Journal of Memory and Language 140 (2025) 104585

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Electrophysiological correlates of incidental L2 word learning from dialogue

Kristin Lemhöfer *, Anqi Lei, Anne Mickan

Donders Institute for Brain, Cognition and Behaviour, Radboud University, Thomas van Aquinostraat 4, 6525 GD Nijmegen, The Netherlands

ARTICLE INFO

Keywords:
Incidental L2 word learning
ERPs
Long-term memory
Late Positive Component
Subsequent memory effect

ABSTRACT

We aimed to determine the electrophysiological correlates of incidental L2 word learning during dialogue, bridging memory and second language acquisition research in a realistic, but strictly controlled experimental paradigm. Native Dutch speakers of L2 English learned English words previously unknown to them (as confirmed in a 'hidden' pretest) through auditory input in a dialogue-like setting revolving around price comparisons, while we measured their EEG. Hearing an unknown as compared to a known word elicited an early and sustained negativity, as well as a later LPC that was actually predictive of subsequent learning success. Notably, in a second block, we found that ERPs to novel words that had just been learned in the previous block were already undistinguishable from those for known words, while not yet learned novel words still showed similar ERP signatures as in block 1. This lends support for a fast learning mechanism in adults incidentally 'picking up' new L2 words.

Introduction

Knowing a second language next to one's mother tongue, often English as a lingua franca for those who do not already speak it as native speakers, has become a common requirement of modern life. While learning a second language (L2) often starts out in school or language courses, it is a lifelong process. Our knowledge of words, the most basic meaningful building blocks of language, increases until late adulthood in the first language (L1; Christian & Paterson, 1936; Keuleers et al., 2015) and most likely (though not yet documented in a normative way) also in the second language, if used regularly (Miralpeix, 2020). A large part of this vocabulary growth is typically a result of *incidental* learning, i.e., learning spontaneously without the explicit instruction to learn, like picking up words in conversations or during reading. Despite a large body of research into L2 word acquisition (e.g., de Vos et al., 2018; Schmitt & Schmitt, 2020; Webb et al., 2020), the neural mechanisms of incidentally 'picking up' a new L2 word (or failing to do so) have hardly been investigated, particularly in an approximately realistic setting.

The present study sets out to pinpoint the decisive moment of hearing a previously unknown L2 word in a realistic setting, to identify its electrophysiological correlates in the brain, and to determine in how far these correlates are indicative of learning success. The realistic setting we use is a (partly scripted) dialogue game; our previous studies have confirmed that this paradigm, which revolves around price comparisons, is highly successful in concealing the learning aim of the study, thus giving rise to 'incidental' (as opposed to intentional) learning (de Vos, Schriefers, & Lemhöfer, 2019; de Vos, Schriefers, ten Bosch, & Lemhöfer, 2019). We focus on three correlates: First, event-related potential (ERP) effects of encountering a previously unknown L2 word for a familiar object; second, ERP correlates of this encounter depending on whether the word will subsequently be learned or not; and third, correlates of a second encounter with such a novel L2 word that has just been learned or not (yet) learned, each in comparison to L2 words known already before participating in the experiment.

We hope that the results of the current study will inform us about in how far the neural processes during the incidental encounter and possible subsequent learning of a novel L2 word resemble those processes that we know from, on the one hand, linguistic (non)word processing, and on the other hand, intentional or incidental memory encoding processes. So far, research on second language acquisition (SLA) has typically been conducted within applied linguistics and in complete separation of (neuro-) cognitive memory research; therefore, we know very little about in how far general principles of memory processing apply to the learning of novel (L2) words (but see, e.g., Chen & Chen, 2023; Havas et al., 2018; Lindsay & Gaskell, 2010, for exceptions). If word learning is governed by largely the same neural processes

* Corresponding author at: Thomas van Aquinostraat 4, 6525 GD Nijmegen, The Netherlands.
E-mail address: kristin.lemhofer@donders.ru.nl (K. Lemhöfer).

https://doi.org/10.1016/j.jml.2024.104585
Received 6 March 2024; Received in revised form 3 October 2024; Accepted 8 November 2024
Available online 14 November 2024
0749-596X/© 2024 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

as learning in general, ERP signatures should also resemble each other.

A further issue is whether spontaneous, thus incidental learning follows the same 'memory route' as explicit or intentional learning does. While most memory studies use explicit learning paradigms, involving a strategy- and attention-governed process that resembles, for instance, classroom learning, incidental learning is supposed to be the product of spontaneously and effortlessly 'picking up' new information in a similar way as children acquire language (Hulstijn, 2003). It has also been shown that under incidental learning conditions, words can be acquired by memory-impaired patients with damage to the central neural hub for long-term declarative memory, the hippocampus (so-called 'fast mapping'; Sharon et al., 2011; Swingley, 2010). It seems thus plausible that incidental vs. explicit learning routines involve different neurophysiological processes. Therefore, we will use an incidental learning paradigm and compare it, where possible, with established findings from explicit learning paradigms.

For each of the three ERP correlates we are going to study, we will briefly review what we might expect to find, based on memory and psycholinguistic studies.

Neurophysiological correlates of encountering a novel (L2) word

We know of three ERP studies that examined the moment of encountering a new, as compared to an already known, word. All these studies replaced existing words in sequences of written sentences with the pseudowords; for instance, the pseudoword *meeve* replaced the word *cloud* (Batterink & Neville, 2011). The sentences were either constraining with respect to meaning, such that the meaning of the 'novel word' could be inferred from the context (M+ condition); or this was not the case, either because sentence context was not constraining, or because the meanings were inconsistent across several sentences containing the same 'novel word' (M- condition). Sentences in which the real words were preserved served as baseline (R condition). Remarkably, though, following the psycholinguistic rather than the memory research tradition, two of these studies looked only at the N400 (typically measured between 300 and 500 ms post-word onset), the ERP component commonly attributed to semantic integration, ignoring possible later memory-related effects like the late positive component (LPC, starting from about 500 ms after word onset).

In the first of these studies in L1 Spanish, Mestres-Missé et al. (2007) reported 'a posterior N400′ during reading of a real word as compared to a pseudoword (both M+ and M-), which however became indistinguishable for M + pseudowords vs. real words by the third sentence that contained a given target. As mentioned above, other time windows were not analyzed. In contrast to these findings, Chen et al. (2017), in L2 English, observed a more canonical N400 over the whole scalp when pseudowords were first encountered as compared to real words. In the only study that also considered later LPC-effects, Batterink and Neville (2011), presenting L1 English speakers with a total of ten English sentences per target (pseudo)word, also observed an N400 for pseudowords as compared to real words at first presentation; this difference diminished with subsequent presentations. The LPC effects seem somewhat blurred by the many presentations, but there was a greater LPC for M+ than for M- pseudowords that increased with more repetitions. Interestingly, the study also split conditions into pseudowords that were subsequently learned vs. not learned, a set of results we will return to later on (for comparison 2).

These three studies, although not very consistently, seem to indicate larger N400s when 'novel words' as compared to existing words were presented, and no (or not analyzed) later effects. However, in an experimental situation where the replacement of real words with pseudowords must have been fairly transparent to the participants, it is questionable in how far 'learning' these pseudowords resembles incidental, real-life learning of existing, but previously unknown words.

In the psycholinguistic word recognition literature, a larger N400 to pseudowords as compared to real words in, for instance, a lexical decision task, is a common finding (e.g., Barber et al., 2013; Bermúdez-Margaretto et al., 2015; Holcomb et al., 2002; Leinonen et al., 2009; Supp et al., 2004; Ziegler et al., 1997); however, these experiments do not typically have a learning component. In a word learning study using a semantic classification task during EEG measurement, Bakker et al. (2015) also found larger N400 components for novel vs. existing words, but the effect carried on into the LPC window (500–700 ms), i.e. there was a smaller (positive) LPC for novel words.

Looking at the literature of so-called recognition memory, where items are first studied and then presented among new items ('lures') for recognition, the comparison of the encounters with new vs. old (i.e. previously studied) items during the recognition test may be regarded as similar as the comparison between perceiving a novel ('new') vs. an existing word ('old'). In this literature, it is commonly observed that (correctly classified) new items give rise to more negative ERPs than old items, both (left) parietally from about 500 to 800 ms after stimulus onset, i.e., in the LPC window, and frontally between approximately 300 and 500 ms (sometimes called "FN400 old/new effect"), the classical N400 window (though the N400 is typically distributed across the whole scalp). The parietal effect between 500 and 800 ms and the FN400 have been attributed to qualitatively different recognition processes based on 'recollection' and 'familiarity', respectively (e.g., Curran & Cleary, 2003; Duarte et al., 2004; Rugg & Curran, 2007; Rugg et al., 1998). Possibly, the lingering "N400" effects for new vs. existing words observed by Bakker et al. (2015) and Bermúdez-Margaretto et al. (2015) were in fact neural correlates of recollection as already described in the memory literature. Regarding a possible difference between incidental and intentional (explicit) encoding, old/new effects have typically been observed to be highly similar across these modes of encoding, at least for subsequently successfully recognized 'old' trials (e.g., Boehm & Sommer, 2005; Haese & Czernochowski, 2015; van Hooff, 2005).

Returning to our study, it is not quite clear what to expect in terms of ERPs when L2 learners encounter a previously unknown (as opposed to a known) L2 word incidentally. Psycholinguistic research would most likely point to an N400 effect caused by the 'pseudoword' status of these novel words, while memory research, based on findings for 'new' items in the recognition paradigm, would suggest one or multiple negativities that may be more stretched out across time and the scalp.

Neurophysiological correlates of subsequent learning success

Our second focus is on whether ERP signatures during encountering a previously unknown L2 word are predictive of later learning success. Sanquist et al. (1980) were the first to observe that ERPs collected during encoding of to-be-remembered stimuli differ depending on whether they are later successfully remembered or not, so-called subsequent memory effects (SME's). Since then, many studies have corroborated and extended this observation (e.g., Friedman & Johnson, 2000; Mecklinger & Kamp, 2023; Paller & Wagner, 2002). The most frequent finding seems to be that of a positivity for later remembered vs. not remembered items starting about 500 ms post stimulus onset during encoding, i.e., in the LPC window (e.g.,Paller et al., 1988; Sanquist et al., 1980; Van Petten & Senkfor, 1996), but effects of earlier latency and different polarity have also been reported (Mecklinger & Kamp, 2023).

Regarding the difference between intentional and incidental encoding, Cycowicz and Friedman (1999) observed SME's only for intentionally, not for incidentally encoded sounds; however, other studies (e.g., Otten & Rugg, 2001; Paller et al., 1987) did observe robust SME's in different incidental encoding tasks.

Accordingly, in the word 'learning' study during reading by Batterink and Neville (2011) described above, collapsing across all ten presentations of a target pseudoword, ERPs during reading those pseudowords that were later correctly recalled (i.e., by coming up with the corresponding real word) were more positive in the 500–900 ms LPC window than the pseudowords not correctly recalled. The effect was present over the whole scalp, but largest at posterior sites. In an (explicit) word learning study in which L1 Chinese speakers learned Russian words through repeated exposure to translation pairs, Zhang et al. (2023) also showed that LPC amplitudes correlated with learning success.

