Lexical access versus lexical decision processes for auditory, visual, and audiovisual items: Insights from behavioral and neural measures.

Rocío A. López Zunini[a*], Martijn Baart[a,b], Arthur G. Samuel[a,c,d] and Blair C. Armstrong[a,e]

Affiliations:

[a] BCBL, Basque Center on Cognition, Brain and Language, San Sebastian, Spain

[b] Tilburg University, Dept. of Cognitive Neuropsychology, Tilburg, The Netherlands

[c] IKERBASQUE, Basque Foundation for Science, Spain

[d] Stony Brook University, Dept. of Psychology, Stony Brook, NY, United States

[e] University of Toronto, Dept. of Psychology and Centre for French & Linguistics, Toronto, ON, Canada

*Corresponding author at Basque Center on Cognition, Brain and Language, Paseo Mikeletegi 69, 2nd floor, 20009, San Sebastian, Spain. Email address: lopez.zunini@gmail.com

Abstract

In two experiments, we investigated the relationship between lexical access processes, and processes that are specifically related to making lexical decisions. In Experiment 1, participants performed a standard lexical decision task in which they had to respond as quickly and as accurately as possible to visual (written), auditory (spoken) and audiovisual (written+spoken) items. In Experiment 2, a different group of participants performed the same task but were required to make responses after a delay. Linear mixed effect models on reaction times and single trial Event-Related Potentials (ERPs) revealed that ERP lexicality effects started earlier in the visual than auditory modality, and that effects were driven by the written input in the audiovisual modality. More negative ERP amplitudes predicted slower reaction times in all modalities in both experiments. However, these predictive amplitudes were mainly observed *within* the window of the lexicality effect in Experiment 1 (the speeded task), and shifted to post-response-probe time windows in Experiment 2 (the delayed task). The lexicality effects lasted longer in Experiment 1 than in Experiment 2, and in the delayed task, we additionally observed a "re-instantiation" of the lexicality effect related to the delayed response. Delaying the response in an otherwise identical lexical decision task thus allowed us to separate lexical access processes from processes specific to lexical decision.

Keywords: visual word recognition, auditory word recognition, single-trial Event-Related Potentials, lexical decision, lexical processing

Introduction

Language use has played an enormous role in the advances that humans have made over the millennia, allowing people to share knowledge in a way that is not available to other species. Perhaps the most critical component of language is that one person can generate words to express a thought, and a second person can recognize those words, and thus the thought behind them. From this perspective, it is clear that understanding how word recognition operates provides important information about fundamental human cognitive processes.

One of the most straightforward ways to investigate word processing in a laboratory setting is to compare how words are perceived relative to pseudowords. The latter are items that are phonotactically legal in a language and could thus technically be a word, but are not (e.g., *flyke* or *dobry* in English). There is a very well-established "lexicality effect" in which words are recognized more quickly and more accurately than pseudowords (Reicher, 1969; Taft, 1986; Wheeler, 1970). One of the most widely used tasks that captures this effect is the lexical decision task (LDT) in which participants indicate whether items are real words or not (i.e., they make a lexical decision). In this paradigm, participants are often asked to "respond as quickly and accurately as possible", with the items presented auditorily or visually, as spoken input or print.

However, explicit lexical decisions are not necessary to investigate how the lexicon operates, because lexical processing occurs in response to linguistic input, even in the absence of a specific task. Electrophysiological techniques such as EEG are capable of revealing the time-course of lexical processing without using tasks in which overt responses are collected on each trial (e.g., Laszlo & Federmeier, 2007; Laszlo & Federmeier, 2012, 2014; van den Brink, Brown, & Hagoort, 2001; van den Brink & Hagoort, 2004). If no response is required, the EEG signal is not contaminated by (pre)motor activity related to preparing, generating, or executing a response;

it is also free of activity that is specifically related to task demands. However, overt tasks have advantages as well, as they allow researchers to single out trials that were perceived as intended (i.e., that are responded to correctly) (e.g., López Zunini, Renoult, & Taler, 2017; Rabovsky, Sommer, & Abdel Rahman, 2012; Taler, Kousaie, & López Zunini, 2013) and can unravel the relationship between cognitive processes and the perceptual outcome that is measured via a response (e.g., Baart, Armstrong, Martin, Frost, & Carreiras, 2017).

For lexical processing, one particularly relevant and well-established electrophysiological marker is the N400 (Holcomb, 1993; Kutas & Hillyard, 1980, 1983, 1984). The N400 is a negative deflection in the Event Related Potential (ERP) that is observed approximately 400 ms after stimulus presentation. Depending on the context and task at hand, smaller N400s can be associated with stronger lexical activation. For example, in word-pseudoword comparisons, or when comparing familiar to unfamiliar words, smaller N400s index stronger lexical activation (Bentin, Mouchetant-Rostaing, Giard, Echallier, & Pernier, 1999; Laszlo & Federmeier, 2007; Rugg & Nagy, 1987) (Rugg, 1990; Van Petten & Kutas, 1990). In the context of orthographic similarity, pseudowords with high neighborhood size produce larger N400s than those with low neighborhood size, and this effect can be ascribed to activation of lexical information similar to the pseudoword (e.g. Holcomb, Grainger, & O'Rourke, 2002).

In this study, we will measure lexical activation by observing the well established word/pseudoword N400 effect (e.g., Kounios & Holcomb, 1994; López Zunini et al., 2017; Muller, Duñabeitia, & Carreiras, 2010; Rabovsky et al., 2012; Taler et al., 2013). It is critical to realize that the cognitive process of interest – which we will refer to as *lexical access* – may be confounded with (potentially very similar) processes related to making an explicit lexical decision. The decision system requires lexical evidence (generated via lexical access) to drive a

response that requires involvement of (pre)motor areas. The underlying processes are therefore difficult to identify uniquely: the measures may reflect the (partial) outcome of lexical access, the actual decision, the response, or a combination of these elements. Here, we do not intend to disentangle all of these possibilities, but we do want to distinguish lexical access from the decision and response aspects. We will refer to the processes related to making an explicit lexical decision as *LDR processes* (LDR for Lexical Decision Response), which we will contrast with *lexical access*.

Lexical access and LDR processes are assumed to be independent systems, each responsible for different aspects of performance (e.g., Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Plaut, McClelland, Seidenberg, & Patterson, 1996) but at the same time, they have to be closely related. As noted, LDR processes require input from lexical access processes that accumulate evidence about the lexical status of an item. However, lexical access may not need to be fully completed before LDR processes are initiated. If so, then even if lexical access and LDR processes are independent, they can overlap in time. Here, we seek to tease the two processes apart by measuring EEG while participants make lexical decisions. In two experiments, we presented participants with words and pseudowords that should provide a clear lexicality effect in the ERPs. The critical difference between the two experiments was the nature of the task. In Experiment 1, participants had to make a lexical decision as quickly and accurately as they could (a speeded task, see for example Holcomb et al., 2002; López Zunini et al., 2017; Rabovsky et al., 2012; Taler et al., 2013) whereas in Experiment 2, we use a non-speeded version of the task. We assumed that the time constraints imposed by the speeded task would force the lexical access and LDR processes to co-occur in time, whereas LDR processes are pushed away from lexical access in the non-speeded task (Experiment 2).

Our primary interest is in the N400 effect that has repeatedly been found for the word-pseudoword manipulation. There have been a number of observations of lexicality effects prior to the N400 (e.g., Baart & Samuel, 2015; MacGregor, Pulvermüller, van Casteren, & Shtyrov, 2012; van den Brink et al., 2001; van den Brink & Hagoort, 2004), but the N400 does not seem to be entirely post-lexical (Deacon, Hewitt, Yang, & Nagata, 2000). Although N400 amplitudes are modulated by neighbourhood size and the number of lexical associates – indicating semantic processes are at play (e.g. Laszlo & Federmeier, 2012) - important parts of the lexical process are nevertheless reflected in the N400. In our analyses, we do not make any *a priori* assumption about the timing of lexical processes. Instead, we use a fine-grained approach based on linear mixed effects regressions on single trial data, starting from stimulus onset.

