-categories revealed the same pattern of results again (High-CP vs. Low-CP: $p < 0.001$; High-CP vs. Int-CP: $p = 0.008$; and Low-CP vs. Int-CP: $p < 0.001$). Relative to the SL-ON, for the SL-OFF indices, we had data from two additional individuals (sample size $n = 31$); however, excluding these two participants from the analysis did not change the above pattern of results.

3.3 | Effects of Working Memory and Processing Speed

In the following analyses, we focused on the potential explanatory effect of WM and processing speed. In the following analyses, we did not include the CP-category × SL-ON interaction term, to avoid overfitting, as the interaction was not significant in the above analyses. WM was entered in the models in interaction with both CP-category and SL-ON, because it was expected to influence both factors. Auditory reaction time and visual processing speed were entered in interaction only with SL-ON.

3.3.1 | Working Memory

The analysis of the effect of WM (N400 ~ CP-category × WM + SL-ON × WM + (1|Participant) + (1|Item)) revealed a significant main effect of WM ($b = -0.28$, SE = 0.13, $F(1, 25.0) = 4.62$, $p = 0.041$) on the N400, while both CP-category ($F(2, 1086.4) = 20.27$, $p < 0.001$) and SL-ON ($F(1, 25.1) = 5.42$, $p = 0.028$) remained significant. The interaction between WM and SL-ON also turned out to be significant ($F(1, 25.0) = 5.26$, $p = 0.030$), but there was no interaction between CP-category and WM ($F(2, 3423.2) = 0.18$, $p = 0.836$). Follow-up analyses revealed that all CP category comparisons remained significant.

In order to explain the interaction of SL-ON and WM, we median split the data by SL-ON (into low SL-ON and high SL-ON subgroups). The analysis revealed a marginally significant effect of WM on the N400 amplitudes in the high SL-ON group ($F(1, 12.0) = 3.24$, $p = 0.097$) but no effect of WM in the low SL-ON group ($F(1, 13.0) = 0.63$, $p = 0.442$). Results of the analyses in the case of high SL-ON and low SL-ON are illustrated in Figure 3.

3.3.2 | Auditory Reaction Time

The auditory reaction time (aRT) model (N400 ~ CP-category + SL-ON × aRT + (1|Participant) + (1|Item)) revealed the significant main effect of CP-category ($F(2, 1077.9) = 20.88$, $p < 0.001$) and of SL-ON ($F(1, 25.0) = 6.61$, $p = 0.016$), with no main effect of aRT or any interactions between aRT and SL-ON ($b = 0.05$, SE = 0.17, $F(1, 25.0) = 0.08$, $p = 0.785$) found. CP-category comparisons all remained significant (High-CP vs. Low-CP: $p < 0.001$; High-CP vs. Int-CP: $p = 0.005$; and Low-CP vs. Int-CP: $p < 0.001$).

3.3.3 | Visual Processing Speed

Entering visual processing speed (VPS) in the model (N400 ~ CP-category + SL-ON × VPS + (1|Participant) + (1|Item)) revealed again a significant main effect of CP-category ($F(2, 1077.9) = 20.89$, $p < 0.001$) and SL-ON ($F(1, 25.0) = 5.52$, $p = 0.027$), while no effect of VPS or any interactions were found. Differences between CP-categories remained the same.

3.3.4 | Exploratory Analyses of Specific SL Indices

In an exploratory follow-up analysis, we examined whether rule-based and chunk-based SL can be dissociated, and whether segmentation and artificial grammar learning (AGL) tasks engage different underlying mechanisms; and if so, whether this distinction is reflected in different patterns of relationship with N400 responses. Results showed partial divergence of specific indices in their association with the ERP signal: the aggregated SEGM-OFF score, the SEGM production, the aggregated AGL-OFF, and the AGL 2AFC sentence rule indices showed an interaction with CP-Category: the SEGM-OFF and SEGM production indices showed a stronger negative, while the AGL-OFF and AGL 2AFC sentence rule indices showed a stronger positive relationship with N400 responses in the Low-CP condition compared to the High-CP condition. However, the interaction between SL and condition did not reach statistical significance in the case of AGL-OFF and AGL 2AFC sentence rule. Due to the limited number of trials contributing to these measures and suboptimal model diagnostics, these findings are reported in Data S1 (https://osf.io/4nm8v), where they are intended to inform future research rather than to serve as the basis for firm conclusions, and are therefore not discussed further in the main text.

4 | Discussion

In this study, we set out to investigate whether the sensitivity of the N400 ERP response to statistical linguistic information (e.g., Molinaro et al. 2010; Laszlo and Federmeier 2011) involves statistical learning mechanisms in predictive processes during sentence comprehension (cf. Rabovsky et al. 2018). The purpose of our experiment was twofold: first, we aimed to examine what role statistical learning abilities may play in N400 amplitudes evoked by the processing of sentences of varying predictability, and second, to provide a clearer characterization of the mediating effects of cognitive capacities related to statistical learning (such as working memory, auditory and visual processing speed) in language comprehension and specifically the mechanisms behind the N400 ERP response. We predicted that the amplitude of the N400 response may be affected by individual differences in statistical learning abilities, based on reported associations between SL and various behavioral aspects of language processing (Christiansen et al. 2012; Conway et al. 2007, 2010; Mainela-Arnold and Evans 2014; Daltrozzo et al. 2017; Kidd 2012; Kidd and Arciuli 2016; Misyak and Christiansen 2012; Torkildsen et al. 2019; Arciuli and Simpson 2012; Spencer et al. 2015) and by recent results and theoretical models emphasizing the role of predictive processes in language comprehension (Federmeier 2007; Pickering and Gambi 2018).