To our knowledge, no previous study has investigated subsequent memory effects during 'truly' incidental L2 word learning. If this kind of learning proceeds using the same mechanisms as those at work during the highly artificial memory paradigms, we would expect a larger LPC for subsequently learned vs. not learned novel words. However, given that encoding conditions in our paradigm are incidental, effects may also be absent, as an earlier study for auditory learning has failed to find incidental encoding SME's (Cycowicz & Friedman, 1999).

ERP correlates of word processing after learning

Finally, we will look at the ERP signature of novel L2 English words that have been learned, or that have not been learned just one experimental block ago, each in comparison to existing words that were known before. We expect that ERPs for those novel words that have just been learned will resemble those for previously known words more than is the case for novel words that have not yet been learned.

We do not know of any studies that looked at ERP effects to novel words (or other items) that had just been 'trained' before, depending on the learning outcome of this training (learned vs. not learned words). However, various studies showed that ERPs for trained or learned novel (L1 or L2) words gradually get to resemble those for previously known words (Bakker et al., 2015; Shtyrov et al., 2010; Soskey et al., 2016). Frishkoff et al. (2010) compared two training methods, only one of which involved a high-constrain context that could actually lead to learning, and found that only novel words trained in this condition, but not those trained in a low-constraint, 'meaningless' training condition, became word-like in terms of their ERP signature two days later. Bermúdez-Margaretto et al. (2015) showed that ERPs for words and pseudowords that initially differed in terms of an N400 extending into the LPC window for pseudowords largely disappeared after six blocks of presentation. This convergence of ERP signals for (repeated) pseudowords and words occurred especially in the LPC window and for central-parietal electrodes. Similarly, Mestres-Missé et al. (2007) showed that ERPs on pseudowords repeatedly put in a meaningful contrast (M+) became indistinguishable from those for real words by the third presentation in terms of an N400 (300–500 ms), the only time window analyzed, while M- pseudowords not embedded in a meaningful context remained more positive than M + and real word conditions.

In contrast, the similar study by Batterink and Neville (2011) does not report of ERP signatures of pseudowords (M+ or M-) approaching those of real words with more (up to 10) repetitions.

Thus, despite the lack of studies that employed a similar comparison, we expect that novel English words that have just been learned one block earlier will show more word-like ERP signatures than those that have not yet been learned, despite the same degree of exposure (once before).

The price comparisons dialogue paradigm

We developed a method to study incidental L2 word learning in which participants are invited to take part in a study that revolves about the subjective value of objects and their comparison, in order to conceal the word learning aspect of the study (de Vos, Schriefers, ten Bosch, et al., 2019). When participants came to the lab, it was revealed to them that the study would be in their second language, in this case English, with the pretended reason that students of all nationalities were invited to the study (which was not true; the participant recruitment system offered the study only to native speakers of Dutch). In the paradigm, after a pretest that also is about subjective prices, the participant and a partner take turns in comparing objects in price ("a rose is cheaper than a book"). While the original method, as described in de Vos et al. (2019), involves a dialogue between a real person (the experimenter) and the participant, this was replaced by a 'dialogue' with pre-recorded phrases in the present study, due to the need for maximal standardization of the auditory signal in an EEG context.

After listening to a price comparison by the (real or simulated) partner, the participant has to press a button indicating whether they agree with the other's statement. Critically, some of the objects are typically unknown in English in our population of L2 speakers (e.g., *whisk, float*, or *bib*), which will be confirmed for each individual participant by the pretest. All these words occur first in an utterance by the partner, i.e., the pre-recordings, in this case. Hearing these words said by the 'other person' gives the participant the chance to learn the word and be able to use it in a later trial involving the same object. The method has been proven successful both in learning rates and in concealing the learning aim of the study (de Vos, Schriefers, ten Bosch, et al., 2019; for a variant of the paradigm, see de Vos, Schriefers, & Lemhöfer, 2019). We will slightly adapt and use the method in an EEG setting to study the three questions framed above.

Methods

All experimental procedures were performed in compliance with the law and institutional guidelines and were approved by the Ethics Committee of the Faculty of Social Sciences at Radboud University, ECSW2014/0109/245. All participants gave informed consent before participation.

Participants

Initially, 64 speakers with Dutch as only mother tongue and proficient in L2 English, mostly students at Radboud University (the Netherlands), were recruited without referring to English as the language of the experiment. This was done by offering the study to only those participants from the university's participant recruitment system SONA who were registered to have Dutch as native and English as a second language (which is the case for basically all Dutch students). We tested as many participants as we could afford given our time and money resources. All the participants were right-handed, had corrected-to-normal or normal vision and no hearing impairment.

Five participants who either knew too many novel (target) words (12 or more, = 30 %), or too few of the presumably known control and filler words (less than 84, = 70 %) in the pretest were sent home after the pretest, since they would have to learn considerably fewer, or more words (including control / filler words that we assumed to be known) than other participants, and because there was a large risk of some cell sizes being too small for reliable EEG comparisons.

For the final analyses, remaining participants with too small cell sizes in any of the relevant conditions (known words, novel words, novel words learned and not learned in the first block) had to be excluded. This means in particular that participants who learned too few or too many words, or who had too many EEG artefacts, could not be included. We excluded participants with less than seven trials in any of the cells of the comparisons we would make. This was the case for 34 participants, so that the remaining sample consisted of 25 participants for the first block (17 females, age 19–34, mean age = 23.4). For the last two analyses in the second block, distinguishing novel words that had vs. had not been learned in the prior first block, we had to exclude three additional participants due to the small number of usable trials, leaving us with 22 participants for this comparison.

Participants also completed a language background questionnaire and the short English vocabulary test LexTALE (Lemhöfer & Broersma, 2012). The results of these measurements for the 25 remaining participants are summarized in Table 1.

Table 1
Characteristics of participants (N = 25) concerning their language background.

	Mean (*SD*)	Range (Min-Max)
Age	23.4 (3.2)	19–34
Age of first English acquisition	10.7 (2.0)	7–15
LexTALE score (%)	76.6 (7.7)	63–93
Self-ratings (1–7)		
Frequency of using English in daily life	5.0 (1.0)	2–6
Proficiency of reading English	5.4 (0.8)	3–7
Proficiency of writing English	4.6 (0.9)	3–6
Proficiency of speaking English	4.6 (1.0)	2–6
Proficiency of listening to English	5.6 (0.9)	4–7
Number of other foreign languages spoken	0.5 (0.8)	0–3

Materials

For stimulus selection, we conducted a pilot study with 45 Dutch speakers of the same population who would not participate in the final study. In that pilot study, 280 pictures of buyable and depictable objects were shown and participants were asked to name them in English and in Dutch (the latter to make sure they identified the object correctly). The pictures were photos of the respective item, with sufficient resolution and if possible on a white background, taken from the internet and standardized to a uniform size. From this item pool, we selected 40 English words, all non-cognates with Dutch, as to-be-learned target words. The pictures for these words were well recognized, as reflected by Dutch naming responses, but their name in English was known by less than 20 % of the pilot study participants. Examples are *buckle*, *mortar* or *bib*. Furthermore, 120 English words, many of them cognates with Dutch, that were known by 80 % of participants or more were selected as fillers (80) or matched control words (40).

The 40 control words were matched to the target words item-wise in terms of length in syllables and, if possible, onset phoneme (examples: *bottle, mirror, bed* for the above named targets *buckle, mortar, bib*). Word length was between one and three syllables (mean = 1.7, *SD* = 0.6 for both targets and controls). None of the target or control words were plurals or compound words, given that learning a new combination of already known constituent words presumably represents a different process than learning entirely new words.

80 fillers, also known by at least 80 % of the pilot study participants, were used to fill up the trial scheme (see *Word Learning Part* below). They had comparable characteristics to the target and control words (mean length 1.7 syllables, *SD* = 0.8), but some were compound words (e.g. *microwave, gameboy*). The complete list of items can be found in Appendix A.

For the audio recordings of the price comparisons presented in the listening trials of the dialogue game, the objects were internally ordered in a roughly plausible manner according to price, and the sentences containing the pairwise price comparisons were made according to that order. That is, a *fridge* would have been regarded as more expensive than a *flower,* but maybe as cheaper than a *table.* This way, we hoped to elicit occasional disagreements in the acceptability judgments following the listening trials. The sentences were recorded as they were read out by a young female native speaker of standard southern British English. In each of the sentence recordings containing a target or matched control word, the time was measured from the start of the recording to the onset of the respective word using Praat (Version 6.0.23; https://www.praat.org/). These onset values were later imported into the EEG signal file to signal the onset of the relevant auditory input.

Procedure

When participants entered the lab, they were told they were to take part in a study on the stability of subjective price estimates, to be conducted in English because one of the experimenters did not speak Dutch (which was indeed true). There were four parts of the experiment: 1) the pretest, 2) the main learning part with EEG measurements, 3) the administration of the LexTALE vocabulary test and of the language background questionnaire (see Table 1), and 4) a posttest.

Pretest

In order to assess for each individual participant which target items were already known to them, or which control and filler items were not known, a pretest was conducted. All 160 pictures that were to appear in the word learning (EEG) experiment were printed on cards and sorted in a fixed, pseudorandomized order in which there were no more than three (presumably unknown) target words in a row. The participant was asked to estimate a price for each object and to utter this estimate in a complete sentence because the experimenter, seated behind a computer screen, could not see which object they were referring to, e.g. "An apple costs 50 cent." They were told that if they did not know the name of the object in English, they should describe it, like (for a *whisk*) "the thing to whip cream with costs 5 Euros".

The experimenter pretended to note down these prices, however what they actually noted was whether the object name was known or unknown to the participant. As 'known', we counted correct productions as well as slightly wrong variations for which a correct 'core' was clearly present (e.g., *floater* instead of *float*), or words with inaccurate pronunciation. Incorrect or missing responses such as descriptions (*the thing to whip cream with*), placeholders (*the thing*), or the Dutch word (*de garde*) were regarded as 'unknown'.

Based on this pretest, we included only those trials in the final analyses in which a) the target word was unknown to the respective participant, b) the control word that had been paired with it in the respective trial was known, and c) the target word's matched control word (appearing in a different trial) was also known. After the pretest, the EEG cap was fitted and all preparations were conducted for the subsequent EEG measurements.

Word learning part with EEG

The main word learning part took the form of a dialogue game with a pre-recorded 'partner' that revolved around price comparisons. The use of pre-recordings was necessary to standardize the auditory input for each participant, as well as to facilitate that the EEG signal could be linked to the exact onset of critical words in the heard utterances. Participants were instructed that, in every other trial ('speaking trial'), they were to make price comparisons between two objects for which they had previously given price estimates (i.e., in the pretest); their price comparisons should be in agreement with those previously given estimates. In the intervening trials ('listening trials'), they would hear the price comparisons of their 'partner' and should indicate whether they agreed with them. Listening trials were the ones from which we would analyze the EEG, looking at neural correlates of word learning. Speaking trials were used to identify whether words had been learned from the prior auditory exposure to the novel words or not. Participants were instructed not to blink and to only look at the fixation cross while they were listening to the audio recording.