Crossed with our manipulation of when people could respond, we manipulated how soon participants had the necessary input to recognize a word. As noted, LDTs can be administered with spoken items as well as visual items in the form of text. The semantic properties of a word are presumably the same whether it is spoken or written (Rogers et al., 2004), but the input codes for activating word meanings are quite different, reflecting the different constraints on word perception across modalities. For example, a letter string – and thus, the lexical information – is presented all at once, and research has shown that people can indeed take in the entire word in one look (provided it is not particularly long; see Kutas & Van Petten, 1994 for a review). This contrasts with spoken words, in which the stimulus unfolds over time, as does the degree to which the lexicon can provide conclusive evidence of its lexical status.

We also included items in which the participant both hears the item, and sees it in orthographic form, at the same time. Since writing systems are grounded in auditory language systems that have been shaped by evolution, such audiovisual stimuli provide an interesting test

case. On the one hand, the visual signal may dominate the perceptual response (on a behavioral and/or neural level) because lexical information is more readily available in terms of stimulus presentation time in printed items than spoken ones. On the other, the auditory element of the stimulus reflects the dominant linguistic modality: For millennia, humans used spoken language without any written language (young children still do), and orthographic input is often argued to be recoded into phonology (see e.g., Coltheart et al., 2001; Frost, 1998, for extensive reviews).

Across the two experiments our goal is to separate the core lexical access process from the processes related to making an explicit lexical decision. In Experiment 1, we examine the relationship between moment-by-moment ERP amplitude and response times in order to determine whether lexical access and LDR processes indeed overlap when responses are speeded. In Experiment 2, we use a delayed response procedure to force the LDR processes away from the lexical access process, allowing us to examine both processes without their mutual confounding.

Experiment 1

Methods

Participants

Twenty-one participants (5 males; mean age: 22 years; age range 18-28) were recruited through the Basque Center on Cognition, Brain and Language participant database. They gave written informed consent prior to testing. The experiment was conducted in accordance with the declaration of Helsinki and approved by the BCBL internal ethics committee.

All participants were right-handed native speakers of Spanish who received 10€/h for their participation. Prior to the testing session, they completed a self-report health and history

questionnaire. They did not suffer from any disorders or neurological problems, and were not taking medication that could affect cognitive function. All participants had self-reported (corrected to) normal vision and adequate hearing.

Participants from experiment 1 did not differ from participants in experiment 2 in age, education or language skills. Table 1 presents the demographic and language characteristics.

Table 1. Demographic information and language skills. Each variable is represented by the mean and standard deviation in parenthesis.

Variable	Speeded LDT (N=21)	Delayed LDT (N=21)	Independent Sample t tests
Age	22.47(3.14)	22.90(2.53)	$t(40) = -0.49, p = 0.63$
Education	15.90(1.99)	15.95(1.59)	$t(40) = -0.09, p = 0.93$
Reading*	9.71(0.56)	9.80(0.41)	$t(39) = -0.56, p = 0.58$
Understanding*	9.76(0.44)	9.75(0.44)	$t(39) = 0.08, p = 0.93$
Writing*	9.57(0.68)	9.70(0.66)	$t(39) = -0.62, p = 0.54$
Speaking*	9.71(0.46)	9.76(0.54)	$t(39) = -0.31, p = 0.76$

* Data missing from one participant

Stimuli

The experimental items consisted of 300 words and 300 pseudowords (see Appendix A for a full list). The stimuli were a subset of those used in a previous study (Baart et al., 2017), which were down-sampled to reduce the total length of the experiment. All words were non-homophones with word frequencies between 1 and 20 per million, and had one noun meaning (although ~10% of the items could also be associated with other grammatical categories). Descriptive statistics for the experimental words are presented in Appendix A. The normative data for these items were obtained from the EsPal database (Duchon, Perea, Sebastian-Galles, Marti, & Carreiras, 2013), supplemented by additional positional bigram frequency data (http://www.blairarmstrong.net/tools/#Bigram).

Phonotactically plausible Spanish pseudowords were generated with the Wuggy nonword generation tool (Keuleers & Brysbaert, 2010). Pseudowords with the lowest orthographic

Levenshtein distance that did not appear to be a misspelling or mispronunciation a word (determined by a native Spanish speaker) were selected, and orthographic accents were added to increase word-likeness. A detailed description of the procedure and descriptive statistics for the pseudowords are presented in Appendix A. Independent sample t-tests indicated that the word stimuli had higher average bigram frequencies ($t(598) = 2.03$, $p = .04$), and more neighbors than pseudowords, as supported by smaller Orthographic Levenshtein Distance values, ($t(598) = 12.30$, $p < .001$).

The audio stimuli were recorded by a female Spanish native speaker who read the stimuli one at a time in a randomly ordered list. The list was read twice, once from start-to-end and once from end-to-start, to obtain two recordings of each item. Both versions were then cut with Audacity (Mazzoni, 2013) and the more natural sounding item (judged by a native speaker) was used in the study.

Task

The experiment was implemented using PsychoPy v1.84.1 (Pierce, 2007), and run on a standard desktop computer equipped with a 19-inch CRT monitor running at 100 Hz (screen resolution of 1024 px. $\times$ 768 px). Sounds were delivered at ~65 dBA (measured at ear level) via two front facing computer speakers (JBL by Harman, Duet) placed on both sides of the monitor. Visual stimuli were displayed in Arial font (font height was 5% of the display height, or 38 px), visual angle 0.7 or 0° 42' 0.97".

The task consisted of twelve experimental blocks of 25 words and 25 pseudowords, yielding a total of 600 trials (300 words and 300 pseudowords). Stimuli were presented in three sensory modalities: visual (V), audio (A), and audiovisual (AV). Four blocks were V (100 words and 100 pseudowords in total), four were A, and four were AV. Two successive blocks were never from the same modality. Trial order was randomized, with the constraint that no more than

three words or pseudowords could be presented in a row to reduce trial carry-over effects

(Armstrong & Plaut, 2011, 2016). Before the experimental task started, participants completed

three practice blocks (one for each sensory modality) with four trials each (two words and two

pseudowords). Self-paced breaks were allowed between blocks.

For each participant, the order of experimental items and conditions was determined as

follows: first, we generated six counterbalanced lists for the different orderings of the 600 trials

in the A, AV, and V conditions. The same ordering of conditions was used across sets of 12

blocks for each participant (e.g., for one participant, the order was 50 trials of AV, 50 trials of V,

50 trials of A, 50 trials of AV, 50 trials of V, 50 trials of A, etc.). We then randomly ordered all

of the word and pseudoword items in a list, subject to constraints on item type repetition (no

more than 3 words or pseudowords in a row). For each of the six counterbalanced lists of

conditions, we then merged the information about the condition orders with the randomly

ordered list of items to determine which items appeared for each participant in each condition.

This perfectly counterbalances the appearance of each item across all experimental conditions

over a series of six participants. After we had generated the list of items and associated

conditions for six participants, we re-randomized the list of items so that a new ordering of

experimental items was used. This regeneration avoided any systematic list effects or list order

effects. In so doing, we aimed to maximize the generalizability of our results while avoiding

systematic bias across experimental lists. Thus, all participants were presented with the same

words and pseudowords, but potentially in different modalities. Across participants, the words

and pseudowords were presented equally often in each modality.

Each trial began with a white fixation cross (+) that was presented for 750 ms in the

center of a black screen, which was followed by a black screen that was randomly jittered in

duration (between 1500 ms and 2000 ms) before stimulus onset. During V and AV trials, white text in was presented in the center of the screen for 2000 ms, while the screen remained blank for A stimuli. On AV trials, visual and auditory onsets were simultaneous, and the text was always congruent with the auditorily presented stimulus. Participants indicated whether the stimulus was a word or pseudoword by pressing the right or left control keys, respectively, and were instructed to respond as quickly as possible. A warning message indicating they were too slow was displayed if participants did not respond within 2000 ms. The next trial began 250 ms after a response was collected. The task took approximately 50 minutes to complete.

Testing protocol

Participants performed the task while the electroencephalogram (EEG) was recorded. They were seated approximately 80 cm from the computer monitor in a sound-attenuated, dimly lit, and electrically shielded booth.