We found that online SL abilities influence the N400 response. This effect is statistically independent from the classical N400 manipulation of varying sentence ending predictability. We found no evidence for the modulation of the N400 by auditory/visual processing speed, and working memory only had an influence on it in learners with high online statistical learning performance. These results corroborate the observation that the N400 is likely a multi-faceted, multi-generator ERP, and that SL abilities are associated with N400 amplitudes on an individual level.

These significant, but statistically independent main effects of SL-ON and CP are consistent with the view that the N400 reflects the contribution of multiple predictive processes. One of these relies on the statistical learning abilities of the individual to track transitional probabilities in the input, while the other involves higher-level representational structures (i.e., world knowledge) reflected in sentence-ending predictability as indexed by CP. These predictive processes may be understood in more than one way.

Although we do not assume that prediction is a necessary component of statistical learning (SL), we consider it likely that predictive mechanisms were engaged during the sequential SL tasks employed here, particularly as reflected in the online measures, which may explain its connection to the N400. This interpretation is supported by previous findings showing that SL can give rise to predictive processing in similar linguistic task contexts (e.g., Batterink et al. 2015; Karuza et al. 2014). At the same time, SL does not inherently involve prediction or anticipation, especially in the case of non-sequential visual statistical learning: the relationship between SL and predictive processing likely varies depending on task characteristics and cognitive domain (e.g., Siegelman et al. 2018; Fiser and Aslin 2001; Turk-Browne et al. 2005).

While the effects of SL-ON and Cloze Probability appeared to be statistically independent, as indicated by the absence of interaction or correlation, this does not necessarily imply that separate predictive mechanisms are at play. One possible interpretation of semantic predictability, as indexed by CP, is that it is influenced by statistical regularities inherent to language (Conway et al. 2010; Miller and Selfridge 1950). In this light, SL-ON and CP may reflect different kinds of probabilistic predictability. This perspective aligns with probabilistic models of language processing (Kuperberg and Jaeger 2016) which suggest that predictive processes operate at multiple levels within the language system.

Further support for this integrated view comes from recent advances in generative AI, where models trained on large-scale language corpora demonstrate that sensitivity to distributional regularities is sufficient to produce contextually coherent and semantically meaningful output (Radford et al. 2019; Contreras Kallens et al. 2023). Such findings suggest that even complex semantic processes can emerge from predictive mechanisms grounded in statistical associations. This view is consistent with Rabovsky et al.'s (2018) Sentence Gestalt Model, which conceptualizes the N400 as reflecting continuous updates to probabilistic semantic representations based on incoming linguistic input.

The absence of a significant interaction between the effects of SL-ON and CP-controlled sentence ending predictability does not necessarily imply that the two effects are functionally unrelated or independent: expectations during language comprehension may be shaped by SL-based processes, which, over time, shape expectations during sentence comprehension, suggesting some degree of connection between the two, even if this relationship is not directly reflected at the level of online sentence processing. It is in this sense that N400 amplitudes may be influenced both by the predictability of individual lexical items (CP main effect) and by individual differences in sensitivity to the distributional structure (SL main effect) of language.

Within the framework of the Sentence Gestalt Model (Rabovsky et al. 2018), the N400 reflects the update of an internal probabilistic representation during comprehension. In this context, the magnitude of the update and the amplitude of the N400 response may be shaped by both external input (e.g., predictability as captured by CP) and internal factors (e.g., the comprehender's statistical learning ability). While Rabovsky et al. argue that N400 amplitude can, in principle, be accounted for by statistical learning-based prediction, the independent effect of CP observed here suggests that additional mechanisms may be involved. One possibility is that SL-ON and CP index distinct, though potentially complementary, predictive processes: SL-ON may reflect sensitivity to local transitional probabilities, while CP captures the influence of higher-level contextual or semantic constraints.

Another possible interpretation is that while the N400 is influenced by statistical information, other generators process complex semantic knowledge, irrespective of statistics. The SL main effect captures the former, while CP effects are also influenced by the latter. For example, semantic taxonomy, the complex hierarchical organization of semantic concepts, is known to influence N400 amplitudes (Federmeier and Kutas 1999). Within-category violations of expected sentence endings (e.g., "Along the highway they planted rows of pines" as opposed to the expected "palms", in San Diego) elicit smaller N400 responses compared to between-category violations (e.g., definitely unexpected "tulips"). This effect can be dissociated from association- (i.e., frequency-)based thematic information, which is a raw-material of statistical learning (Honke et al. 2020). World-knowledge also directly impacts the N400 amplitude (Hagoort et al. 2004), even of fictive worlds (Troyer et al. 2020; Troyer and Kutas 2020). The N400 responds differentially to the typicality of the location (e.g., "arena") of actions ("skating") in different past tenses (Ferretti et al. 2007), and it is impacted by quantifiers such as most/some (Urbach and Kutas 2010), and pronouns such as a/an, as well (DeLong et al. 2005; Urbach et al. 2020). While prediction likely plays a role in CP, our findings suggest that the N400 cannot be fully reduced to prediction as defined by anticipatory pre-activation. Not all forms of prediction are rooted in statistical learning; some may arise from explicit knowledge or isolated exposure. Moreover, the ongoing debate on whether the N400 reflects expectancy, prediction, surprisal, integration or contextual facilitation (e.g., Hodapp and Rabovsky 2021; Krieger et al. 2024; Lai et al. 2021; Rabovsky et al. 2018; Szewczyk and Federmeier 2022) lies beyond the scope of this study. Importantly, the statistical independence of SL-ON and CP effects indicates that some generators of the N400 may engage semantic knowledge in ways that are not strictly statistical or predictive in nature.