During each trial, the participants first saw two pictures side by side on the screen, on a white background. After two seconds, the pictures disappeared and a fixation cross was shown in the middle of the screen, to avoid eye movements to the two objects during the critical EEG measurement. In listening trials, the audio recording of a sentence referring to the two pictures was played simultaneously with the fixation cross. Afterwards, the participants were asked to press one of two buttons on the buttonbox (left for "No" or right for "Yes") to indicate if they agreed with the statement made in the audio recording. The button press triggered the appearance of two new pictures on the screen, which would be a speaking trial. Participants were instructed to start their own price comparisons with the picture on the left and to choose the adjective (*cheaper / more expensive than*) accordingly. Their answer was recorded for later coding. After they finished their sentence, participants pressed a button to start the next (listening) trial.

This main study part consisted of 320 trials containing 640 items, divided into four blocks of 80 trials, separated by breaks. Additionally, six practice trials that contained filler words were given at the beginning. Each of the 40 target words appeared once in a listening trial and once in a speaking trial in the first half of this experimental part (i.e. blocks 1 or 2), and the same for the second half (blocks 3 or 4), i.e. four times in total. The 120 control and filler words also appeared four times throughout the experiment (twice in a listening, twice in a speaking trial). List construction was such that each target word appeared in a listening trial first, giving the participant the chance to learn; five trials later, it appeared in a speaking trial to test for the success of learning.

Target words were located once as first and once as second noun of a listening trial and the corresponding speaking trial (five trials later) in a counterbalanced fashion. They were always paired with a control or filler word, and never with another target word. Furthermore, if such a word accompanying a target word was unknown, the target word was excluded from analyses, because being exposed to two novel words in one trial might involve different processes. None of the pairwise item combinations of any trial was repeated in the experiment.

Post-experiment measurements and posttest

After the EEG session, we asked participants what they thought the goal of the study was. Out of the 25 included participants, seven mentioned that they thought it had to do with word learning; for one participant, a record of this response is missing; and the remaining 17 did not mention anything about word learning, but believed that the study was about price estimates. Afterwards, participants were debriefed about the real goal of the study.

Subsequently, the English vocabulary test LexTALE (Lemhöfer & Broersma, 2012) was administered, as well as the language background questionnaire that provided the data for Table 1. Afterwards, a final posttest was conducted to assess longer-term learning. Pictures for all target words were shown again in a fixed order, and had to be named in English.

EEG recording, preprocessing and analysis

The EEG signal was recorded with 27 active Ag-AgCI electrodes (actiCAP, Brain Products) embedded in an elastic cap and two additional electrodes placed on both mastoids (see Fig. 1 for their location on the scalp). The left mastoid electrode was used as online reference, but the EEG was later re-referenced off-line to the average of the left and right mastoid electrodes. Additionally, two electrodes were placed on the outer side of both eyes, one below and one above the right eye to be able to identify eye movements artefacts. The impedance for the electrodes was kept below 15 kΩ. The EEG signal was amplified and sampled at a frequency of 500 Hz. The signal was filtered online with a low cutoff of 0.01 Hz and a high cutoff of 125 Hz.

After recording, the onset times of target and matched control words that had been measured in the audio files were imported in the EEG signal file. Based on the participant's behavior, each of these words received a marker that indicated in which condition it fell, i.e., whether it had been known (including partially known) or unknown in the pretest, and whether it had been produced correctly after input in the first and the second half of the experiment (i.e., for target words, whether they had been learned).[1] As 'partially known', we regarded answers that

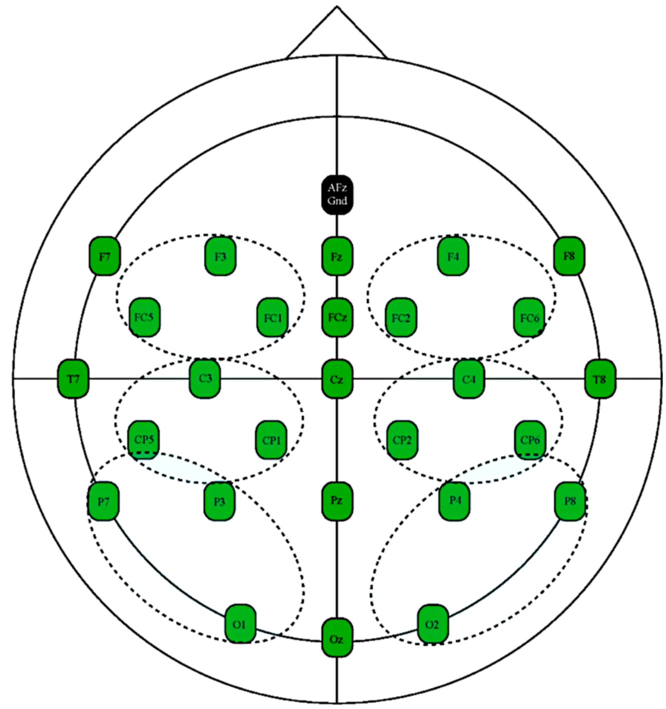

Fig. 1. Schematic representation of the position of the electrodes on the EEG cap.

clearly showed that a part of the word had been learned (e.g., *floater* instead of *float*).

EEG data were preprocessed using the FieldTrip toolbox (http://www.fieldtriptoolbox.org) in MATLAB (Oostenveld et al., 2011). Trials were segmented from −500 ms to 1500 ms relative to the onset of target/control words in each sentence. A low-pass filter of 40 Hz was applied to the EEG data (excluding EOG channels) to remove high frequency noise. A Butterworth band-pass filter (1–15 Hz, filter order of 4) was applied to the EOG data to allow for better detection of ocular artifacts. Independent component analysis (ICA) was applied to correct eye blinks and eye movements. Following ICA correction, artifact rejection of the remaining noisy trials and channels was conducted manually based on visual inspection. 10.3 % of all trials were rejected across participants. Interpolation was performed to replace rejected channels with the average of their neighboring channels. Baseline correction was applied using the − 200 ms to 0 ms window prior to the target/control word onset.

The following analyses were planned: 1) *known* vs. *novel words*: previously unknown (targets) vs. known words (control words) in the first block. We will refer to this factor as *Word Status* in the following; 2) *subsequent learning effect*: previously unknown words (targets) that were vs. were not subsequently learned in block 1 (a factor we will call *Learned Status*), each compared with their respective matched known controls. Thus, we are looking for the 2 x 2 interaction of the two factors Word Status und Learned Status; and 3) *recent learning effect*: unknown words that had or had not been learned in block 1, but now encountered in block 2, compared with their respective matched control words, i.e., the Word Status x Learned Status interaction for block 2. Average voltages were plotted using FieldTrip.

EEG data were analyzed using nonparametric cluster-based permutation tests (Maris & Oostenveld, 2007). As a data-driven analysis approach, cluster-based permutation tests require fewer a priori restrictions on the selection of time windows and electrodes for analysis, and they control for multiple comparisons without reducing the statistical power (Candia-Rivera & Valenza, 2022). This method is ideal for the current study as we had very few a priori expectations on the time

[1] It is possible that a word was passively known to a participant, but could not be actively produced in the pretest. We have no means to assess this in the present data; therefore, it should be kept in mind that what we regard as 'word learning' in the present study is the transition from being unable to being able to produce a word, given its picture or concept.

window and distribution of effects, given the lack of literature on ERP correlates of auditory word learning. Following the FieldTrip documentation (https://www.fieldtriptoolbox.org/tutorial/cluster_permutation_freq/), the cluster-based permutation method first determined spatiotemporally adjacent clusters that show a difference across experimental conditions via dependent-samples *t*-tests (at an alpha threshold of 0.05). This was followed by the permutation test, in which each cluster's test statistic (the sum of all *t*-values within each cluster) was compared to a null distribution of cluster statistics generated by 1000 Monte-Carlo permutations (i.e., randomization of data labels). P-values of the observed clusters were calculated as the proportion of randomizations that had a larger cluster statistic than the observed effect. The critical alpha level for permutation tests was set to $p < .05$.

Results

Behavioral results

As mentioned above, we had to exclude participants with too high or too low learning rates, or with too many EEG artifacts, because this would leave us with too few trials per cell to analyze. However, in Appendix B, we report on the behavioral results of the complete sample of 58 participants. This large group of participants showed a qualitative similar learning pattern to the selected group of 25 participants, but had a slightly lower percent correct rate in the pretest, slightly less steep learning, but less forgetting than the selected group.

Fig. 2 shows the percentage of correctly produced and partially correctly produced target words through the course of the experiment (pretest, the two blocks of the learning phase, and posttest). As can be seen from the figure, there was a steep increase of words produced fully correctly, and a slight increase of partially correct words until the end of the learning phase (block 2). Towards the posttest, there was a drop of the percentage of words produced correctly due to forgetting, but a slight increase in partially correct words.

Performance on fillers and control words was 95.60 % correct in the pretest in the selected sample (range 74–100 %, $SD = 5.08$).[2]

Because the target words in the responses were embedded at several places in sentences, response latencies were not measured.

ERP results

We carried out the comparisons described above. Mean and ranges of trial numbers per cell can be found in Appendix C.

Comparison 1: Known vs. Novel words, block 1

The grand-averaged ERP waveforms comparing previously unknown and known words in the first block can be found in Fig. 3 (A), along with a plot visualizing the scalp topography (the bottom row).

The dependent-samples *t*-test revealed a significant negative cluster for the effect of Word Status, i.e. for previously unknown relative to known words from 0 to 714 ms ($p < .01$) and two significant positive clusters at later intervals (766–1094 ms, $p < .01$; 1112–1260 ms, $p = .03$) with an 18 ms gap in between them. The effects were distributed over the whole scalp, with the effect in the second time window focusing on posterior sites (see scalp maps). Thus, we see a negativity that includes the N400 time window (~300–500 ms), but which unexpectedly starts already immediately after onset of the target word, followed by a

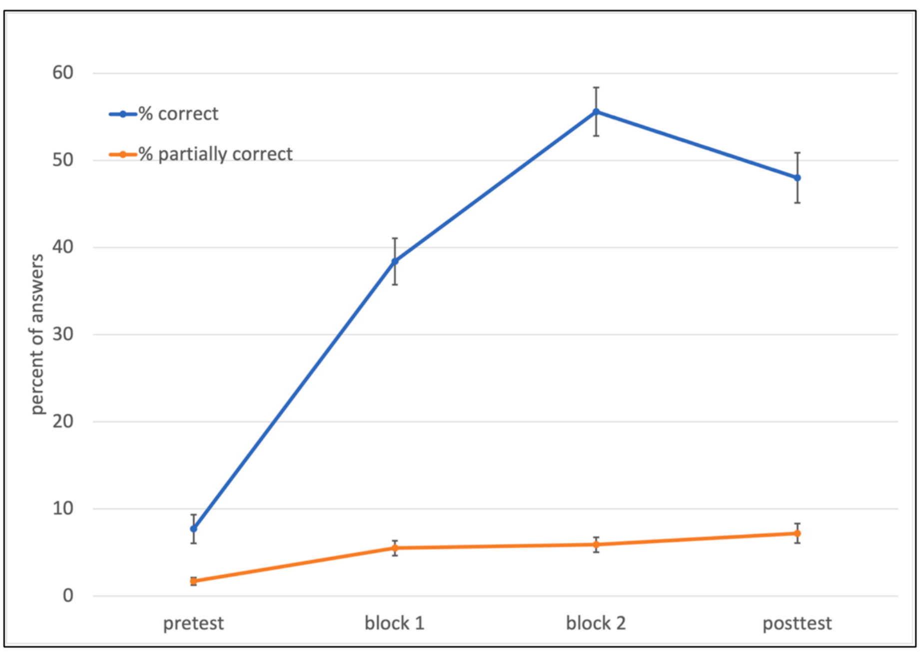

Fig. 2. Mean naming accuracy (blue or dark: % of words produced fully correctly, orange or light: % of words produced partially correctly) on the 40 target words throughout the experiment by the selected 25 participants. Error bars represent standard errors. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

[2] If these easy words were not produced correctly, this was often due to a misinterpretation of the picture (shown for the first time here).