Performance analyses

Reaction time of accurate responses (RT) and accuracy analyses were conducted with linear mixed-effect models (Baayen, Davidson, & Bates, 2008) using the lme4 package in R (Bates, Maechler, Bolker, & Walker, 2015). After removing all RTs smaller than 200 ms, RT was modeled with a Gaussian distribution both on the raw data and inverse RT. Accuracy was modeled with a binomial distribution. Significance was assessed with lmerTest package in R, which uses the Satterthwaite´s approximation method for degrees of freedom and p values (Kuznetsova, Brockhoff, & Christensen, 2017).

EEG recording, processing and analyses

EEG was recorded with a 32 channel BrainAmp system (Brain Products GmbH, Munich, Germany) at a sampling rate of 250 Hz. Twenty-seven Ag/AgCl electrodes placed in an EasyCap recorded the EEG from sites Fp1, Fp2, F7, F3, Fz, F4, F8, FC5, FC1, FC2, FC6, T7, C3, Cz, C4, T8, CP5, CP1, CP2, CP6, P7, P3, Pz, P4, P8, O1 and O2. An electrode at FCz served as ground and an electrode on the left mastoid served as on-line reference. Additional electrodes were placed on the right mastoid, above and below the right eye (to record vertical electro-oculogram; EOG), and the left and right canthi (to record horizontal electro-oculogram). Impedances were set below 5 kΩ for the mastoids and cap electrodes and below 10 kΩ for the horizontal and vertical EOG electrodes.

The EEG signal was processed off-line using Brain Vision Analyzer 2.0 (Brain Products GmbH, Munich, Germany). The signal was re-referenced to an average of the two mastoids and was digitally filtered at a low cutoff of 0.1 Hz at 24 dB per octave. Next, the signal was decomposed into independent components (Jung et al., 2000) with restricted infomax based on the entire data set. Components that captured blinks or horizontal eye-movements and EMG bursts, identified through visual inspection based on components' energy and topography, were removed. The mean number of removed components was 4.8. Next, the data were filtered with a high cutoff of 40 Hz at 24 dB per octave, and an additional 50 Hz notch filter was applied to remove residual electrical interference.

ERPs were time locked to the onset of the stimuli and segmented into 2700 ms epochs with a 200 ms pre-stimulus baseline. Segments with artifacts were rejected. Artifacts were defined as: 1) activity < -100 or $> +100$ µV in the entire epoch, 2) a > 100 µV difference in 200 ms intervals, and 3) activity < 0.5 µV in a 100 ms interval. On average, 4.6 % of the trials were rejected. In order to calculate single trial mean amplitudes, the segments were then imported into

MATLAB v.2014b (MathWorks, Inc., Natick, Massachusetts, United States) and read using

EEGlab (Delorme & Makeig, 2004). For each participant, the single-trial segments were divided

into 50 ms bins (i.e., there were 50 segments: 0-50 ms, 50-100 ms, … 2450-2500 ms post-

stimulus) and mean amplitudes for each trial were calculated with the 'meanepoch function' in

ERP lab (Lopez-Calderon & Luck, 2014).

The ERP analyses were performed in two ways: first, we performed a series of linear

mixed effect regression models on all electrodes (27 in total) and 50 time bins (from 0 to 2500

ms post stimulus onset). This analysis provides a comprehensive assessment of the results.

Second, to provide a more focused assessment, we identified a central cluster of electrodes in

which N400 amplitudes for words (and pseudowords) did not significantly differ from each

other. To identify an appropriate central cluster, we first conducted a 27 x 2 repeated ANOVA

for each modality. The within-subject factors were: Electrode (27 levels: all electrodes) and

Lexicality (2 levels: words and pseudowords). Given the N400 timing differences in the visual

and audiovisual modalities relative to the auditory modality, the analyses were conducted on

mean amplitudes from 300-500 ms post-stimulus onset for the visual and audiovisual modalities,

and from 500-1000 ms for the auditory modality. Only correct trials were analyzed. For each

modality, the ANOVA revealed an interaction with electrode and lexicality (V: $F(26, 936) =$

$5.81, p < .001$; A: $F(26, 936) = 4.92, p = .001$; AV: $F(26, 936) = 3.10, p = .014$). Follow-up pair-

wise comparisons showed that for words, amplitudes at C3, Cz and C4 did not significantly

differ; this was also the case for pseudowords (all p's $> .5$). We therefore averaged amplitudes

across these three electrodes (the central cluster) and conducted linear mixed effect models on

those data. In all models, significance was assessed with the lmerTest package in R, which uses

the Satterthwaite´s approximation (Kuznetsova et al., 2017).

Results

Behavioral results

The models for RT and accuracy contained crossed random effects (intercepts) of participants and items, and fixed effects of lexicality (words were considered baseline) and modality (with the V modality acting as a baseline for contrasts with the other two modalities, i.e, V vs. A; V vs. AV). Lexicality and modality were allowed to interact. An additional model was performed with A as the modality baseline instead, in order to include the A vs. AV contrast in the analyses. The model on the raw RT data provides the estimates in milliseconds and represents the most straightforward analysis of the data. However, the residual variance in the model was not normally distributed according to a Kolmogorov-Smirnov test of normality. For this reason, we also report an analogous model applied to inverse RTs (i.e., 1/RT), which does not violate the normality assumption and which also offers an intuitively interpretable measure (i.e., a measure of response speed). For completeness, we therefore report the results of both models, which produced qualitatively similar results (note that the sign of the tests is expected to reverse when analyzing inverse RTs because shorter RTs correspond to higher response speeds). Figure 1 displays mean RT and accuracy for words and pseudowords broken down by presentation modality. Tables 2 and 3 present the statistics from the regression models.
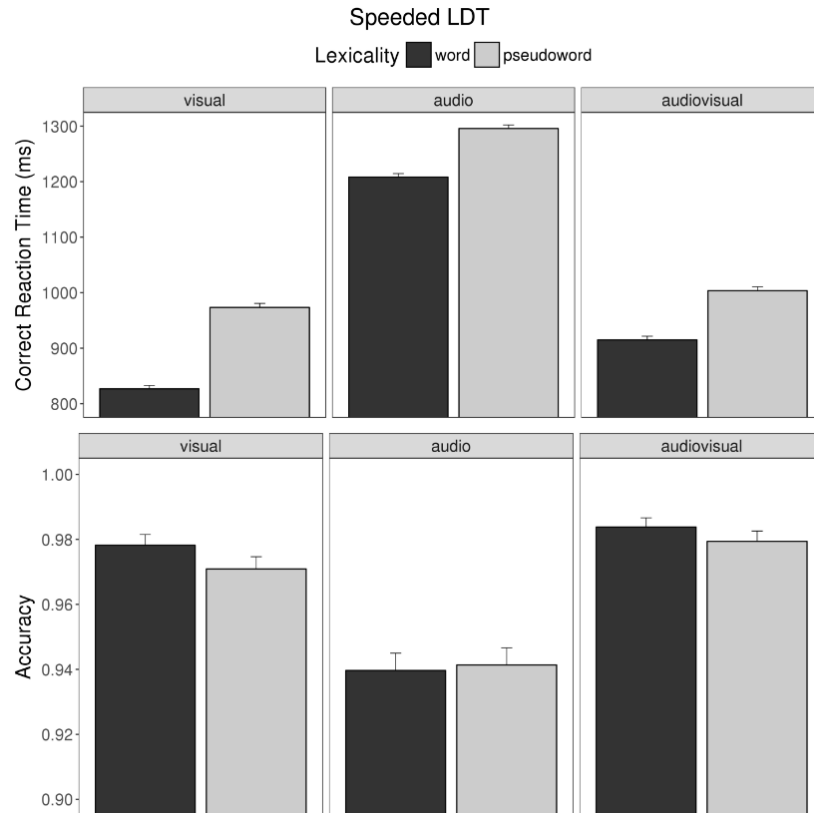
Table 2. Summary of LMER statistics for RT.