The absence of a statistically significant SL-ON × CP-measured predictability interaction may also stem from our choice of SL measures. Given that SL is a componential capacity of multiple abilities which are known to vary across modalities even within individuals, for operational reasons, SL is best described as a factor of multiple different measures. The inherent complexity and variability of SL components, coupled with the fluctuating reliability of assessment tools, complicate the formulation of a comprehensive SL metric that accurately reflects an individual's complete set of SL capabilities. Consequently, it is plausible that the SL indices employed in our study failed to capture certain statistical predictive cognitive functions that were incidentally assessed by CP metrics.

In sum, our findings suggest that N400 amplitude reflects the combined influence of multiple processes, which may operate at different representational and computational levels, supporting the multi-componential view of this neural response. While SL-ON and CP-based predictability effects emerged independently, this pattern likely reflects the complexity of both statistical learning and predictive mechanisms in language comprehension, rather than a strict functional separation.

In hindsight, the absence of a significant effect of SL-OFF performance on N400 response size was not surprising in the light of criticism regarding offline SL measures (e.g., Isbilen et al. 2017). It is increasingly apparent from the SL literature that SL is not a monolithic, unified ability, but rather constitutes a componential capacity of multiple facets or even a collection of various abilities. This implies that sensitivity to environmental probabilities may vary across different modalities even on the level of the individual. Consequently, SL ability may be best described by a factor of performance indices on multiple tasks measuring implicit abilities. Certain tasks and indices employed to measure statistical learning exhibit larger consistency and reliability in representing individual abilities than others. Offline measures provide information on the outcome of the learning process, and mostly rely on grammatical well-formedness judgments that participants have to make after the training phase. On the other hand, online tasks and measures, with metrics such as reaction time and accuracy, allow the researcher to tap into the learning process as it is ongoing (Lukács et al. 2021), arguably providing more reliable and detailed indices than offline methods (Batterink et al. 2015; Lammertink et al. 2019; Lukics and Lukács 2021). The lack of a significant association between SL-OFF and N400, both in the primary model and with the addition of general cognitive covariates, and the presence of such associations with SL-ON measures offer further support for the notion that online measures provide more sensitive indices of SL abilities than offline ones.

As predicted, the effect of SL-ON on N400 was modulated by WM performance, which also had a significant effect on N400 responses. Additional analysis into this interaction revealed that in participants with better SL-ON capabilities, higher WM performance was associated with a larger evoked N400 response. These findings further refine those of Van Petten et al. (1997), who demonstrated that in sentence-level semantic context, higher WM capacity in good readers generates a larger N400 response compared to poorer readers with lower WM performance. Our results reveal that the relationship between WM and the N400 holds only for individuals with better SL-ON abilities. A larger WM capacity may allow for the manipulation of more sentential context information at any given time, which in turn enables the comprehender to make improved updates to their internal predictive models, which is reflected in a larger N400 response compared to less skilled statistical learners. Conversely, an increased WM capacity may not benefit individuals with lower SL-ON abilities by setting a "predictive ceiling": despite having sentential context information available in WM, they cannot utilize it in a way that would further improve their internal predictive models.

In this study, we observed no significant effects of auditory reaction time or visual processing speed on N400, nor were there any interactions with either of the SL metrics. During a conversation, minute details of communicative information have to be tracked continuously and simultaneously in visual and auditory modalities. Recent results have shown that both visual and auditory processing speed facilitate sentence comprehension when participants are presented with predictable sentences, but not in the case of unpredictable ones (Hintz et al. 2024). These findings highlight the primary role of automatic stimulus processing (and its speed) in cases when upcoming elements are predictable in sentences, but not in cases when predictive processing has a less important role. Our results, however, suggest that processing speed differences do not have an additional explanatory effect on the probability-based language comprehension processes we measured on the individual level by the evoked N400 responses.

5 | Limitations and Future Directions

The present study has limitations which invite further research. First of all, we failed to find any effect of offline statistical learning indices on individuals' N400 ERP responses. Besides the earlier discussed methodological concerns—namely, that online metrics are deemed more reliable and detailed indices than their offline counterparts (Batterink et al. 2015; Lammertink et al. 2019; Lukics and Lukács 2021)—our results could be indicative of a deeper divergence of offline and online SL metrics, in what they actually measure. If one considers that sentence comprehension involves active, statistical learning based predictive processes, which can be tapped into by online SL measures, but not by offline ones, it begs the question: is it because there are qualitative differences on a cognitive level measured by these separate metrics?

One possible explanation for the differential effects of the online and offline SL measures lies in the level of explicitness each task entails. The online task assesses implicit processing during exposure, whereas the offline task requires explicit judgment. Previous studies have shown that implicit learning can give rise to measurable behavioral or neural effects even when indices of explicit recognition fail to reveal such effects (Reber 1989; Nissen and Bullemer 1987; Isbilen et al. 2020; Batterink et al. 2015). The observed pattern may therefore reflect differences in how implicit and explicit learning mechanisms relate to neural responses such as the N400. It is also possible that the active SL-based processes that we observed affecting the N400 response via online metrics cannot be tapped into by offline metrics because they do not have a large enough effect outside of sentence comprehension, thus acting in a more localized manner than we initially expected. So the lack of SL-OFF effect on the N400 found raises not only methodological but conceptual questions about the SL-based predictive processes involved in comprehension.