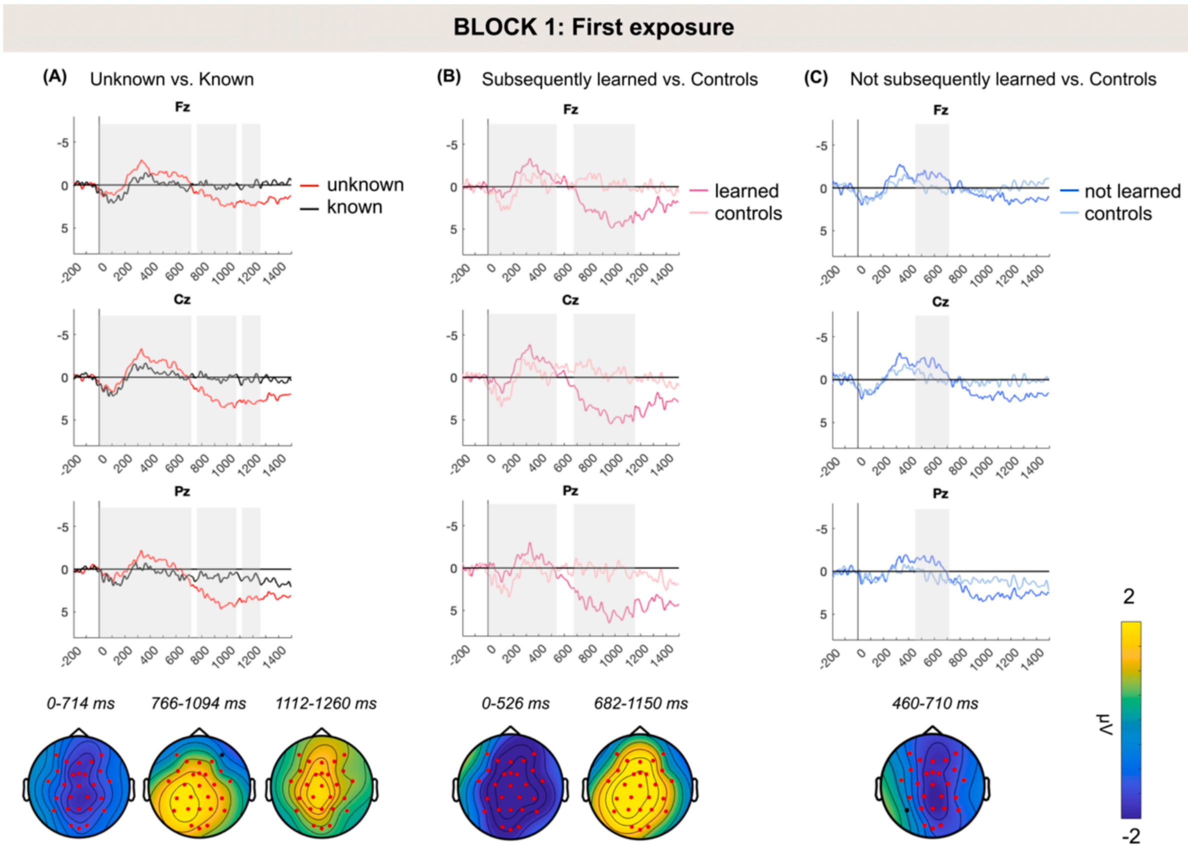

Fig. 3. Grand-averaged ERPs for the first block at selected frontal, central, and posterior midline electrodes (Fz, Cz, Pz), comparing between (A) previously unknown (red) and known (black) words, (B) spontaneously learned target words (dark pink) and matched control words (light pink), and (C) not learned target words (dark blue) and matched control words (light blue)(N = 25). Significant clusters revealed by the permutation tests were shaded in gray. Topographic plots were included for each significant cluster (bottom row). Electrodes that contribute to the significant clusters are shown in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

late positive component (LPC) at 766–1260 ms.[3]

Comparison 2: Subsequent learning effect, block 1

For comparison 2, we compared the effects depicted in panels (B) and (C) in Fig. 3: the effects of Word Status for targets (compared to their known matched controls) that later turned out to be learned in block 1, as evident from the respective production trial following the input (listening) trial in that same block, compared to the respective effect for those words that had not been learned. Thus, we looked for an interaction of Word Status and Learned Status.

The cluster-based permutation test revealed a significant interaction between the two factors over the whole scalp in three time windows (negative cluster: 54–254 ms, $p < .01$; positive cluster 1: 750–904 ms, $p = .01$; positive cluster 2: 944–1110 ms, $p = .01$). The first negative cluster is caused by the early negativity (but with a much smaller time window than in Comparison 1) having an earlier onset for learned targets (compared to their matched controls) than for not learned targets (also compared to their control words). The two positive clusters were caused by the LPC, which was also larger for learned than for not learned targets.

To explore this interaction, pairwise *t*-tests comparing target words and controls for each condition separately (learned vs. not learned) were conducted. We chose to do this for the complete time window, because the exact time of effects resulting from permutation tests cannot be taken completely literal (Sassenhagen & Draschkow, 2019). For learned target words, we found an early negative cluster relative to matched control words from 0 to 526 ms ($p < .01$), followed by a positive cluster at a later time window from 682 to 1150 ms ($p < .01$; see Fig. 3, B). The effects were distributed over the whole scalp. For target words not learned in block 1, we found a significant negativity from 460 to 710 ms ($p < .01$) across the whole scalp, but we did not observe any significant late positive cluster (see Fig. 3, C).

Thus, first of all, these data show that the surprisingly early negativity of Comparison 1 was mainly caused by subsequently learned words (relative to their controls) in the very early window, followed by a N400-like effect that however came quite late for not learned words (starting at about 460 ms). Second, and most importantly, the LPC was only significant for words that later turn out to be learned. The occurrence of an LPC can thus be seen as a marker for subsequent learning success.

Comparison 3: Recent learning effect, block 2

Finally, we compared the EEG signatures of target words presented in block 2 that had been learned in the previous block (as apparent in the production trial following the input trial in block 1) with those target words that had not been learned. Keep in mind that the number of participants that could be included in this comparison was reduced to 22. Again, we compared each target word group (learned vs. not learned) with their matched (known) control words as a baseline. Thus, we were looking for the interaction between Word Status and Learned Status, but then in the presentation following the possible learning

[3] This effect is briefly interrupted by an 18 ms period of lacking significance, probably due to the permutation method we have used; had we selected the LPC window beforehand as most studies do, this would have gone unnoticed.

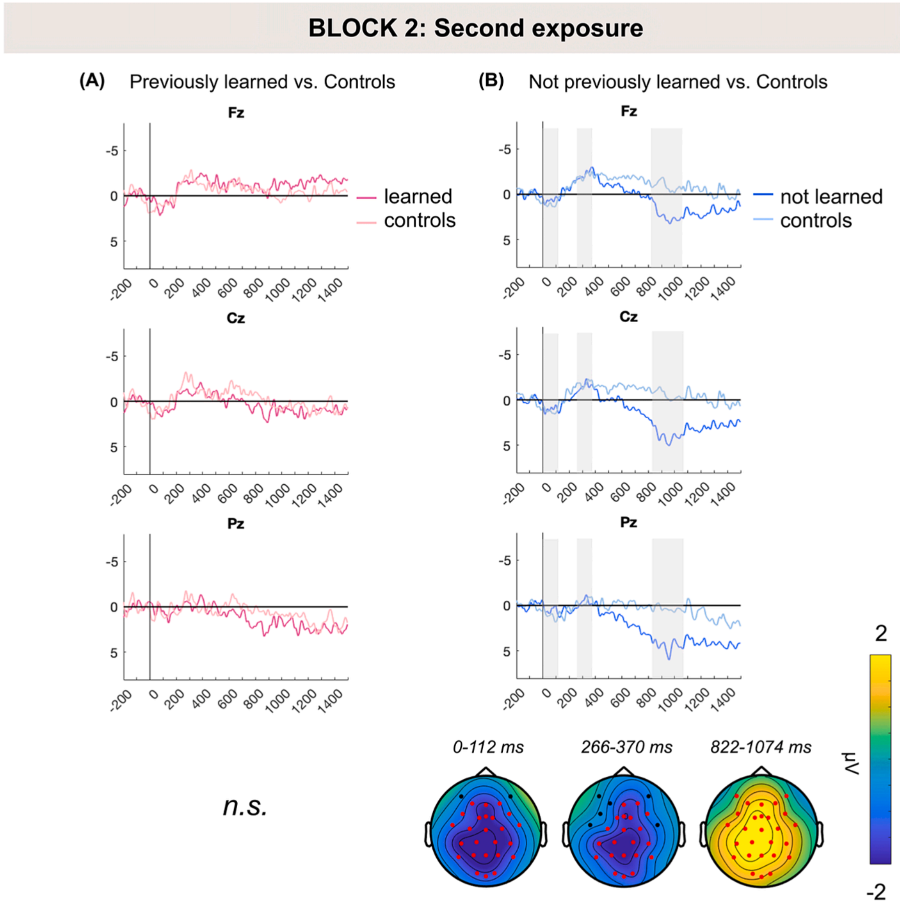

Fig. 4. Grand-averaged ERPs for the second block at selected frontal, central, and posterior midline electrodes (Fz, Cz, Pz), comparing between (A) previously learned target words (dark pink) and matched control words (light pink), and (B) not learned target words (dark blue) and matched control words (light blue)(N = 22). Significant clusters revealed by the permutation tests are shaded in gray. Topographic plots are included for each significant cluster. Electrodes that contribute to the significant clusters are shown in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(block 2). Fig. 4 shows the EEG activity in target vs. matched control words for those that had previously been learned (panel A) or not (panel B). Thus, the partition of trials into conditions is the same as in Fig. 3 B and C, but now sampled during the second presentation during a listening trial in block 2.

The cluster-based permutation test revealed two clusters with a significant interaction effect of Word Status and Learned Status: one early negative cluster (266–342 ms, $p = .04$) which, in principle, should signify either a larger negativity or a smaller positivity for learned words (which is hard to see in the graphs), and a late positive cluster (906–1060 ms, $p < .01$) signaling a larger LPC for not learned words.