Model 1 (words and visual modality as baseline)	Raw RT			Inverse RT		
Fixed effects	b	SE	T	b	SE	T

Lexicality (pseudoword)	147	9	16.43**	-0.0002	0.000009	-21.95**
Modality (auditory)	383	7	49.54**	-0.00045	0.000007	-62.03**
Modality (audiovisual)	86	8	11.26**	-0.00011	0.000007	-15.35**
Lexicality X Modality (auditory)	-61	11	-5.62**	0.00012	0.00001	12.73**
Lexicality X Modality (audiovisual)	-55	2	-5.14**	0.00007	0.00001	7.15**
Model 2 (words and auditory as baseline)						
Modality (audiovisual)	-296	8	-38.4**	0.0003	0.000007	46.78**
Lexicality X Modality (audiovisual)	5.68	11	0.52	-0.00005	0.00001	-5.63*

*p<.01, **p<.001

Visual words were responded to 147 ms faster than visual pseudowords. In addition, words in the visual modality were responded to 383 ms faster than words in the auditory modality and 86 ms faster than words in the audiovisual modality. The significant interaction reflected a lexicality effect that was 61 ms larger for V than for A, and 55 ms larger for V than for AV. A and AV significantly interacted with lexicality – indicating a slightly larger effect in the AV modality than in A – but only in the model with inverse RT.

Figure 1. Mean reaction time (top) and accuracy (bottom) for words and pseudowords. Error bars represent standard errors.

Table 3. Summary of LMER statistics for accuracy.

Model 1 (words and visual modality as baseline)			
Fixed effects	**b**	**SE**	**z**
Lexicality (pseudoword)	-0.35	0.23	-1.55

Modality (auditory)	-1.15	0.18	-6.19**
Modality (audiovisual)	0.29	0.24	1.23
Lexicality X Modality (auditory)	0.38	0.25	1.53
Lexicality X Modality (audiovisual)	0.07	0.32	0.21
Model 2 (words and auditory as baseline)			
Modality (audiovisual)	1.01	0.16	6.12**
Lexicality X Modality (audiovisual)	0.18	0.23	0.80

**p<.001

For visual items, accuracy was comparable for words and pseudowords. The lexicality effect for A and AV did not differ from the V condition. For words, accuracy was higher for V than for A but similar to AV.

ERP results

Figure 2 presents the ERP results, and clearly shows a robust lexicality effect for all three presentation modalities. In all three panels there is a substantially more negative waveform for pseudowords than for words. This broad similarity is accompanied by differences in the onset and duration of the lexicality effect across modalities. For visual and audiovisual items, the word and pseudoword waveforms begin to diverge at about 300 ms. In contrast, for auditory items, this effect begins approximately 400 ms later. The onset timing for the Audiovisual case suggests that the lexical effect is initiated by the visual input.

The ERP data were first analyzed in a set of linear mixed effects analyses on the averaged ERP amplitudes in 50 ms time bins (the entire epoch contained 2500 ms, corresponding to 50 bins) of all electrodes, and then in a second set of analyses at the central cluster. We constructed separate models for each modality of presentation. Each included crossed random effects (intercepts) of participants and items, and fixed effects of lexicality (words as baseline), and modality (V as baseline)[1]. For the analyses on the full set of electrodes, there were 1350 comparisons per modality (27 electrodes x 50 time bins)[2], and for the analyses on the central cluster, there were 50 comparisons per modality (i.e., 50 time bins). FDR correction (Benjamini & Hochberg, 1995) was applied to correct for multiple comparisons.

---

[1] Given the substantial differences between modality of presentation, an initial omnibus model produced significant differences in the contrasts between modalities at several time bins, which is why we analyzed the data for each modality separately.

[2] In additional models, bigram frequency and orthographic neighborhood density were included as covariates and allowed to interact with lexicality. Neither variable significantly affected ERP amplitudes. Thus, we do not include these factors in the analyses.
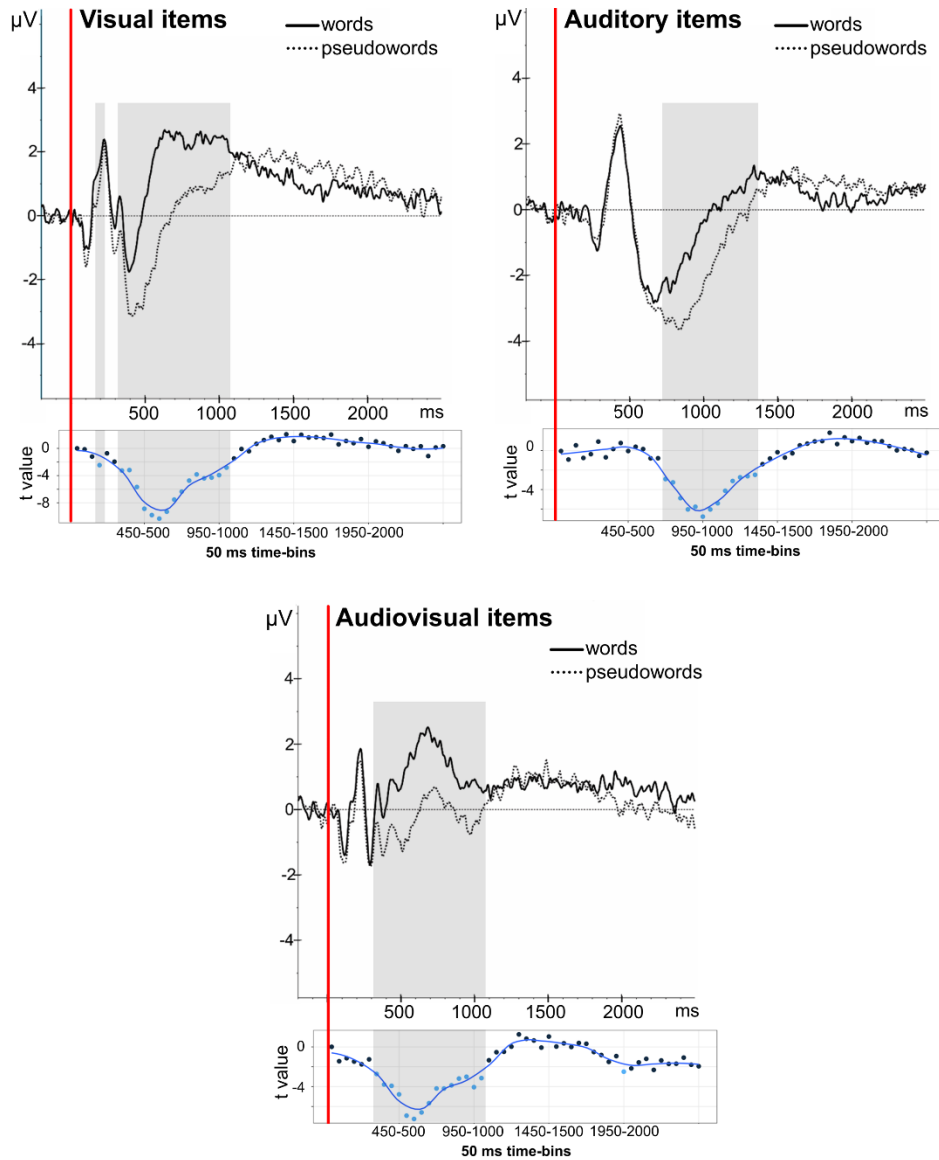
Figure 2. ERPs for words and pseudowords are presented for each modality of presentation. The red lines indicate stimulus onset. The panels below each ERP plot indicate the t-values obtained in the linear mixed effects analyses for each 50 ms bin. Dark blue dots represent non-significant lexicality effects, light blue dots represent significant effects obtained after FDR correction. The gray shaded areas show the time-windows where the lexicality effect was significant.

All electrodes analyses

The results on all electrodes are summarized in Figure 3. Blue tiles represent significant lexicality effects and red tiles represent a reverse lexicality effect (words are more negative than pseudowords). In each modality of presentation, the lexicality effect can be observed at all electrodes (except at F7, P8 and O2 in the auditory items). In the visual and audiovisual modalities, the lexicality effect consistently appears on most electrodes starting 300-350 ms, lasting approximately up to 1000 ms. In the auditory modality, the effect is observed later than in the visual and auditory items and it consistently appears on most electrodes at approximately 800-850 ms, and lasts approximately up to 1100-1150 ms. In all modalities, long lasting lexicality effects that last the entire epoch be observed at some electrodes (e.g., C4 in all modalities, FC6 in visual and audiovisual). The reverse lexicality effect occurs in left lateralized electrodes and is observed in late time windows, approximately after 1300 ms. Note that because "word" responses were made with the right hand, and "pseudoword" responses with the left hand, this reverse lexicality effect may reflect some premotor activation for the responses (the single-trial linear mixed effects analyses across all electrodes indicate that this possibility does not affect our core questions).