Another possible limitation of this study is its use of CP as the only measure of N400 mechanisms and predictability. While CP served effectively to define categorical conditions and generate stimuli that elicited graded N400 responses, recent findings by Szewczyk and Federmeier (2022) suggest that distributional models such as GPT-2 may offer a useful complementary metric, especially for capturing variability among low-predictability sentence endings. Incorporating such models in future research could help clarify how stimulus-based predictability shapes the engagement of predictive mechanisms. At the same time, the inherent noisiness and interpretive ambiguity of CP should be considered when selecting and comparing predictability measures.

Building on this, the increasing availability of large language models (GPT-3, RoBERTa, and ALBERT; Michaelov et al. 2023) presents an opportunity to use potentially less noisy and more fine-grained metrics of sentence predictability. Recent behavioral and electrophysiological results (e.g., Goldstein et al. 2022, 2024) draw attention to parallels between predictive processing in the human brain and in deep language models (DLMs), demonstrating that both systems generate probabilistic expectations about upcoming words prior to their onset when processing sentences. Goldstein and colleagues also found compelling electrocorticographic evidence showing that these pre-onset predictions are evaluated against the incoming word to calculate the post-onset surprise response, which peaks around 400 ms after stimulus presentation (Goldstein et al. 2022). These findings, together with the results reported here, suggest that pre-onset predictive processes involved in language comprehension rely on the individual's SL abilities. The extent to which these predictive processes influence the N400 response and whether SL-based predictive processes contribute to post-onset integration are all important questions worthy of pursuit in future research.

While the present study focused on the N400 and its relationship to statistical learning (SL), previous research has identified other ERP components sensitive to statistical regularities. A mid-latency positivity resembling the P300 has been reported in several studies (Daltrozzo et al. 2017; Eghbalzad et al. 2021; Jost et al. 2015; Singh et al. 2017, 2018), interpreted as reflecting an attention-dependent predictive mechanism. These findings suggest that SL may engage multiple neural processes beyond those indexed by the N400. Future work could explore how these components interact to provide a more comprehensive account of prediction and adaptation in language processing.

6 | Conclusion

Our findings demonstrate that individual differences in statistical learning abilities affect N400 ERP magnitudes, providing support for a predictive generator component of N400 and suggesting the active engagement of statistical learning processes during language comprehension. Additionally, they indicate that in the case of skilled statistical learners, who can utilize contextual information effectively, working memory performance may contribute to this effect by setting the limit to the amount of contextual information available for such predictive processes. Observations of independent effects of statistical learning and CP on the N400 may index different aspects of a hierarchically organized predictive processing system. SL performance likely captures sensitivity to local transitional probabilities, and CP may reflect broader contextual constraints. Taken together, the results are consistent with the view that the N400 reflects the combined influence of predictive processes operating at multiple levels of representation.

Author Contributions

Márton Munding: conceptualization, data curation, formal analysis, methodology, visualization, writing – original draft, writing – review and editing. **Bálint Forgács:** conceptualization, data curation, formal analysis, funding acquisition, methodology, visualization, writing – review and editing. **Krisztina Sára Lukics:** data curation, formal analysis, investigation, methodology, visualization, writing – original draft, writing – review and editing. **Ágnes Lukács:** conceptualization, funding acquisition, methodology, project administration, writing – original draft, writing – review and editing.

Acknowledgments

This work was supported by the Lendület Research Grant of the Hungarian Academy of Sciences (Momentum 96233 'Profiling learning mechanisms and learners: individual differences from impairments to excellence in statistical learning and in language acquisition', PI: Ágnes Lukács) and by Grant TKP2021-NVA-02, and by a DFG grant of the German Research Council (CI247/7-1, Bálint Forgács). We are also grateful to Noémi Lehóczki, Brúnó Schütz, and Zsolt Mayer for their help in sentence stimulus material creation. We are especially grateful to Karolin Török-Suli for her help in data collection, and to Kornél Németh for his help in data collection and visualization. We are also grateful to Márta Zimmer for her valuable comments on the manuscript.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are openly available on OSF at http://doi.org/10.17605/OSF.IO/F2R4N.

References

Arciuli, J., and I. Simpson. 2012. "Statistical Learning Is Related to Reading Ability in Children and Adults." *Cognitive Science* 36: 286–304. https://doi.org/10.1111/j.15516709.2011.01200.x.

Aslin, R. N. 2017. "Statistical Learning: A Powerful Mechanism That Operates by Mere Exposure." *Wiley Interdisciplinary Reviews: Cognitive Science* 8, no. 1–2: e1373.

Baayen, R. H. 2008. *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge University Press.

Baayen, R. H., D. J. Davidson, and D. M. Bates. 2008. "Mixed-Effects Modeling With Crossed Random Effects for Subjects and Items." *Journal of Memory and Language* 59, no. 4: 390–412.

Barr, D. J., R. Levy, C. Scheepers, and H. J. Tily. 2013. "Random Effects Structure for Confirmatory Hypothesis Testing: Keep It Maximal." *Journal of Memory and Language* 68, no. 3: 255–278.

Bates, D. 2016. lme4: Linear Mixed-Effects Models Using Eigen and S4. R Package Version, 1, 1.

Bates, D., R. Kliegl, S. Vasishth, and H. Baayen. 2015. "Parsimonious Mixed Models." arXiv preprint arXiv:1506.04967.

Batterink, L. J., P. J. Reber, H. J. Neville, and K. A. Paller. 2015. "Implicit and Explicit Contributions to Statistical Learning." *Journal of Memory and Language* 83: 62–78.

Berkum, J. J. V., P. Hagoort, and C. M. Brown. 1999. "Semantic Integration in Sentences and Discourse: Evidence From the N400." *Journal of Cognitive Neuroscience* 11, no. 6: 657–671.