The follow-up t-test that compared only learned target words and their matched controls did not reveal any significant cluster (see Fig. 4., A). However, when comparing not learned target words and their matched controls, we found two negative clusters at early time windows (0–112 ms, $p = .01$; 266–370 ms, $p = .04$) as well as one positive cluster at a later interval (822–1074 ms, $p < .01$) (see Fig. 4, B).[4]

In our view, both early negativities can be seen as instances of the longer early negativity already observed for comparison 1: previously unknown words (in this case, words encountered once, but not learned) give rise to early negativities compared to well-known control words.

The much larger LPC for not yet learned vs. previously learned words as visible in Fig. 4 was confirmed by the permutation tests, albeit only in a restricted time window (interaction: 906–1060 ms, not learned words only: 822–1074 ms). This is very likely due to power issues, given that on the basis of their learned / not learned distribution, only 22 participants could be included in this comparison, and cell sizes were rather low in some cells (see Appendix C). Fig. 4, however, shows a pattern that is clear and meaningful: target words already learned in the previously block already 'behaved' like long-known words in terms of their EEG signature, while target words that had been encountered equally often, but not successfully learned, showed basically the same ERP signature as completely novel words (see Fig. 3, A).

Discussion

In the present study, we used a paradigm based on scripted dialogue to study the electrophysiological correlates of incidental L2 word learning. The underlying aim was to simulate the spontaneous 'picking up' of new words in a conversation in a second language. The study differed in many respects from prior studies. For instance, this paradigm resembles natural dialogue as closely as possible in a controlled EEG-laboratory setting. The majority (71 %) of participants did not detect

[4] The negativities in the two early time windows are hard to see in the three central electrodes in Fig. 4, in particular the second negativity, which was the least significant of all found effects (p =.04). To understand this result better, Appendix D shows an electrode (O2) where this effect is more visible, with a slight negative inflection for not learned words compared to their matched controls.

the learning aspect of the study.[5] Importantly, encoding (or learning) and retrieval (or testing) of the new words were embedded in the same dialogue game. For the sake of the probably most face-valid definition of truly 'knowing' a word (i.e., being able to produce it), retrieval was assessed by way of production or 'cued recall', rather than recognition as in the majority of memory studies. Unlike most previous word learning studies using EEG, the newly learned words in our study were real English words, to be learned by advanced L2 English speakers, who were 'covertly' individually pre-tested on their (missing) knowledge of these target words. Word presentation was only auditory, as is natural for a dialogue, and not visual as in most memory or word recognition studies.

Overall, we were interested in (1) ERP correlates of encountering a spoken novel word compared to an already known word in general, (2) whether these correlates were predictive of subsequent remembering, and (3) how learned vs. not learned novel words 'behaved' during a second exposure (relative to already known words). In what follows, we will discuss how all these results are related to findings from both word processing and memory literature.

For all three comparisons, we would like to acknowledge that our final sample size, after having recruited more than 60 participants, is on the low side (N = 25 and N = 22). The reasons for this are specified above, and are due to the fact that for our performance-based condition splits, only 'intermediate' learners with a good balance of learned and not learned words could be used. No prior word learning study, to our knowledge, has performed the same split, maybe due to this extreme restriction, but leading to a gap in our knowledge about learning-dependent ERP signatures of L2 word acquisition that we aimed to fill; future studies may want to turn to a word recognition rather than production study to avoid the enormous loss of participants (see, e.g., Bakker et al., 2015, who used a word recognition / matching paradigm). It should however be noted that the ERP field has only recently started to use sample sizes above 25; for a comparison, the study by Bakker et al. (2015) had 22 participants, and that by Batterink & Neville (2011) had 20. Among all 14 studies cited in the present paper that used ERPs, only two have a sample size clearly above 25 (Chen & Chen, 2023: 35; Zhang et al., 2023: 50).

Incidentally encountering new (previously unknown) L2 words compared to already known words

In the first comparison, we aimed at identifying the ERP correlates of hearing, for the first time, an L2 word that one did not know before. The previous psycholinguistic and memory literature, especially in the context of L1 word processing and learning, had mainly pointed at negativities, either earlier in an N400 window or later in the LPC window, for novel words (compared to already known words) or 'new' stimuli (compared to known words or previously studied stimuli).

We observed a sustained and broadly distributed negativity for unknown relative to known words in block 1 that spanned the usual N400 window (300–500 ms after onset of the critical word), but occurred much earlier and lasted longer (from 0 to ~ 700 ms), followed by a positivity from about 750 to 1250 ms. First, while the negativity for novel words was expected, its time course is uncanonical. It resembles "mismatch negativity" effects from auditory speech segmentation as observed by De Diego Balaguer et al. (2007), where (rule-violating) nonwords elicited a similarly early negativity compared to words until about 400 ms, which was also followed by a larger LPC from about 600 ms onwards. However, in that study, this pattern was interpreted as a correlate of phonotactic rule extraction, which we did not manipulate.

In (L1) word recognition studies, Ziegler et al. (1997) observed an "N400" for pseudowords vs. words in a semantic categorization task as early as 75 ms, while Supp et al. (2004), in an auditory lexical decision task, reported an "N400" as late as 500–700 ms. Conversely, Baart and Samuel (2015) observed lexicality effects during auditory word processing that started 200 ms after onset of the syllable that determined lexicality, and that (even though later windows were not analyzed) seemed to last until about 700 ms. Thus, in the word recognition literature, the typically observed N400 effect for unknown lexical items (pseudowords) as compared to words has more often been seen to begin and / or end outside the classical N400 time window.

On the other hand, the recognition memory literature identifies two negativities typically elicited by new vs. old items, a frontal negativity from 300-500 ms (also called 'FN400′) and a left parietal negativity from about 500 to 800 ms (Curran & Cleary, 2003; Duarte et al., 2004; Rugg & Curran, 2007). It should however be considered that these effects are obtained during retrieval and not, as in our study, during encoding. Furthermore, considering that in our study, we had auditory presentations of complete sentences tied to a complex and incidental-learning price comparison task, it is conceivable that our observed effects are not as focused as in visual single-item learning or lexical decision studies (with contextually isolated word stimuli). For instance, the early onset of the negativity might be partly due to the fact that participants already saw the pictures of an upcoming listening trial two seconds before the onset of the auditory phrase. That is, participants could already anticipate on the occurrence of an unknown and to-be-learned word, possibly accelerating the onset of the detection of such a target item. In general, however, the negativity itself fits both with the idea of increased effort for lexical-semantic processing that is usually attributed to the N400 in language studies (Kutas & Federmeier, 2010; Lau et al., 2008; Rabovsky et al., 2018), as well as with failed retrieval from long-term memory when comparing novel to already known words. The broad distribution of our observed effect over the whole scalp, however, fits more with a 'linguistic' N400 than with the negativities reported in the recognition memory literature (frontal or left-parietal).

Late in the analysis window (~750 ms), our observed negativity for novel words turned into a positivity, which was rather unexpected on the grounds of earlier psycholinguistic or recognition memory studies. As mentioned above, recognition memory studies typically observe only a negativity for 'new' vs. 'old' items in the retrieval phase, which is however different from the current situation where the EEG measures were collected during encoding. In word learning and processing studies, Holcomb et al. (2002) observed an N400 for pseudowords that extended into a later time window (550–800 ms). Similarly, Bakker et al. (2015), Borovsky et al. (2010), in a semantically high-constraint condition) and, at least descriptively, Mestres-Missé et al. (2007; the later time window was not analyzed) observed more negative ERPs in an 'LPC window' (500–700 ms) for novel vs. existing words. Thus, all these studies would lead to the expectation of a negativity, not a positivity, for novel words at later time windows. However, none of these studies looked at a time window after 700 or 800 ms. Batterink and Neville (2011), in their study of pseudowords replacing real words in sentences, did not find a difference between real and pseudowords in an LPC window from 500 to 900 ms at first presentation of an item.

Yet, our observed LPC effect seems to be compatible with some studies on word repetition in discourse. Van Petten et al. (1991) observed that words that had not previously been mentioned in a text gave rise to an increased N400 and a larger LPC compared to repeated words. A similar result, albeit restricted to certain conditions (items congruent with sentence context, or short repetition lag, non-living items, and young adults as participants), was obtained by Besson et al. (1992) and by Zhou et al. (2015). Analogously, it is conceivable that the already known words in our study might be considered as 'repeated' words (as they are probably used in participants' daily life), while the unknown words can be likened to 'new' words in these studies.

[5] This percentage was much higher (98%) in the original version of the experiment (de Vos, Schriefers, ten Bosch, et al., 2019), where the dialogue partner was a real person. A dialogue with a real confederate was however not feasible, and not sufficiently standardized in terms of auditory input, in an EEG context.

Finally, to anticipate on the discussion of the following comparison, the 'subsequent memory' effect, it is clear that the LPC can indicate the submission of new information to declarative memory, with its size correlating with later recall success (Fernández et al., 1999). In that context, it is not surprising that our novel words elicited this ERP component, as we will discuss in the following section.

Successfully vs. Unsuccessfully encoding novel words

In the second comparison, we split the trials of encountering novel words in block 1, as described above, into those that later turned out to have led to successful learning (because the respective target word was produced correctly in a later production trial) and those that did not. Both categories of novel words were compared to their respective matched known words ('matched controls'). We observed that the larger LPC for novel relative to known words was restricted to those novel words that were subsequently learned, which showed a large effect from about 700 to 1150 ms post word onset. In contrast, there was no LPC effect for not learned novel words. We also found that the early onset of the earlier negativity (0 ms, lasting until ~ 550 ms) was also restricted to the learned targets, while the not learned targets showed, as only effect, a later negativity from ~ 450 to 700 ms.

While the fact that both groups of novel words showed a negativity is compatible with the idea of the negativity signaling failed lexical access (i.e., retrieval from word memory) for these previously unknown words, we do not have an explanation for the difference in the latency of the effect. One bold speculation is that an early anticipation of an upcoming to-be-learned novel word on the basis of the previously presented picture pair, as mentioned above, may have been more conducive to subsequent learning than no such anticipation.

More importantly in terms of the assumed memory mechanisms in the current incidental L2 word learning paradigm, we replicated the 'subsequent memory' effect on the LPC that has previously been reported in recognition memory studies (Friedman & Johnson, 2000; Paller et al., 1987; Paller et al., 1988; Schott et al., 2002; Van Petten & Senkfor, 1996), in a recent explicit L2 word learning study (Zhang et al., 2023), as well as in one L1 word learning study in which pseudowords replaced real words in sentences (Batterink & Neville, 2011). Because of the dependence of the LPC on learning outcomes, this ERP component can be linked to successful declarative memory formation, which is in accordance with its neural sources like the hippocampus (Fernández et al., 1999; Friedman & Johnson, 2000). In other words, the size of the LPC, obtained in a naturalistic situation where novel words appeared 'covertly' in a partner's utterance, predicted the likelihood of these novel words being picked up by the L2 speaker. That is, neither the fact that this learning was incidental, nor that it was about L2 words presented in a much more 'noisy' task than in typical list learning experiments, nor that we used cued recall rather than recognition as a test, prevented the occurrence of such subsequent memory effects.

While it seems unsurprising that declarative memory processes are involved in the present task, incidental L2 word learning in a near-realistic situation like dialogue has never been investigated from this angle. In fact, most L2 word learning studies to date have investigated only the consequences, not the process, of learning (e.g., whether it gives rise to semantic priming effects), and few of them adopted a memory perspective. This may also be partly due to the fact that assessing the subsequent memory effect requires a balanced distribution of novel word trials that lead and do not lead to learning, which is only the case for some participants and can only be determined after testing, resulting in a large number of excluded participants (as in our case).