Visual Items

Auditory Items

Audiovisual Items
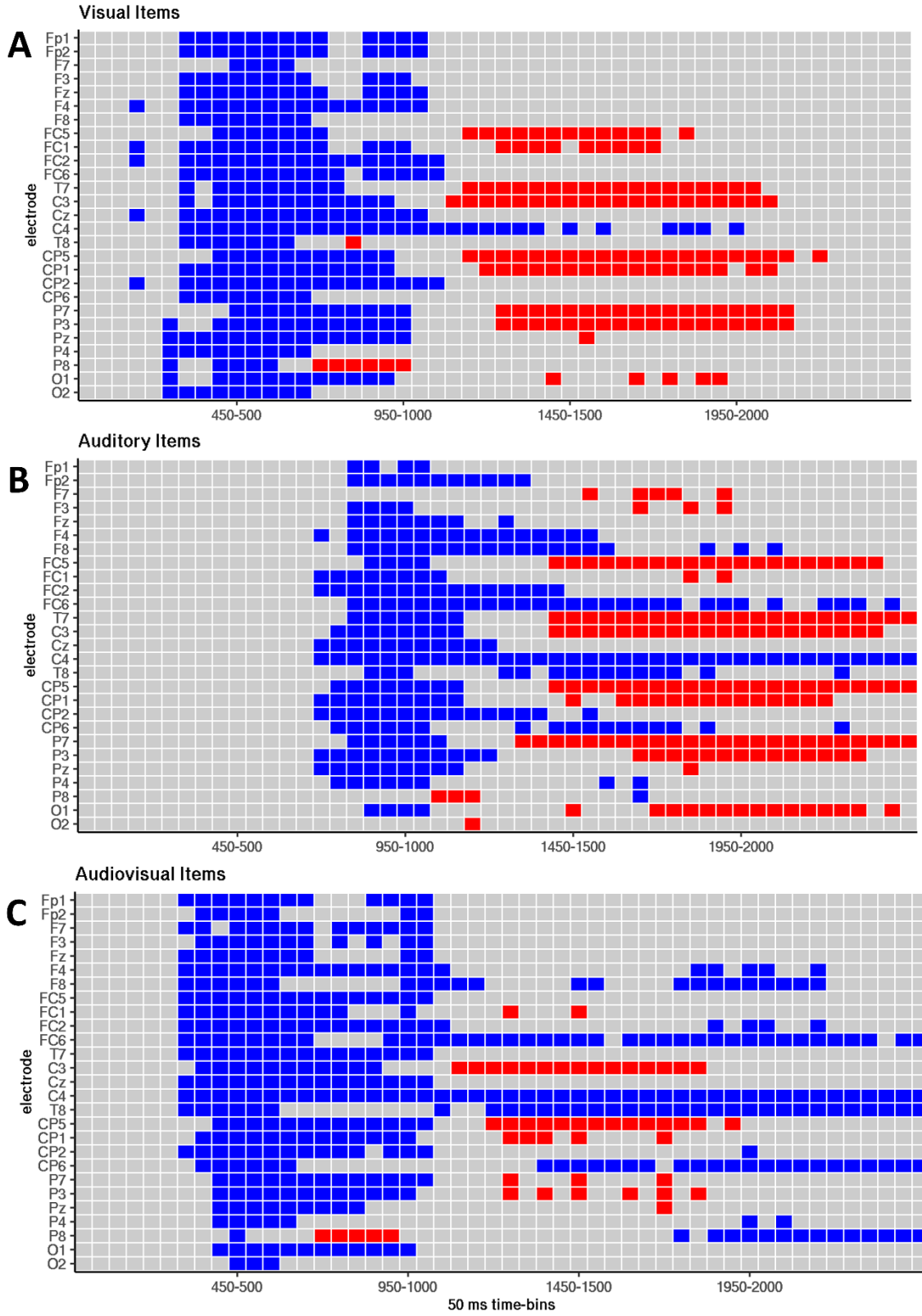
Figure 3. Each tile represents a significant or non-significant p-value from a linear regression model for a particular electrode (y-axis) and time bin (x-axis) in the speeded LDT. Gray tiles are non-significant effects, blue tiles are significant lexicality effects and red tiles are significant reverse lexicality effects obtained after FDR correction. Panel A displays visual items, panel B displays auditory items and panel C displays audiovisual items.

Central cluster analyses

The results are summarized below the ERPs in Figure 2. Each dot in each panel represents a *t*-score that was obtained by comparing the averaged activity in a 50 ms bin for words versus pseudowords. As can be seen, the trajectories of the significant results paralleled those from the grand-average ERPs, suggesting that lexicality effects are responsible for much of the ERP morphology.

For visual items, the difference between words and pseudowords was consistently present from 300 to 1050 ms (we consistently report the lower boundary of the time bin where the effect started, and the upper-boundary of the bin where the effect ended). The time course was quite different for auditory items, with the effect starting at 700 ms and remaining reliable until 1350 ms. The lexicality effect for audiovisual items was similar to that for Visual items, starting around 300 ms and lasting through 1050 ms. Again, this similarity (and its dissimilarity to the pattern for Auditory items) suggests that the AV processing is being primarily driven by the visual information.

ERP amplitudes versus response times

Next, we investigated whether ERP amplitudes during the lexical decision task could predict RT and thus shed light on the dynamics of lexical access and LDR processes[3]. Given that RT analyses revealed response time differences for words and pseudowords as well as for modality, we constructed separate models for each lexical type and each modality. Each

---

[3] We performed similar analyses on accuracy rather than RT, but these results were not significant.

included ERP amplitude as an independent variable and RT as the dependent variable. The

models contained crossed random effects (intercepts) of participants and items, and a fixed effect

of ERP amplitude at the central cluster (averaged amplitudes for each 50 ms bin). The results of

these analyses are summarized in Figure 4. As can be seen, the time-points at which ERP

amplitude could predict RT largely overlapped with the window of the lexicality effect,

indicating that lexical access and LDR processes co-occurred in time, as anticipated.

This was particularly true for visual items. Visual ERP amplitudes predicted RT from 450

ms to 950 ms for words, and from 550 to 1200 ms for pseudowords. The somewhat later onset of

significant pseudoword effects relative to significant word effects is in line with the standard

lexicality effect observed behaviorally, and with the mean response times for words and

pseudowords in the current experiment. As the figure shows, there is a significant relationship

between electrophysiological activity and behavior, starting about 400 ms before the average

response times. For auditory items, the patterns are very similar, but somewhat noisier and

noticeably later (consistent with the slower processing generally seen for A versus V). ERP

amplitudes predicted RT starting at 800 ms and up to 2050 ms for words, and from 1150 ms to

1350 ms for pseudowords. In addition, there is a positive relationship in some of the earlier bins

for the pseudowords, with more positive amplitudes at the 250-300 ms bin, and from 450 to 800

ms predicting slower (i.e., larger) RT for pseudowords. For audiovisual items, amplitudes

predicted reaction time from 250 to 350 ms, and from 450 to 1300 ms for words, while