Boeve, S., H. Zhou, and L. Bogaerts. 2023. "A Meta-Analysis of 97 Studies Reveals That Statistical Learning and Language Ability Are Only Weakly Correlated." *L'Année Psychologique* 124, no. 3: 283–316.

Brouwer, H., H. Fitz, and J. Hoeks. 2012. "Getting Real About Semantic Illusions: Rethinking the Functional Role of the P600 in Language Comprehension." *Brain Research* 1446: 127–143.

Bulf, H., S. P. Johnson, and E. Valenza. 2011. "Visual Statistical Learning in the Newborn Infant." *Cognition* 121, no. 1: 113–127.

Chatrian, G. E., E. Lettich, and P. L. Nelson. 1985. "Ten Percent Electrode System for Topographic Studies of Spontaneous and Evoked EEG Activities." *American Journal of EEG Technology* 25, no. 2: 83–92 2.

Christiansen, M. H., C. M. Conway, and L. Onnis. 2012. "Similar Neural Correlates for Language and Sequential Learning: Evidence From Event-Related Brain Potentials." *Language & Cognitive Processes* 27: 231–256.

Clark, A. 2013. "Whatever Next? Predictive Brains, Situated Agents, and the Future of Cognitive Science." *Behavioral and Brain Sciences* 36, no. 3: 181–204.

Contreras Kallens, P., R. D. Kristensen-McLachlan, and M. H. Christiansen. 2023. "Large Language Models Demonstrate the Potential of Statistical Learning in Language." *Cognitive Science* 47, no. 3: e13256.

Conway, C. M. 2020. "How Does the Brain Learn Environmental Structure? Ten Core Principles for Understanding the Neurocognitive Mechanisms of Statistical Learning." *Neuroscience & Biobehavioral Reviews* 112: 279–299.

Conway, C. M., A. Bauernschmidt, S. S. Huang, and D. B. Pisoni. 2010. "Implicit Statistical Learning in Language Processing: Word Predictability Is the Key." *Cognition* 114: 356–371.

Conway, C. M., and M. H. Christiansen. 2006. "Statistical Learning Within and Between Modalities: Pitting Abstract Against Stimulus-Specific Representations." *Psychological Science* 17, no. 10: 905–912.

Conway, C. M., and M. H. Christiansen. 2009. "Seeing and Hearing in Space and Time: Effects of Modality and Presentation Rate on Implicit Statistical Learning." *European Journal of Cognitive Psychology* 21, no. 4: 561–580.

Conway, C. M., J. Karpicke, and D. B. Pisoni. 2007. "Contribution of Implicit Sequence Learning to Spoken Language Processing: Some Preliminary Findings With Hearing Adults." *Journal of Deaf Studies and Deaf Education* 12: 317–334.

Daltrozzo, J., S. N. Emerson, J. Deocampo, et al. 2017. "Visual Statistical Learning Is Related to Natural Language Ability in Adults: An ERP Study." *Brain and Language* 166: 40–51. https://doi.org/10.1016/j.bandl.2016.12.005.

Deacon, D., S. Hewitt, C. M. Yang, and M. Nagata. 2000. "Event-Related Potential Indices of Semantic Priming Using Masked and Unmasked Words: Evidence That the N400 Does Not Reflect a Post-Lexical Process." *Cognitive Brain Research* 9, no. 2: 137–146.

DeLong, K. A., L. Quante, and M. Kutas. 2014. "Predictability, Plausibility, and Two Late ERP Positivities During Written Sentence Comprehension." *Neuropsychologia* 61: 150–162.

DeLong, K. A., T. P. Urbach, and M. Kutas. 2005. "Probabilistic Word Pre-Activation During Language Comprehension Inferred From Electrical Brain Activity." *Nature Neuroscience* 8, no. 8: 1117–1121.

Delorme, A., and S. Makeig. 2004. "EEGLAB: An Open Source Toolbox for Analysis of Single-Trial EEG Dynamics Including Independent Component Analysis." *Journal of Neuroscience Methods* 134, no. 1: 9–21.

Eghbalzad, L., J. A. Deocampo, and C. M. Conway. 2021. "How Statistical Learning Interacts With the Socioeconomic Environment to Shape Children's Language Development." *PLoS One* 16, no. 1: e0244954. https://doi.org/10.1371/journal.pone.0244954.

Federmeier, K. D. 2007. "Thinking Ahead: The Role and Roots of Prediction in Language Comprehension." *Psychophysiology* 44, no. 4: 491–505.

Federmeier, K. D., and M. Kutas. 1999. "A Rose by Any Other Name: Long-Term Memory Structure and Sentence Processing." *Journal of Memory and Language* 41, no. 4: 469–495.

Ferreira, F., A. Foucart, and P. E. Engelhardt. 2013. "Language Processing in the Visual World: Effects of Preview, Visual Complexity, and Prediction." *Journal of Memory and Language* 69: 165–182.

Ferretti, T. R., M. Kutas, and K. McRae. 2007. "Verb Aspect and the Activation of Event Knowledge." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 33, no. 1: 182–196.

Fiser, J., and R. N. Aslin. 2001. "Unsupervised Statistical Learning of Higher-Order Spatial Structures From Visual Scenes." *Psychological Science* 12, no. 6: 499–504.

Goldstein, A., A. Grinstein-Dabush, M. Schain, et al. 2024. "Alignment of Brain Embeddings and Artificial Contextual Embeddings in Natural Language Points to Common Geometric Patterns." *Nature Communications* 15, no. 1: 2768.