Immediate consequences of L2 word learning

In our third comparison, we turned to block 2 and compared novel and already known L2 words depending on whether the former had just, in block 1, been learned or not. In the absence of any comparable prior studies, our results were even more clearcut than we had expected: Novel words that had been successfully learned (and produced) in the previous block became statistically indistinguishable from known words ('matched controls') in terms of ERPs. In contrast, novel words that had been presented, but not yet learned, displayed a similar pattern as novel words presented for the first time, namely an early (but now not as sustained) negativity, and a positivity in a late LPC window.

This result is in accordance with findings showing a convergence of ERP signatures for known words vs. novel words or pseudowords as the latter are repeated or trained (Bakker et al., 2015; Frishkoff et al., 2010; Mestres-Missé et al., 2007). However, these studies all involved explicit rather than incidental learning, and a much more intensive training, or a higher number of repetitions, than a single exposure as in the present study. More importantly, none of these studies included a condition in which presented, but not learned novel words were also compared to already known words. In other words, none of the earlier studies took learning success for the novel words into account. It is this inclusion of not learned novel words – that still showed differences to already known words in a second exposure – that lets us rule out the possibility that the missing effects for successfully learned novel words vs. known words are due to a lack of power (given that the number of participants qualifying for this comparison was reduced to 22).

One surprising aspect of this finding is that the ERP convergence of just learned and long-known words happened in block 2, i.e. after only one exposure in block 1. It has to be considered, though, that for each novel word that qualified as a 'target' for a participant, the respective participant had shown a failed naming attempt in the pretest (otherwise, the target would have been excluded). We have previously shown that such failed naming attempts produce the noticing of 'vocabulary holes' in L2 learners, which promote subsequent learning (de Vos, Schriefers, & Lemhöfer, 2019). Furthermore, each word that contributed to this condition had been successfully pronounced by the learner in the previous block. Possibly, our fast learning results would not have been quite as fast if it had not been for this pretest and the successful naming attempt.

In the psycholinguistic word learning literature, it is typically assumed that full lexicalization, i.e. the integration of new words into (semantic) verbal memory, requires over-night consolidation (see Palma & Titone, 2021, for a review). Most evidence for this consolidation claim comes from behavioral studies, which use, for instance, semantic priming or form-related inhibition as indices for lexicalization, showing that these connections between old and newly learned words only surface fully after sleep (Clay et al., 2007; Davis et al., 2009; Dumay & Gaskell, 2007; Tamminen & Gaskell, 2013). In contrast to this account, some researchers have argued, especially for a certain form of incidental word learning called 'fast mapping procedure', for the existence of a fast learning mechanism that integrates novel words into semantic word memory without a lengthy consolidation period or even without hippocampal involvement (Atir-Sharon et al., 2015; Coutanche & Thompson-Schill, 2014; Merhav et al., 2015; Sharon et al., 2011; but see also Cooper et al., 2019; Gaskell & Lindsay, 2019).

It is currently unclear what either the consolidation or the Fast Mapping accounts would predict in terms of ERP signatures, but our result of fast convergence of just learned novel and already known words in terms of ERPs seems compatible with fast lexical learning; for instance, Bakker et al. (2015) and Kaczer et al. (2018) show that ERP signatures reflect the status of consolidation in lexical memory. These studies, however, include a second session after one or two days, which we did not, nor did we collect established behavioral indices of lexicalization like form competition or semantic priming. The two studies also identify the N400 more than the LPC as a reflection of lexicalization, a finding that we do not really replicate (our negativity for not yet learned words is very small, descriptively). All in all, our data do not directly speak to the issue of consolidation and lexicalization.

What does the LPC during exposure to novel words reflect?

Given the significance of the LCP as a marker of episodic memory retrieval (e.g., Rugg & Curran, 2007; Peters & Daum, 2009), one could be led to assume that our LPC was reversely related to the strength of the episodic memory representation for the previous encounters with a word. If a novel word is forgotten, the reason for this is probably the too weak memory trace it has left in episodic memory, compared to a novel remembered word, or a word known before (which also leaves a trace for the episode during which it was presented). This would explain the effects we see in block 2 (comparison 3), but it falls short of explaining the subsequent memory effect in block 1 (comparison 2): because this effect is predictive of subsequent learning, it appears to be an effect of memory *encoding* rather than retrieval.

There are however two more possibilities we see at this moment for the interpretation of the LPC in our data. First, it is possible that the LPC on not yet learned novel words in block 2 is in fact a subsequent memory effect for a potential next block, just as it was in block 1. The behavioral data show that after block 1, about 60 % of target words remained to be learned, of which participants, on average, learned about 30 % (taking the remaining 60 % as the new total). Analyzing these two groups of novel words separately in terms of the EEG to calculate a new subsequent memory effect was impossible due to too small cell sizes. This would mean that the LPC for all not yet learned novel words in block 2 (Fig. 4 B) would have been driven by those words learned in this block.

The third and last, but possibly related option is that all not-yet learned new words received a special kind of 'attention for learning' (and, possibly, more so than in block 1), since learners were probably aware that they had not yet succeeded in producing them; already learned novel words did not need to receive this attention any more. This would be compatible with an account of LPC as "a potential marker of high-quality, long-lasting learning attention" (Turk et al., 2018, p. 1327). This interpretation is not only in line with memory findings, especially regarding the subsequent memory LPC effect, but also with results from non-memory studies showing that the LPC is larger during the exposure to stimuli that are more task-relevant (Eugster et al., 2014; Kissler et al., 2009; Yang et al., 2019).

The two latter accounts are reconcilable with each other when one assumes that the kind of attention that the LPC signals promotes learning; the more attention is allocated to a learning episode, e.g. because of a recently experienced failed naming attempt, the more likely it is that it will be learned. This is also in line with a visual inspection of the block 1 data (Fig. 3) in which not subsequently learned novel words (panel C) seem to display a small LPC descriptively, which however does not reach significance. Still, these novel words might have received a slightly larger portion of attention than already known words, but not enough to enable learning up to being able to produce the word.

A note on the incidental nature of learning in the present paradigm

In the current study, we regard word learning as 'incidental' because, first, there was no explicit instruction to learn, and second, by using the price comparisons as a cover task, we concealed the learning goal of the study. In the memory literature, learning is typically regarded as 'incidental' when participants, prior to the learning phase, are not made aware of the (specific) learning goal, e.g. by the announcement of a subsequent memory test (e.g., Eagle & Leiter, 1964; Paller et al., 1987); our use of the term is consistent with this. However, we do acknowledge that the nature of the dialogue paradigm, which alternated between listening (learning) and speaking (testing) trials, and which was moreover preceded by a covert pretest, likely caused a certain intrinsic intention to learn in the participants. We could only have avoided this by turning the task into a purely perceptive listening task with a subsequent, surprising memory test. This would however have fallen short of simulating the dialogue situation we were interested in and would most likely have led to much less learning (see de Vos et al., 2019, for learning results in such a variant of the task). Thus, whether it can be called 'truly incidental' or not, our aim was to study the processes during the 'picking-up' of a word in a dialogue-like situation, be it with or without an intrinsic intention to learn. Finally, the fact that only a minority our participants (29 %) thought that the study was about word learning showed that most participants' major intention was on producing and judging the price comparisons, and not on learning words. While the present, already small sample could not be further divided, it would however be very interesting for future research to address the question whether these participants showed different results from the other group of 29 % who suspected a certain word learning goal in the study (see de Vos, Schriefers, & Lemhöfer (2019) for such an analysis by learning awareness).

Conclusion

The current study examined the electrophysiological correlates of incidental L2 word learning in dialogue. In particular, we adopted a perspective that bridges the classical memory and classical word processing literature to understand what happens in the brain during the incidental encounter and, possibly, learning of an L2 word that has previously been unknown to the speaker. Our results show, first, that hearing novel words elicits a negativity that is broadly compatible with both the N400 reported in the psycholinguistic literature for pseudowords and novel words, and for 'new' stimuli in the recognition memory literature. In line with the memory literature, but unexpected in terms of the word recognition literature, there also was an LPC effect for novel words that was predictive of subsequent learning. Finally, the ERP results of the second block show that ERP signatures for newly learned words became indistinguishable from those for 'old' words already after one exposure, while old-new differences, especially the LPC, were retained for new, not yet learned words. One possible explanation for this finding is the interpretation of the LPC as a marker of 'attention for learning'. These results advance our understanding of the neural underpinnings of the process of L2 speakers 'picking up' new words spontaneously in daily-life situations.

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Kristin Lemhöfer: Writing – review & editing, Writing – original draft, Visualization, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Anqi Lei:** Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Formal analysis, Data curation. **Anne Mickan:** Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Parts of these data have been collected in the context of a Master thesis by Julia Egger. We also thank Iris Verpaalen for help with data collection and Jochem Koopmans for help with data management.

This work was supported by the Dutch Research Council (NWO) (*vidi* grant to K. Lemhöfer, no. 276-89-004).

Appendix A

Table A1
Novel (target) words and their matched control words.

Target word	Matched control word	Target word	Matched control word
anvil	pencil	raft	tent
apron	apple	rake	sock
bib	bed	rattle	racket
bobbin	balloon	saucer	helmet
bridle	book	scooter	necklace
buckle	bottle	scythe	soap
chisel	candle	skimmer	muffin
colander	camera	squeegee	barbie
float	fork	swatter	wallet
funnel	flower	syringe	shower
hinge	ring	tassel	toilet
ladle	table	thimble	hammer
level	lighter	top	cup
maraca	banana	tripod	laptop
mortar	mirror	turnip	towel
onesie	tissue	valve	key
peeler	guitar	vise	vase
peg	bag	whisk	clock
pennant	painting	wrench	watch
plunger	pillow	yarn	car

Filler words

airplane, arrow, backpack, ball, battery, bell, belt, bicycle, bikini, blouse, boat, boomerang, boot, bra, bread, brush, bucket, bus, button, cd, chain, chair, couch, cross, crown, curtain, doll, door, dreamcatcher, dress, earring, egg, fridge, frisbee, gameboy, glass, glove, goal, helicopter, jacket, keyboard, knife, lamp, lipstick, magnet, microphone, microwave, mouse, newspaper, orange, pan, paperclip, pen, perfume, piano, plant, plate, pumpkin, radio, rope, sausage, scarf, shoe, skateboard, skirt, smartphone, spoon, stamp, straw, suit, suitcase, swing, t-shirt, tape, telephone, television, tie, toothbrush, train, umbrella.

Appendix B

Table B1
Mean naming accuracy on the 40 target words throughout the experiment of the complete set of 58 participants.

	Pretest		Learning phase, block 1		Learning phase, block 2		Posttest	
	Mean (*SD*)	Range	Mean (*SD*)	Range	Mean (*SD*)	Range	Mean (*SD*)	Range
% correct	5.52 (7.45)	0–27.50	31.62 (20.24)	5.00–77.50	43.73 (24.56)	7.50–95.00	39.44 (23.64)	5.00–97.50
% partially correct	1.29 (1.94)	0–7.50	3.93 (3.87)	0–15.00	4.87 (4.15)	0–17.50	5.34 (4.53)	0–25.00

Table B2
Mean naming accuracy on the 40 target words throughout the experiment of the remaining N = 25 participants.