amplitudes from 550 to 1550 ms predicted reaction time for pseudowords.
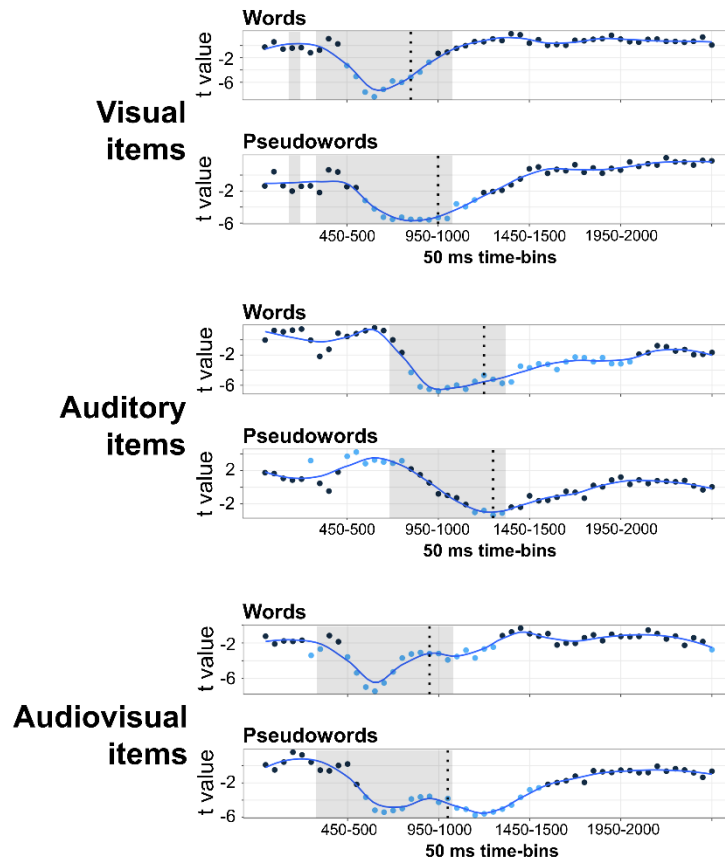
Figure 4. Effect of averaged ERP amplitude in each time bin on RTs for each modality. Dark blue dots represent non-significant effects, and light blue dots represent significant effects obtained after FDR correction. The gray shaded areas correspond to the overall lexicality effect (see Figure 2) and the vertical lines correspond to mean RTs.

Experiment 1, Discussion

In Experiment 1, we presented auditory, visual and audiovisual words and pseudowords and investigated lexical processing using behavioral and electrophysiological measures. We used a speeded version of the LDT, the version typically used in the field. We anticipated that under speeded conditions, lexical access and LDR processes would co-occur in time.

As expected, participants responded faster to visual than auditory items, and the ERP lexicality effect started earlier for visual than auditory items. RTs for audiovisual items were

comparable to those obtained with visual items, as was the onset of the lexicality effect. We anticipated an earlier onset of the lexicality effect and faster RTs for visual than auditory items because the lexical information in printed items is available all at once; for spoken items, lexical access takes more time because the signal unfolds over time. A less predictable finding was that the audiovisual effects mirrored those obtained with print only rather than sound only. From both an evolutionary and a developmental perspective, the primary modality for language perception is audition. Writing systems only appeared a few thousand years ago, and a typically developing child learns to read and write after the auditory language system is in place. Yet, when print and speech are combined, the system follows the information source that is most readily available, not the information that is primary for language perception.

The mean RTs fell in the time-windows where the lexicality effects were observed, about 600 ms after the overall onset of the lexicality effect (averaged across conditions). Thus, on many trials, we can be confident that lexical access and the LDR system were simultaneously engaged. This is corroborated by the relationship between the average ERP amplitudes and RTs.

Perhaps surprisingly, the relationship between ERP amplitude and RT was the same for words and pseudowords, across all modalities of stimulus presentation: more *negative* ERP amplitudes were related to *later* responses. There is however an important distinction to make between effects that fall *within* the window of the lexicality effect, and effects that fall *outside* the window of the lexicality effect. When ERP amplitude predicts RT within the window of the lexicality effect, it is safe to assume that both lexical access and LDR processes occurred simultaneously. However, when RT is predicted by amplitude of the ERP after the lexicality effect, the ERP is most likely not related to lexical access per se, and could instead reflect processes related to LDR.

For visual items, virtually all ERP amplitudes that predicted RT fell within the lexicality effect window (see Figure 4). As noted in the introduction, we assume that LDR processes require at least some prior lexical processing – the lexical decision response must be based on some lexical activation information. It is thus plausible that the later part of what we considered to be a lexicality effect is actually a reflection of LDR processes. In other words, the lexical access processes and LDR processes are overlapping in this time period, and may be combining in ways that boost the magnitude of the apparent lexicality effect. For example, the process of accessing a word representation that has been established through extensive learning may not only be faster, but also less variable than that of accessing a pseudoword representation. The differences in variability in accessing a representation and in reaching a stable representation for the words relative to the nonwords may therefore also lead to differential advantages in reaching a response for each item type in a decision system (Joordens, Piercey, & Azarbehi, 2009). If so, pushing the LDR towards a later point in time – as we will do in Experiment 2 – should *decrease* the length of the (apparent) lexicality effect for visual items by removing the later, LDR-based, section.

For auditory items, a significant portion of the region in which ERP amplitudes predicted RT fell outside the window of the lexicality effect (see Figure 4). For example, for words, the lexicality effect ended at around 1300 ms, but negative ERP amplitudes predicted RTs as late as 1900 ms. In thinking about this pattern, it is useful to remember that in general more positive ERPs are found for words than for pseudowords (i.e., the N400 is bigger for pseudowords). More negative ERP amplitudes for words thus indicate a more 'pseudoword-like' state – a potential error. It stands to reason that such a compromised lexical process does not occur in the window of the lexicality effect, because that effect depends on correct word versus pseudoword

processing. Indeed, the more negative 'pseudoword-like' ERPs for words that were related to slower RTs occurred after the lexicality effect, indicating that sub-optimal lexical processing may have slowed down the response.

The same logic can explain the early positive effect for auditory pseudowords, where more positive ERP amplitudes (i.e., a more 'word-like' signature) predict relatively late RTs. In this case, upon hearing an item, the system presumably started to accumulate lexical evidence, producing relatively positive ERPs. At some point however, it became clear that the item was not a word. The initially gathered lexical evidence needs to be revised, slowing down the correct pseudoword response. Overall, the temporal overlap between the ERP-RT relationship and the lexicality effect is weaker for auditory items than for visual items. Since we argued that the apparent lexicality effect for visual items should become shorter if LDR processes are pushed away to a later point in time, we expect this temporal reduction to be smaller for auditory than visual items because of the weaker relationship for the auditory case.

For the audiovisual items, we observed that ERP amplitudes within the window of the lexicality effect predicted RT (again indicating that the lexical access and decision processes overlap), but we also found a significant portion of RTs predicted by the ERP amplitudes outside the window of the lexicality effect. However, as indicated in Figure 4, these effects aligned with the lexicality effect for auditory items, and are perhaps related to (lexical) processing of the auditory component of the stimulus. Thus, while lexicality effects seem to be mainly driven by the written input, the ERP amplitudes that predict RT seem to be driven by both modalities. These results suggest that these two effects reflect different neurocomputational processes, one which focuses on (attempted) lexical access, and another which focuses on evaluating the degree to which the lexical access process succeeded (thus supporting a word response) or failed (thus

supporting a pseudoword response). Importantly, because the AV lexicality effect only partially

overlapped with the ERP amplitudes that predicted RT, we can again predict that pushing the

LDR to a later point in time will not shorten the AV lexicality effect to the same degree as for

visual items.

To summarize, we observed that audiovisual items were processed very similarly to

visual items in terms of behavioral and ERP patterns. Despite the fact that spoken language

precedes written language in time (both evolutionarily and developmentally), a letter string

presented all at once provides an advantage when recognizing words. Apparently, the rapid

availability of written stimuli dominates the word recognition process when presented in parallel

with its spoken counterpart. We observed a general negative relationship between ERP

amplitude and RT, which aligns with a variety of decision making theories that model the

decision process as the accumulation of differentially more evidence for one response relative to

another response (e.g., Armstrong, Joordens, & Plaut, 2009; Ratcliff, Gomez, & McKoon, 2004;

Ratcliff, Van Zandt, & McKoon, 1999). To determine how the LDR process affects the apparent

ERP lexicality effect – more specifically, its duration – we separated the two processes in

Experiment 2 by using a non-speeded LDT in an otherwise identical experiment.

Experiment 2

Methods

Participants

Twenty-four new participants (7 males; mean age: 23 years; age range 20-29) were recruited for this experiment. As in Experiment 1, all were right-handed native speakers of Spanish, had (corrected to) normal vision, with self-reported normal hearing and no neurological problems.

Task and stimulus materials

The stimulus materials, general procedures and task were identical to Experiment 1, except that responses were delayed. Specifically, a question mark appeared 1750 ms after stimulus onset, which signaled the participant to make a response. A warning message was displayed if participants pressed an invalid key or if they responded before the question mark. The experiment took approximately 1 hour and 15 minutes. Reaction times were recorded from the onset of the question mark.

The protocols and procedures related to EEG recording, processing, and analyses were the same as in Experiment 1.