Goldstein, A., Z. Zada, E. Buchnik, et al. 2022. "Shared Computational Principles for Language Processing in Humans and Deep Language Models." *Nature Neuroscience* 25, no. 3: 369–380.

Gomez, R. L., and L. Gerken. 1999. "Artificial Grammar Learning by 1-Year-Olds Leads to Specific and Abstract Knowledge." *Cognition* 70, no. 2: 109–135.

Hagoort, P., G. Baggio, and R. M. Willems. 2009. "Semantic Unification." In *The Cognitive Neurosciences*, edited by et al, 4th ed., 819–836. MIT Press.

Hagoort, P., L. Hald, M. Bastiaansen, and K. M. Petersson. 2004. "Integration of Word Meaning and World Knowledge in Language Comprehension." *Science* 304, no. 5669: 438–441.

Hald, L. A., E. G. Steenbeek-Planting, and P. Hagoort. 2007. "The Interaction of Discourse Context and World Knowledge in Online Sentence Comprehension. Evidence From the N400." *Brain Research* 1146: 210–218.

Hintz, F., A. S. Meyer, and F. Huettig. 2020. "Visual Context Constrains Language-Mediated Anticipatory Eye Movements." *Quarterly Journal of Experimental Psychology* 73: 458–467.

Hintz, F., C. C. Voeten, D. Dobó, K. S. Lukics, and Á. Lukács. 2024. "The Role of General Cognitive Skills in Integrating Visual and Linguistic Information During Sentence Comprehension: Individual Differences Across the Lifespan." *Scientific Reports* 14, no. 1: 17797.

Hodapp, A., and M. Rabovsky. 2021. "The N400 ERP Component Reflects an Error-Based Implicit Learning Signal During Language Comprehension." *European Journal of Neuroscience* 54, no. 9: 7125–7140.

Honke, G., K. J. Kurtz, and S. Laszlo. 2020. "Similarity Judgments Predict N400 Amplitude Differences Between Taxonomic Category Members and Thematic Associates." *Neuropsychologia* 141: 107388.

Isbilen, E. S., S. M. McCauley, E. Kidd, and M. H. Christiansen. 2017. "Testing Statistical Learning Implicitly: A Novel Chunk-Based Measure of Statistical Learning." In *The 39th Annual Conference of the Cognitive Science Society (CogSci 2017)*, 564–569. Cognitive Science Society.

Isbilen, E. S., S. M. McCauley, E. Kidd, and M. H. Christiansen. 2020. "Statistically Induced Chunking Recall: A Memory-Based Approach to Statistical Learning." *Cognitive Science* 44, no. 7: e12848.

Janacsek, K., J. Fiser, and D. Nemeth. 2012. "The Best Time to Acquire New Skills: Age-Related Differences in Implicit Sequence Learning Across the Human Lifespan." *Developmental Science* 15, no. 4: 496–505. https://doi.org/10.1111/j.1467-7687.2012.01150.x.

Jost, E., C. M. Conway, J. D. Purdy, A. M. Walk, and M. A. Hendricks. 2015. "Exploring the Neurodevelopment of Visual Statistical Learning Using Event-Related Brain Potentials." *Brain Research* 1597: 95–107. https://doi.org/10.1016/j.brainres.2014.10.017.

Karuza, E. A., T. A. Farmer, A. B. Fine, F. X. Smith, and T. F. Jaeger. 2014. "On-Line Measures of Prediction in a Self-Paced Statistical Learning Task." In *Proceedings of the Annual Meeting of the Cognitive Science Society*, 36. UC Merced.

Kidd, E. 2012. "Implicit Statistical Learning Is Directly Associated With the Acquisition of Syntax." *Developmental Psychology* 48, no. 1: 171–184. https://doi.org/10.1037/a0025405.

Kidd, E., and J. Arciuli. 2016. "Individual Differences in Statistical Learning Predict Children's Comprehension of Syntax." *Child Development* 87, no. 1: 184–193. https://doi.org/10.1111/cdev.12461.

Knowlton, B. J., and L. R. Squire. 1996. "Artificial Grammar Learning Depends on Implicit Acquisition of Both Abstract and Exemplar-Specific Information." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22, no. 1: 169–181.

Krieger, B., H. Brouwer, C. Aurnhammer, et al. 2024. "On the Limits of LLM Surprisal as Functional Explanation of ERPs." In *Proceedings of the Annual Meeting of the Cognitive Science Society*, Vol. 46. Open Access Publications from the University of California.

Kuhl, P. K. 2000. "A New View of Language Acquisition." *Proceedings of the National Academy of Sciences* 97, no. 22: 11850–11857. https://doi.org/10.1073/pnas.97.22.11850.

Kuperberg, G. R., and T. F. Jaeger. 2016. "What Do We Mean by Prediction in Language Comprehension?" *Language, Cognition and Neuroscience* 31, no. 1: 32–59.

Kutas, M., and K. D. Federmeier. 2011. "Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP)." *Annual Review of Psychology* 62, no. 1: 621–647. https://doi.org/10.1146/annurev.psych.093008.131123.

Kutas, M., and S. A. Hillyard. 1980. "Reading Senseless Sentences: Brain Potentials Reflect Semantic Incongruity." *Science* 207, no. 4427: 203–205. https://doi.org/10.1126/science.7350657.

Kutas, M., and S. A. Hillyard. 1984. "Brain Potentials During Reading Reflect Word Expectancy and Semantic Association." *Nature* 307, no. 5947: 161–163.

Kutas, M., R. Kluender, C. Barkley, and B. Amsel. 2016. "Language." In *Handbook of Psychophysiology*, edited by J. T. Cacioppo, L. G. Tassinary, and G. G. Berntson, 511–525. Cambridge University Press.