	Pretest		Learning phase, block 1		Learning phase, block 2		Posttest	
	Mean (*SD*)	Range	Mean (*SD*)	Range	Mean (*SD*)	Range	Mean (*SD*)	Range
% correct	7.70 (8.19)	0–27.50	38.40 (13.25)	12.50–70.00	55.60 (13.87)	32.50–85.00	48.00 (14.40)	17.50–85.50
% partially correct	1.70 (2.13)	0–7.50	5.50 (4.27)	0–15.00	5.90 (4.26)	0–12.50	7.20 (5.56)	0–25.00

Appendix C

Table C1
Mean and range of numbers of trials per condition for the ERP analyses.

Condition	Mean	Range (Min-Max)
Block 1		
Comparison 1: Known vs novel words		
Novel words (targets)	33	21–40
Known words (matched controls)	31	11–39
Comparison 2: Subsequent learning effect		
Subsequently learned targets	13	7–24
Matched controls for subsequently learned targets	13	7–23
Not subsequently learned targets	20	8–29
Matched controls for not learned targets	20	10–29
Block 2		
Comparison 3: Recent learning effect		
Previously learned targets	13	7–23
Matched controls for previously learned targets	12	7–23
Previously not learned targets	21	11–30
Matched controls for previously not learned targets	19	11–30

Appendix D

Illustration of the negativity in the time window 266–370 ms in comparison 3 (block 2) for not yet learned novel words vs. controls.

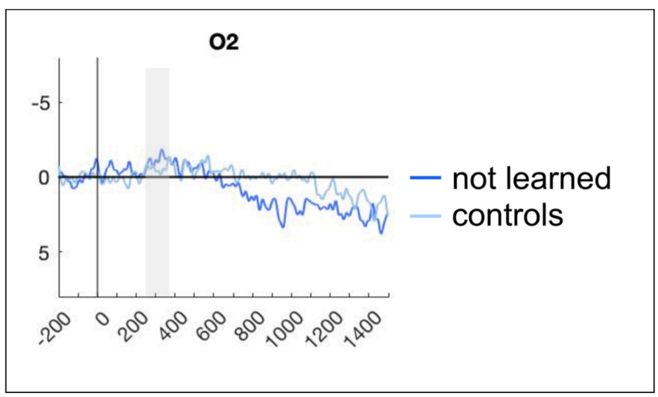

Fig. D1. Grand-averaged ERPs for the first block at electrode O2 comparing between not learned target words (dark blue) and matched control words (light blue) in block 2 (N = 22). The area shaded in gray shows the effect at the time window.

Data availability

The raw data and analysis scripts are available at the Radboud Data Repository (https://doi.org/10.34973/24y8-q773).

References

Atir-Sharon, T., Gilboa, A., Hazan, H., Koilis, E., & Manevitz, L. M. (2015). Decoding the formation of new semantics: MVPA investigation of rapid neocortical plasticity during associative encoding through fast mapping. *Neural Plasticity, 2015*, Article 804385. https://doi.org/10.1155/2015/804385

Baart, M., & Samuel, A. G. (2015). Early processing of auditory lexical predictions revealed by ERPs. *Neuroscience Letters, 585*, 98–102. https://doi.org/10.1016/j.neulet.2014.11.044

Bakker, I., Takashima, A., van Hell, J. G., Janzen, G., & McQueen, J. M. (2015). Tracking lexical consolidation with ERPs: Lexical and semantic-priming effects on N400 and LPC responses to newly-learned words. *Neuropsychologia, 79*, Part A, 33–41. https://doi.org/10.1016/j.neuropsychologia.2015.10.020.

Barber, H. A., Otten, L. J., Kousta, S.-T., & Vigliocco, G. (2013). Concreteness in word processing: ERP and behavioral effects in a lexical decision task. *Brain and Language, 125*(1), 47–53. https://doi.org/10.1016/j.bandl.2013.01.005

Batterink, L., & Neville, H. (2011). Implicit and explicit mechanisms of word learning in a narrative context: An event-related potential study. *Journal of Cognitive Neuroscience, 23*(11), 3181–3196. https://doi.org/10.1162/jocn_a_00013

Bermúdez-Margaretto, B., Beltrán, D., Domínguez, A., & Cuetos, F. (2015). Repeated exposure to “meaningless” pseudowords modulates LPC, but not N(FN)400. *Brain Topography, 28*(6), 838–851. https://doi.org/10.1007/s10548-014-0403-5

Besson, M., Kutas, M., & Petten, C. V. (1992). An event-related potential (ERP) analysis of semantic congruity and repetition effects in sentences. *Journal of Cognitive Neuroscience, 4*(2), 132–149. https://doi.org/10.1162/jocn.1992.4.2.132

Boehm, S. G., & Sommer, W. (2005). Neural correlates of intentional and incidental recognition of famous faces. *Cognitive Brain Research, 23*(2), 153–163. https://doi.org/10.1016/j.cogbrainres.2004.10.008

Borovsky, A., Kutas, M., & Elman, J. (2010). Learning to use words: Event-related potentials index single-shot contextual word learning. *Cognition, 116*(2), 289–296. https://doi.org/10.1016/j.cognition.2010.05.004

Candia-Rivera, D., & Valenza, G. (2022). Cluster permutation analysis for EEG series based on non-parametric Wilcoxon–Mann–Whitney statistical tests. *SoftwareX, 19*, Article 101170. https://doi.org/10.1016/j.softx.2022.101170

Chen, B., Ma, T., Liang, L., & Liu, H. (2017). Rapid L2 word learning through high constraint sentence context: An event-related potential study. *Frontiers in Psychology, 8*(2285). https://doi.org/10.3389/fpsyg.2017.02285

Chen, J., & Chen, B. (2023). Learning new meanings for known L2 words: Long-term semantic representation is updated to integrate new information after consolidation. *Psychophysiology, 60*(5). https://doi.org/10.1111/psyp.14228

Christian, A. M., & Paterson, D. G. (1936). Growth of vocabulary in later maturity. *Journal of Psychology: Interdisciplinary and Applied, 1*, 167–169.

Clay, F., Bowers, J. S., Davis, C. J., & Hanley, D. A. (2007). Teaching adults new words: The role of practice and consolidation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(5), 970–976. https://doi.org/10.1037/0278-7393.33.5.970

Cooper, E., Greve, A., & Henson, R. N. (2019). Little evidence for Fast Mapping (FM) in adults: A review and discussion. *Cognitive Neuroscience, 10*(4), 196–209. https://doi.org/10.1080/17588928.2018.1542376

Coutanche, M. N., & Thompson-Schill, S. L. (2014). Fast mapping rapidly integrates information into existing memory networks. *Journal of Experimental Psychology: General, 143*, 2296–2303. https://doi.org/10.1037/xge0000020

Curran, T., & Cleary, A. M. (2003). Using ERPs to dissociate recollection from familiarity in picture recognition. *Cognitive Brain Research, 15*(2), 191–205. https://doi.org/10.1016/S0926-6410(02)00192-1

Cycowicz, Y. M., & Friedman, D. (1999). The effect of intention to learn novel, environmental sounds on the novelty P3 and old/new recognition memory. *Biological Psychology, 50*(1), 35–60. https://doi.org/10.1016/S0301-0511(99)00004-6

Davis, M. H., Di Betta, A. M., Macdonald, M. J. E., & Gaskell, M. G. (2009). Learning and consolidation of novel spoken words. *Journal of Cognitive Neuroscience, 21*(4), 803–820. https://doi.org/10.1162/jocn.2009.21059

De Diego Balaguer, R., Toro, J. M., Rodriguez-Fornells, A., & Bachoud-Lévi, A.-C. (2007). Different neurophysiological mechanisms underlying word and rule extraction from speech. *PLoS One, 2*(11), e1175.

de Vos, J. F., Schriefers, H., & Lemhöfer, K. (2019). Noticing vocabulary holes aids incidental second language word learning: An experimental study. *Bilingualism: Language and Cognition, 22*(3), 500–515. https://doi.org/10.1017/S1366728918000019

de Vos, J. F., Schriefers, H., Nivard, M. G., & Lemhöfer, K. (2018). A meta-analysis and meta-regression of incidental second language word learning from spoken input. *Language Learning, 68*(4), 906–941. https://doi.org/10.1111/lang.12296

de Vos, J. F., Schriefers, H., ten Bosch, L., & Lemhöfer, K. (2019). Interactive L2 vocabulary acquisition in a lab-based immersion setting. *Language, Cognition and Neuroscience, 34*(7), 916–935. https://doi.org/10.1080/23273798.2019.1599127

Duarte, A., Ranganath, C., Winward, L., Hayward, D., & Knight, R. T. (2004). Dissociable neural correlates for familiarity and recollection during the encoding and retrieval of pictures. *Cognitive Brain Research, 18*(3), 255–272. https://doi.org/10.1016/j.cogbrainres.2003.10.010

Dumay, N., & Gaskell, M. G. (2007). Sleep-associated changes in the mental representation of spoken words. *Psychological Science, 18*(1), 35–39. https://doi.org/10.1111/j.1467-9280.2007.01845.x

Eagle, M., & Leiter, E. (1964). Recall and recognition in intentional and incidental learning. *Journal of Experimental Psychology, 68*(1), 58–63. https://doi.org/10.1037/h0044655

Eugster, M. J. A., Ruotsalo, T., Spapé, M. M., Kosunen, I., Barral, O., Ravaja, N., Jacucci, G., & Kaski, S. (2014). Predicting term-relevance from brain signals. Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval, Gold Coast, Queensland, Australia. https://doi.org/10.1145/2600428.2609594.

Fernández, G., Effern, A., Grunwald, T., Pezer, N., Lehnertz, K., Dümpelmann, M., Roost, D. V., & Elger, C. E. (1999). Real-time tracking of memory formation in the human rhinal cortex and hippocampus. *Science, 285*(5433), 1582. https://doi.org/10.1126/science.285.5433.1582

Friedman, D., & Johnson, R., Jr (2000). Event-related potential (ERP) studies of memory encoding and retrieval: A selective review. *Microscopy Research and Technique, 51*(1), 6–28. https://doi.org/10.1002/1097-0029(20001001)51:1<6::AID-JEMT2>3.0.CO;2-R

Frishkoff, G. A., Perfetti, C. A., & Collins-Thompson, K. (2010). Lexical quality in the brain: ERP evidence for robust word learning from context. *Developmental Neuropsychology, 35*(4), 376–403. https://doi.org/10.1080/87565641.2010.480915

Gaskell, M. G., & Lindsay, S. (2019). Reasons to doubt the generalizability, reliability, and diagnosticity of fast mapping (FM) for rapid lexical integration. *Cognitive Neuroscience, 10*(4), 234–236. https://doi.org/10.1080/17588928.2019.1600487

Haese, A., & Czernochowski, D. (2015). Sometimes we have to intentionally focus on the details: Incidental encoding and perceptual change decrease recognition memory performance and the ERP correlate of recollection. *Brain and Cognition, 96*, 1–11. https://doi.org/10.1016/j.bandc.2015.02.003

Havas, V., Taylor, J. S. H., Vaquero, L., de Diego-Balaguer, R., Rodríguez-Fornells, A., & Davis, M. H. (2018). Semantic and phonological schema influence spoken word learning and overnight consolidation. *Quarterly Journal of Experimental Psychology, 71*(6), 1469–1481. https://doi.org/10.1080/17470218.2017.1329325

Holcomb, P. J., Grainger, J., & O'Rourke, T. (2002). An electrophysiological study of the effects of orthographic neighborhood size on printed word perception. *Journal of Cognitive Neuroscience, 14*(6), 938–950. https://doi.org/10.1162/089892902760191153

Hulstijn, J. H. (2003). Incidental and intentional learning. In C. J. Doughty, & M. H. Long (Eds.), *Handbook of second language acquisition* (pp. 349–381). Blackwell. https://doi.org/10.1002/9780470756492.ch12.