Results

Three participants were removed from the analyses, one because he responded before the response probe on more than half of the trials, and two because the ERPs contained a large proportion of artifacts (> 30%).

Behavioral results

The models constructed for RT and accuracy analyses were the same as those in Experiment 1. Figure 5 displays the lexicality effect in each modality for RT and accuracy. Tables 4 and 5 present the statistics from the regression models. Again, we report both raw RT and inverse RT analyses because the raw RT residuals violated the assumption of normality whereas the inverse RT data did not.
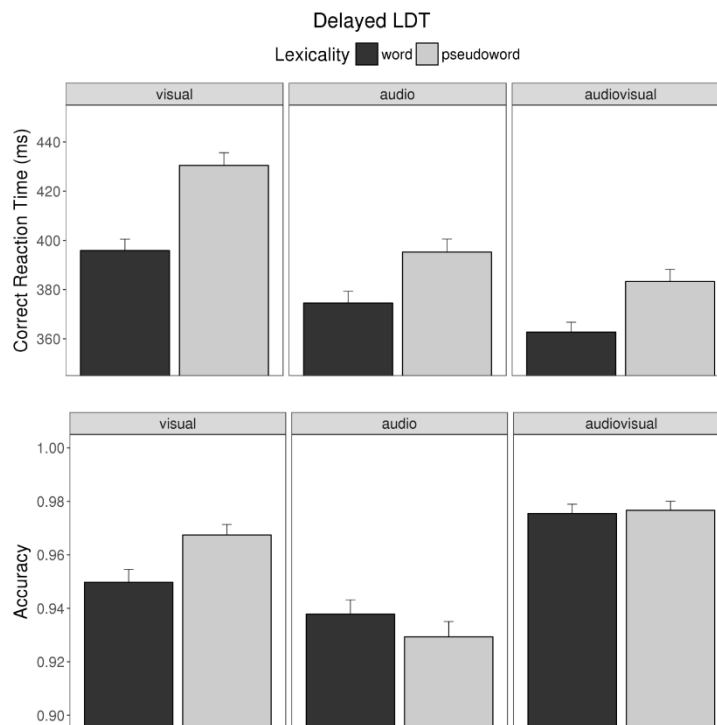
Figure 5. Mean reaction time (top) and accuracy (bottom) for words and pseudowords. Error bars represent standard errors. The RT y-axis represents the mean RT computed from onset of the response probe.

Table 4. Summary of LMER statistics for RT.

Model 1 (words and visual modality as baseline)	Raw RT			Inverse RT		
Fixed effects	b	SE	t	b	SE	t
Lexicality (pseudoword)	33	6	5.22**	-0.0001	0.00003	-5.24**
Modality (auditory)	-19	6	-3.00*	-0.0002	0.00003	6.77**
Modality (audiovisual)	-32	6	-5.12**	0.0002	0.0003	6.88**
Lexicality X Modality (auditory)	-12	9	-1.34	0.00003	0.00004	0.88
Lexicality X Modality (audiovisual)	-10	9	-1.15	0.00005	0.00004	1.36
Model 2 (words and auditory as baseline)						
Modality (audiovisual)	-13	6	-2.10*	0.000001	0.00003	0.05
Lexicality X Modality (audiovisual)	-2	9	-0.20	0.00002	0.00004	0.46

*p<.05, **p<.001

For visual items, the lexicality effect was 34 ms, versus 22 and 20 ms for auditory and audiovisual items. Visual words were responded to faster than pseudowords. Unlike Experiment

1, RTs for words were slower for V than for A, for V than AV, and for AV than A (raw RT

model only). Interactions were not significant.

Table 5. Summary of LMER statistics for accuracy.

Model 1 (words and visual modality as baseline)

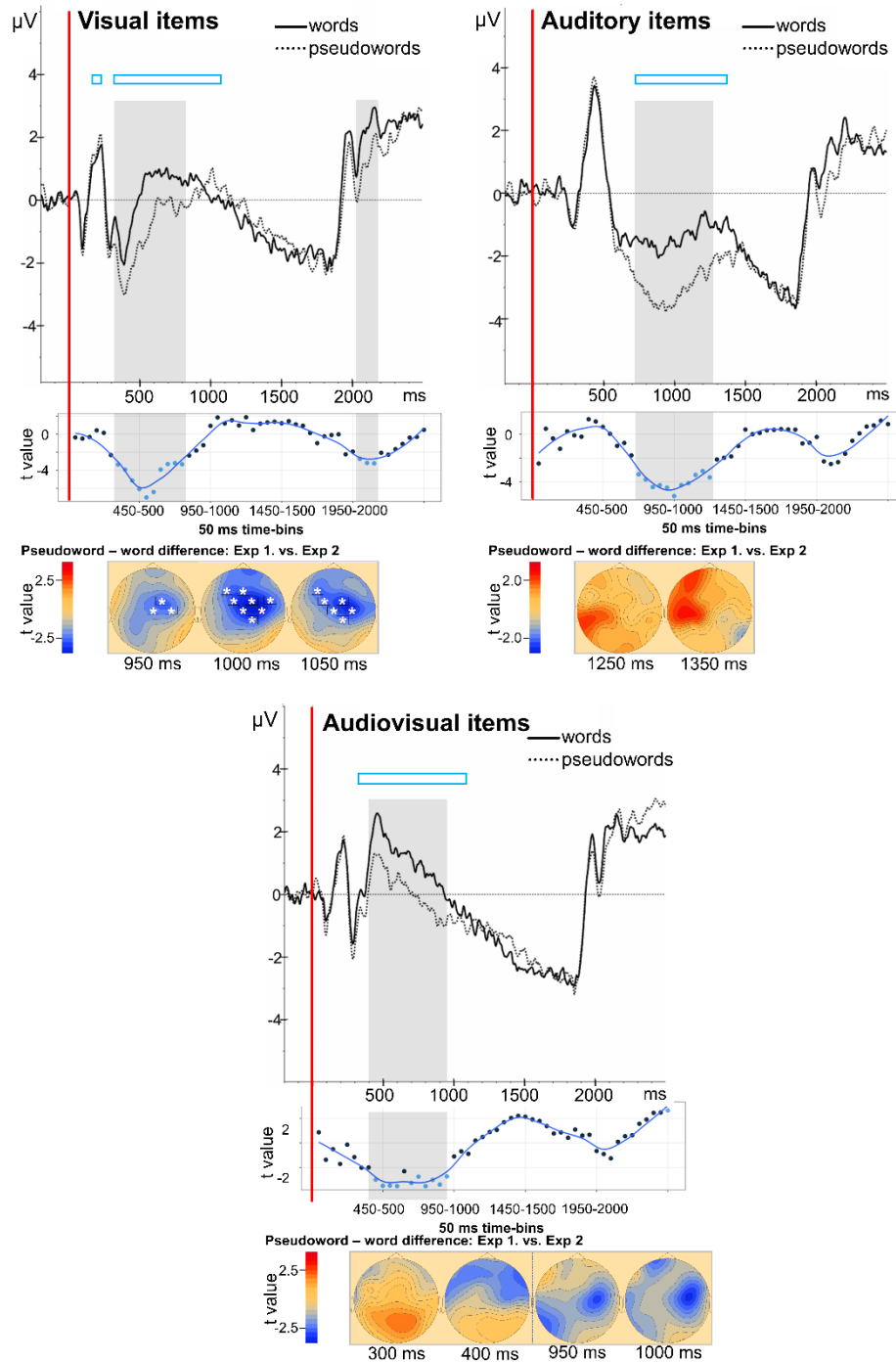| Fixed effects	b	SE	z
Lexicality (pseudoword)	0.48	0.17	2.88*
Modality (auditory)	-024	0.13	-1.86
Modality (audiovisual)	0.76	0.17	4.51*
Lexicality X Modality (auditory)	-0.60	0.20	-2.98*
Lexicality X Modality (audiovisual)	-0.41	0.25	-1.62
Model 2 (words and auditory as baseline)			
Modality (audiovisual)	1.01	0.17	6.12*
Lexicality X Modality (audiovisual)	0.19	0.23	0.80

*p<.01

Accuracy in the V modality was higher for words than for pseudowords. In addition, accuracy for words was (marginally) higher for V than for A, but lower for V than for AV. Finally, the lexicality effect was larger for V than for A but V and AV and A and AV did not differ from each other.