Lai, M. K., J. Rommers, and K. D. Federmeier. 2021. "The Fate of the Unexpected: Consequences of Misprediction Assessed Using ERP Repetition Effects." *Brain Research* 1757: 147290.

Lammertink, I., M. Van Witteloostuijn, P. Boersma, F. Wijnen, and J. Rispens. 2019. "Auditory Statistical Learning in Children: Novel

Insights From an Online Measure." *Applied PsychoLinguistics* 40: 279–302. https://doi.org/10.1017/S0142716418000577.

Laszlo, S., and K. D. Federmeier. 2011. "The N400 as a Snapshot of Interactive Processing: Evidence From Regression Analyses of Orthographic Neighbor and Lexical Associate Effects." *Psychophysiology* 48, no. 2: 176–186.

Lau, E. F., P. J. Holcomb, and G. R. Kuperberg. 2013. "Dissociating N400 Effects of Prediction From Association in Single-Word Contexts." *Journal of Cognitive Neuroscience* 25, no. 3: 484–502.

Lau, E. F., C. Phillips, and D. Poeppel. 2008. "A Cortical Network for Semantics:(de) Constructing the N400." *Nature Reviews Neuroscience* 9, no. 12: 920–933.

Lopez-Calderon, J., and S. J. Luck. 2014. "ERPLAB: An Open-Source Toolbox for the Analysis of Event-Related Potentials." *Frontiers in Human Neuroscience* 8: 213.

Lukács, Á., and F. Kemény. 2015. "Development of Different Forms of Skill Learning Throughout the Lifespan." *Cognitive Science* 39, no. 2: 383–404.

Lukács, Á., K. S. Lukics, and D. Dobó. 2021. "Online Statistical Learning in Developmental Language Disorder." *Frontiers in Human Neuroscience* 15: 715818. https://doi.org/10.3389/fnhum.2021.715818.

Lukics, K. S., D. Dobó, B. J. Ugrin, and Á. Lukács. 2021. *The Effect of Statistical Learning and General Cognitive Skills on Language Processing: A Structural Equation Modeling Study [Poster Presentation].* LingCologne.

Lukics, K. S., and Á. Lukács. 2021. "Tracking Statistical Learning Online: Word Segmentation in a Target Detection Task." *Acta Psychologica* 215: 103271. https://doi.org/10.1016/j.actpsy.2021.103271.

Lukics, K. S., and Á. Lukács. 2022. "Modality, Presentation, Domain and Training Effects in Statistical Learning." *Scientific Reports* 12, no. 1: 20878.

Mainela-Arnold, E., and J. L. Evans. 2014. "Do Statistical Segmentation Abilities Predict Lexical-Phonological and Lexical-Semantic Abilities in Children With and Without SLI?" *Journal of Child Language* 41, no. 2: 327–351. https://doi.org/10.1017/S0305000912000736.

Marcus, G. F., S. Vijayan, S. Bandi Rao, and P. M. Vishton. 1999. "Rule Learning by Seven-Month-Old Infants." *Science* 283, no. 5398: 77–80.

Michaelov, J. A., S. Coulson, and B. K. Bergen. 2023. "So Cloze Yet So Far: N400 Amplitude Is Better Predicted by Distributional Information Than Human Predictability Judgements." *IEEE Transactions on Cognitive and Developmental Systems* 15, no. 3: 1033–1042. https://doi.org/10.1109/TCDS.2022.3176783.

Miller, G. A., and J. A. Selfridge. 1950. "Verbal Context and the Recall of Meaningful Material." *American Journal of Psychology* 63, no. 2: 176–185.

Misyak, J. B., and M. H. Christiansen. 2012. "Statistical Learning and Language: An Individual Differences Study." *Language Learning* 62, no. 1: 302–331.

Mognon, A., J. Jovicich, L. Bruzzone, and M. Buiatti. 2011. "ADJUST: An Automatic EEG Artifact Detector Based on the Joint Use of Spatial and Temporal Features." *Psychophysiology* 48, no. 2: 229–240.

Molinaro, N., M. Conrad, H. A. Barber, and M. Carreiras. 2010. "On the Functional Nature of the N400: Contrasting Effects Related to Visual Word Recognition and Contextual Semantic Integration." *Cognitive Neuroscience* 1, no. 1: 1–7.

Nieuwland, M. S., and J. J. Van Berkum. 2006. "When Peanuts Fall in Love: N400 Evidence for the Power of Discourse." *Journal of Cognitive Neuroscience* 18, no. 7: 1098–1111.

Nissen, M. J., and P. Bullemer. 1987. "Attentional Requirements of Learning: Evidence From Performance Measures." *Cognitive Psychology* 19, no. 1: 1–32.

Pickering, M. J., and C. Gambi. 2018. "Predicting While Comprehending Language: A Theory and Review." *Psychological Bulletin* 144, no. 10: 1002–1044.

Rabovsky, M., S. S. Hansen, and J. L. McClelland. 2018. "Modelling the N400 Brain Potential as Change in a Probabilistic Representation of Meaning." *Nature Human Behaviour* 2, no. 9: 693–705. https://doi.org/10.1038/s41562-018-0406-4.

Radford, A., J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever. 2019. "Language Models Are Unsupervised Multitask Learners." *OpenAI Blog* 1, no. 8: 9.

Raimondo, F., J. E. Kamienkowski, M. Sigman, and D. F. Slezak. 2012. "CUDAICA: GPU Optimization of Infomax-ICA EEG Analysis." *Computational Intelligence and Neuroscience* 2012: 2.