Kaczer, L., Bavassi, L., Petroni, A., Fernández, R. S., Laurino, J., Degiorgi, S., & Pedreira, M. E. (2018). Contrasting dynamics of memory consolidation for novel word forms and meanings revealed by behavioral and neurophysiological markers. *Neuropsychologia, 117*, 472–482. https://doi.org/10.1016/j.neuropsychologia.2018.07.001

Keuleers, E., Stevens, M., Mandera, P., & Brysbaert, M. (2015). Word knowledge in the crowd: Measuring vocabulary size and word prevalence in a massive online experiment. *Quarterly Journal of Experimental Psychology, 68*(8), 1665–1692. https://doi.org/10.1080/17470218.2015.1022560

Kissler, J., Herbert, C., Winkler, I., & Junghofer, M. (2009). Emotion and attention in visual word processing—An ERP study. *Biological Psychology, 80*(1), 75–83. https://doi.org/10.1016/j.biopsycho.2008.03.004

Kutas, M., & Federmeier, K. D. (2010). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology, 62*(1), 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (de) constructing the N400. *Nature Reviews Neuroscience, 9*, 920–933. https://doi.org/10.1038/nrn2532

Leinonen, A., Grönholm-Nyman, P., Järvenpää, M., Söderholm, C., Lappi, O., Laine, M., & Krause, C. M. (2009). Neurocognitive processing of auditorily and visually presented inflected words and pseudowords: Evidence from a morphologically rich language. *Brain Research, 1275*, 54–66. https://doi.org/10.1016/j.brainres.2009.03.057

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods, 44*(2), 325–343. https://doi.org/10.3758/s13428-011-0146-0

Lindsay, S., & Gaskell, M. G. (2010). A complementary systems account of word learning in L1 and L2. *Language Learning, 60*(s2), 45–63. https://doi.org/10.1111/j.1467-9922.2010.00600.x

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods, 164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Mecklinger, A., & Kamp, S.-M. (2023). Observing memory encoding while it unfolds: Functional interpretation and current debates regarding ERP subsequent memory effects. *Neuroscience & Biobehavioral Reviews, 153*, Article 105347. https://doi.org/10.1016/j.neubiorev.2023.105347

Merhav, M., Karni, A., & Gilboa, A. (2015). Not all declarative memories are created equal: Fast Mapping as a direct route to cortical declarative representations. *NeuroImage, 117*, 80–92. https://doi.org/10.1016/j.neuroimage.2015.05.027

Mestres-Missé, A., Rodriguez-Fornells, A., & Münte, T. F. (2007). Watching the brain during meaning acquisition. *Cerebral Cortex, 17*(8), 1858–1866. https://doi.org/10.1093/cercor/bhl094

Miralpeix, I. (2020). L1 and L2 vocabulary size and growth. In S. Webb (Ed.), *Routledge Handbook of Vocabulary Studies* (pp. 189–206). Routledge. https://doi.org/10.4324/9780429291586.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience, 2011*, Article 156869. https://doi.org/10.1155/2011/156869

Otten, L. J., & Rugg, M. D. (2001). Electrophysiological correlates of memory encoding are task-dependent. *Cognitive Brain Research, 12*(1), 11–18. https://doi.org/10.1016/S0926-6410(01)00015-5

Paller, K. A., Kutas, M., & Mayes, A. R. (1987). Neural correlates of encoding in an incidental learning paradigm. *Electroencephalography and Clinical Neurophysiology, 67*(4), 360–371. https://doi.org/10.1016/0013-4694(87)90124-6

Paller, K. A., McCarthy, G., & Wood, C. C. (1988). ERPs predictive of subsequent recall and recognition performance. *Biological Psychology, 26*(1), 269–276. https://doi.org/10.1016/0301-0511(88)90023-3

Paller, K. A., & Wagner, A. D. (2002). Observing the transformation of experience into memory. *Trends in Cognitive Sciences, 6*(2), 93–102. https://doi.org/10.1016/S1364-6613(00)01845-3

Palma, P., & Titone, D. (2021). Something old, something new: A review of the literature on sleep-related lexicalization of novel words in adults. *Psychonomic Bulletin & Review, 28*(1), 96–121. https://doi.org/10.3758/s13423-020-01809-5

Peters, J., & Daum, I. (2009). Frontal but not parietal positivity during source recollection is sensitive to episodic content. *Neuroscience Letters, 454*(3), 182–186. https://doi.org/10.1016/j.neulet.2009.03.019

Rabovsky, M., Hansen, S. S., & McClelland, J. L. (2018). Modelling the N400 brain potential as change in a probabilistic representation of meaning. *Nature Human Behaviour, 2*, 693–705. https://doi.org/10.1038/s41562-018-0406-4

Rugg, M. D., & Curran, T. (2007). Event-related potentials and recognition memory. *Trends in Cognitive Sciences, 11*(6). https://doi.org/10.1016/j.tics.2007.04.004

Rugg, M. D., Schloerscheidt, A. M., & Mark, R. E. (1998). An electrophysiological comparison of two indices of recollection. *Journal of Memory and Language, 39*(1), 47–69. https://doi.org/10.1006/jmla.1997.2555

Sanquist, T. F., Rohrbaugh, J. W., Syndulko, K., & Lindsley, D. B. (1980). Electrocortical signs of levels of processing: Perceptual analysis and recognition memory. *Psychophysiology, 17*(6), 568–576. https://doi.org/10.1111/j.1469-8986.1980.tb02299.x

Sassenhagen, J., & Draschkow, D. (2019). Cluster-based permutation tests of MEG/EEG data do not establish significance of effect latency or location. *Psychophysiology, 56*(6). https://doi.org/10.1111/psyp.13335

Schmitt, N., & Schmitt, D. (Eds.). (2020). *Vocabulary in language teaching* ((2nd ed.).). Cambridge University Press. https://doi.org/10.1017/9781108569057.

Schott, B., Richardson-Klavehn, A., Heinze, H.-J., & Düzel, E. (2002). Perceptual priming versus explicit memory: Dissociable neural correlates at encoding. *Journal of Cognitive Neuroscience, 14*(4), 578–592. https://doi.org/10.1162/08989290260045828

Sharon, T., Moscovitch, M., & Gilboa, A. (2011). Rapid neocortical acquisition of long-term arbitrary associations independent of the hippocampus. *Proceedings of the National Academy of Sciences, 108*(3), 1146. https://doi.org/10.1073/pnas.1005238108

Shtyrov, Y., Nikulin, V. V., & Pulvermüller, F. (2010). Rapid cortical plasticity underlying novel word learning. *Journal of Neuroscience, 30*(50), 16864. https://doi.org/10.1523/JNEUROSCI.1376-10.2010

Soskey, L., Holcomb, P. J., & Midgley, K. J. (2016). Language effects in second-language learners: A longitudinal electrophysiological study of Spanish classroom learning.

Brain Research, 1646(Supplement C), 44–52. https://doi.org/10.1016/j.brainres.2016.05.028.

Supp, G. G., Schlogl, A., Gunter, T. C., Bernard, M., Pfurtscheller, G., & Petsche, H. (2004). Lexical memory search during N400: Cortical couplings in auditory comprehension. *NeuroReport, 15*(7), 1209–1213. https://doi.org/10.1097/01.wnr.0000126219.25235.¡

Swingley, D. (2010). Fast mapping and slow mapping in children's word learning. *Language Learning and Development, 6*(3), 179–183. https://doi.org/10.1080/15475441.2010.484412

Tamminen, J., & Gaskell, M. G. (2013). Novel word integration in the mental lexicon: Evidence from unmasked and masked semantic priming. *Quarterly Journal of Experimental Psychology, 66*(5), 1001–1025. https://doi.org/10.1080/17470218.2012.724694

Turk, K. W., Elshaar, A., & a. A., Deason, R. G., Heyworth, N. C., Nagle, C., Frustace, B., Flannery, S., Zumwalt, A., & Budson, A. E.. (2018). Late positive component event-related potential amplitude predicts long-term classroom-based learning. *Journal of Cognitive Neuroscience, 30*(9), 1323–1329. https://doi.org/10.1162/jocn_a_01285

van Hooff, J. C. (2005). The influence of encoding intention on electrophysiological indices of recognition memory. *International Journal of Psychophysiology, 56*(1), 25–36. https://doi.org/10.1016/j.ijpsycho.2004.09.010

Van Petten, C., Kutas, M., Kluender, R., Mitchiner, M., & McIsaac, H. (1991). Fractionating the word repetition effect with event-related potentials. *Journal of Cognitive Neuroscience, 3*(2), 131–150. https://doi.org/10.1162/jocn.1991.3.2.131

Van Petten, C., & Senkfor, A. J. (1996). Memory for words and novel visual patterns: Repetition, recognition, and encoding effects in the event-related brain potential. *Psychophysiology, 33*(5), 491–506.

Webb, S., Yanagisawa, A., & Uchihara, T. (2020). How effective are intentional vocabulary-learning activities? A meta-analysis. *The Modern Language Journal, 104*(4), 715–738. https://doi.org/10.1111/modl.12671

Yang, H., Laforge, G., Stojanoski, B., Nichols, E. S., McRae, K., & Köhler, S. (2019). Late positive complex in event-related potentials tracks memory signals when they are decision relevant. *Scientific Reports, 9*(1), 9469. https://doi.org/10.1038/s41598-019-45880-y

Zhang, J., Huang, Y., Jiang, C., Xu, Y., Rao, H., & Xu, H. (2023). Dynamic brain responses to Russian word acquisition among Chinese adult learners: An event-related potential study. *Human Brain Mapping, 44*(9), 3717–3729. https://doi.org/10.1002/hbm.26307

Zhou, T., Li, J., Broster, L. S., Niu, Y., & Wang, P. (2015). Reduced late positivity in younger adults, but not older adults, during short-term repetition. *Brain Research, 1594*, 223–232. https://doi.org/10.1016/j.brainres.2014.10.042

Ziegler, J. C., Besson, M., Jacobs, A. M., Nazir, T. A., & Carr, T. H. (1997). Word, pseudoword, and nonword processing: A multitask comparison using event-related brain potentials. *Journal of Cognitive Neuroscience, 9*(6), 758–775. https://doi.org/10.1162/jocn.1997.9.6.758