Reber, A. S. 1989. "Implicit Learning and Tacit Knowledge." *Journal of Experimental Psychology: General* 118, no. 3: 219–235.

Saffran, J. R., R. N. Aslin, and E. L. Newport. 1996. "Statistical Learning by 8month-Old Infants." *Science* 274, no. 5294: 1926–1928.

Saffran, J. R., E. K. Johnson, R. N. Aslin, and E. L. Newport. 1999. "Statistical Learning of Tone Sequences by Human Infants and Adults." *Cognition* 70, no. 1: 27–52.

Siegelman, N., L. Bogaerts, O. Kronenfeld, and R. Frost. 2018. "Redefining "Learning" in Statistical Learning: What Does an Online Measure Reveal About the Assimilation of Visual Regularities?" *Cognitive Science* 42: 692–727.

Siegelman, N., and R. Frost. 2015. "Statistical Learning as an Individual Ability: Theoretical Perspectives and Empirical Evidence." *Journal of Memory and Language* 81: 105–120. https://doi.org/10.1016/j.jml.2015.02.001.

Singh, S., J. Daltrozzo, and C. M. Conway. 2017. "Effect of Pattern Awareness on the Behavioral and Neurophysiological Correlates of Visual Statistical Learning." *Neuroscience of Consciousness* 2017, no. 1: nix020. https://doi.org/10.1093/nc/nix020.

Singh, S., A. M. Walk, and C. M. Conway. 2018. "Atypical Predictive Processing During Visual Statistical Learning in Children With Developmental Dyslexia: An Event-Related Potential Study." *Annals of Dyslexia* 68: 165–179. https://doi.org/10.1007/s11881-018-0161-2.

Singmann, H., and D. Kellen. 2017. "An Introduction to Mixed Models for Experimental Psychology." In *New Methods in Neuroscience and Cognitive Psychology*. Psychology Press.

Spencer, M., M. P. Kaschak, J. L. Jones, and C. J. Lonigan. 2015. "Statistical Learning Is Related to Early Literacy-Related Skills." *Reading and Writing* 28, no. 4: 467–490. https://doi.org/10.1007/s11145-014-9533-0.

Szewczyk, J. M., and K. D. Federmeier. 2022. "Context-Based Facilitation of Semantic Access Follows Both Logarithmic and Linear Functions of Stimulus Probability." *Journal of Memory and Language* 123: 104311.

Taylor, W. L. 1953. "'Cloze Procedure': A New Tool for Measuring Readability." *Journalism Quarterly* 30, no. 4: 415–433.

Team R. C. 2017. "R: A Language and Environment for Statistical Com-Puting." R Foundation for Statistical Computing. https://www.R-project.org/.

Teinonen, T., V. Fellman, R. Näätänen, P. Alku, and M. Huotilainen. 2009. "Statistical Language Learning in Neonates Revealed by Event-Related Brain Potentials." *BMC Neuroscience* 10, no. 1: 1–8.

Torkildsen, J., J. Arciuli, and O. B. Wie. 2019. "Individual Differences in Statistical Learning Predict Children's Reading Ability in a Semi-Transparent Orthography." *Learning and Individual Differences* 69: 60–68. https://doi.org/10.1016/j.lindif.2018.11.003.

Troyer, M., and M. Kutas. 2020. "Harry Potter and the Chamber of What?: The Impact of What Individuals Know on Word Processing

During Reading." *Language, Cognition and Neuroscience* 35, no. 5: 641–657.

Troyer, M., T. P. Urbach, and M. Kutas. 2020. "Lumos!: Electrophysiological Tracking of (Wizarding) World Knowledge Use During Reading." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 46, no. 3: 476.

Turk-Browne, N. B., J. A. Jungé, and B. J. Scholl. 2005. "The Automaticity of Visual Statistical Learning." *Journal of Experimental Psychology: General* 134, no. 4: 552–564.

Urbach, T. P., K. A. DeLong, W. H. Chan, and M. Kutas. 2020. "An Exploratory Data Analysis of Word Form Prediction During Word-By-Word Reading." *Proceedings of the National Academy of Sciences* 117, no. 34: 20483–20494.

Urbach, T. P., and M. Kutas. 2010. "Quantifiers More or Less Quantify On-Line: ERP Evidence for Partial Incremental Interpretation." *Journal of Memory and Language* 63, no. 2: 158–179.

Van Den Brink, D., J. J. A. Van Berkum, M. C. M. Bastiaansen, et al. 2012. "Empathy Matters: ERP Evidence for Inter-Individual Differences in Social Language Processing." *Social Cognitive and Affective Neuroscience* 7, no. 2: 173–183. https://doi.org/10.1093/scan/nsq094.

Van Petten, C. 1993. "A Comparison of Lexical and Sentence-Level Context Effects in Event Related Potentials." *Language & Cognitive Processes* 8, no. 4: 485–531.

Van Petten, C. 2014. "Examining the N400 Semantic Context Effect Item-By-Item: Relationship to Corpus-Based Measures of Word Co-Occurrence." *International Journal of Psychophysiology* 94, no. 3: 407–419.

Van Petten, C., J. Weckerly, H. K. McIsaac, and M. Kutas. 1997. "Working Memory Capacity Dissociates Lexical and Sentential Context Effects." *Psychological Science* 8, no. 3: 238–242. https://doi.org/10.1111/j.1467-9280.1997.tb00418.x.

Yu, C., and L. B. Smith. 2007. "Rapid Word Learning Under Uncertainty via Cross-Situational Statistics." *Psychological Science* 18, no. 5: 414–420.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Data S1:** psyp70125-sup-0001-DataS1.docx.