ERP results

Figure 6 shows the ERPs for words and pseudowords, for the three modalities of stimulus presentation. Again, there are clear differences in the onset and duration of the lexicality effect across modalities. In addition, there was a clear contingent negative variation effect (CNV) (Kononowicz & Penney, 2016) signaling anticipation of the response probe. For visual and auditory items, the CNV was followed by an effect in the same direction as the lexicality effect.

Figure 6. ERPs for words and pseudowords are presented for each modality of presentation. The red lines indicate stimulus onset. The panels directly below each ERP plot indicate the t-values obtained in the linear mixed effects analyses on the 50 ms bins. Dark blue dots represent non-significant lexicality effects, and light blue dots represent significant effects obtained after FDR correction. The gray shaded areas show the time-windows where the lexicality effect was significant. As context, the blue boxes indicate the time-windows in which the lexicality effect was significant in Experiment 1. The topography plots below each panel show the (partial) outcome of between-subject cluster-based permutation tests that tested the pseudoword – word difference waves in Experiment 1 against Experiment 2, for those time-windows where the lexicality effect appeared to be shorter in Experiment 2 than in 1. Asterisks represent electrodes that are part of significant clusters ($p < .05$).

All electrode analyses
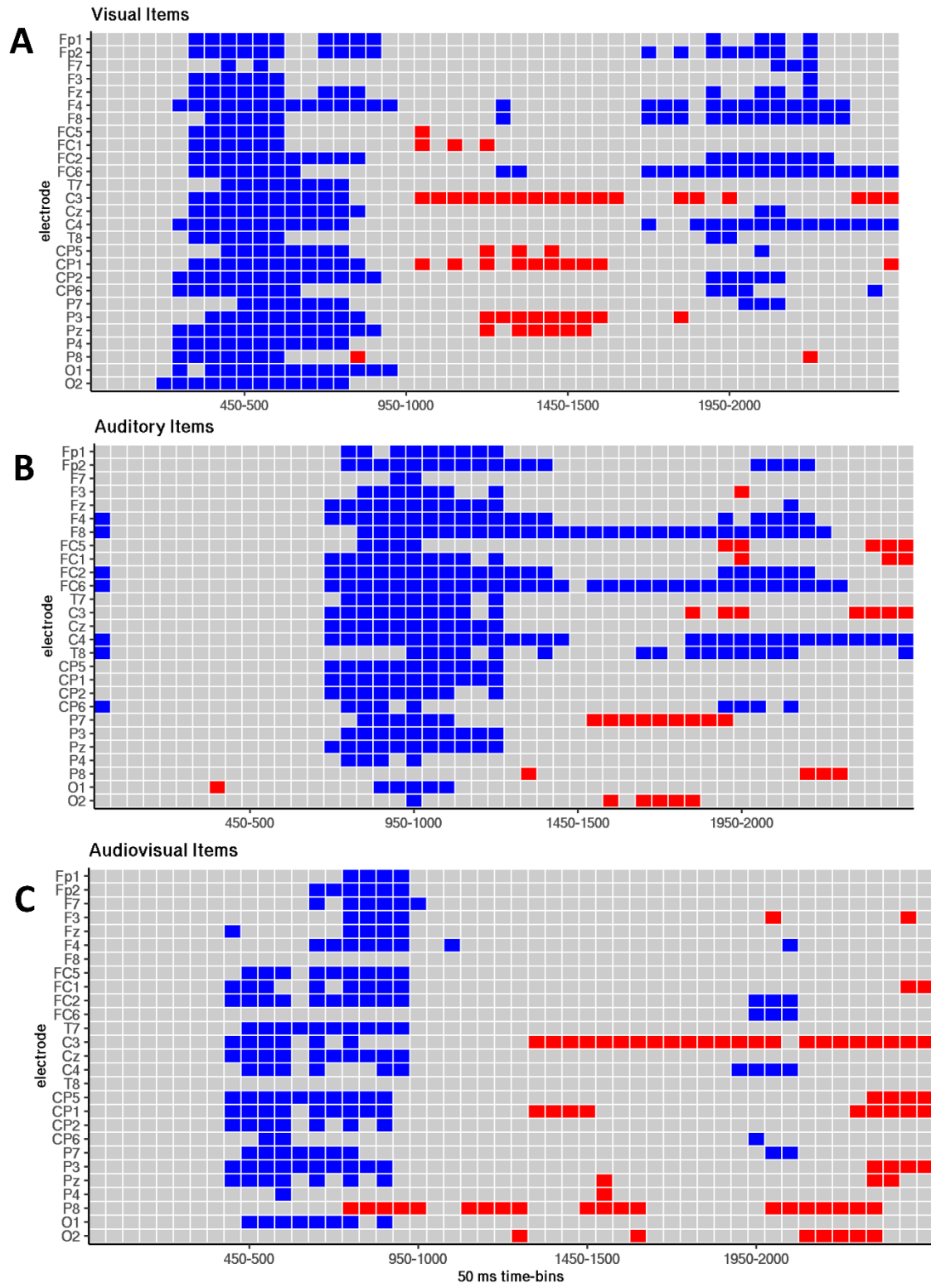
Figure 7. Each tile represents a significant or non-significant p-value from a linear regression model for a particular electrode (y-axis) and time bin (x-axis) in the delayed LDT. Gray tiles are non-significant effects, blue tiles are significant lexicality effects and red tiles are significant reverse lexicality effects obtained after FDR correction. Panel A displays visual items, panel B displays auditory items and panel C displays audiovisual items.

The ERP data were analyzed using the same approach as in Experiment 1. The overall lexicality effect patterns (see Figures 6 and 7) are similar to those found in the speeded task:

A lexicality effect widespread to all electrodes can be observed in each modality. In the V modality, the lexicality effect started at 300 ms and lasted up to approximately 800 ms on most electrodes. In the A modality it started later, at 700 ms, and lasted up to 1250 ms in some electrodes. In the AV modality, the effect was consistent between approximately 400 ms and 850 ms. A left lateralized reverse lexicality effect can be also observed but of shorter duration and in fewer electrodes (e.g., C3 in V and AV) than in the speeded task.

As explained in the Discussion of Experiment 1, we predicted that the lexicality effects would be shorter in Experiment 2 than in Experiment 1, as the delayed task would push LDR processes away from lexical access. In addition, we argued that this reduction should be largest for V, as for visual items, virtually all ERP amplitudes that predicted RT (in Experiment 1) fell within the lexicality effect window (see Figure 4). As can be seen in Figure 6, this qualitative effect was indeed observed. To be more precise, the central cluster window in which the lexicality effect was significant had shortened by 250 ms for V (300-1050 ms in Experiment 1, vs. 300-800 ms in Experiment 2), by 100 ms for A (700-1350 ms in Experiment 1, vs. 700-1250 ms in Experiment 2) and by 200 ms for AV (300-1050 ms in Experiment 1, vs. 400-950 ms in Experiment 2). To quantify these effects, we conducted 2-tailed between-subject cluster-based permutation tests in which we compared the lexicality effects (i.e., the pseudoword – word

difference wave, averaged per condition for each participant, retaining the original sample

frequency) in the time windows where lexicality effects were observed in Experiment 1, but not

in Experiment 2 (i.e., 800-1050 ms for V, 1250-1350 ms for A, and 300-400 and 950-1050 ms

for AV) [4]. The cluster-based analyses were conducted in BESA statistics 2.0, and were based on

1000 permutations per time-window. All channels were included, the cluster-level alpha was set

at .05, and neighbor distance was set at 5 cm, which corresponds to an average neighborhood

size of 2.67 channels for each electrode. As can be seen in the topography of the t-distributions

included in Figure 6, the duration of the lexicality effect indeed significantly decreased in

Experiment 2 as compared to Experiment 1. No significant results were obtained for A and AV,

although the topographical distribution of the later effect in the AV condition closely resembled

the topography observed in V.

For all items, an effect that resembled the lexicality effect (more negative ERPs for

pseudowords than for words) appeared when the response signal was presented, with this effect

reaching significance at several electrodes in the V and A modalities (e.g, Fp2, FC2, C4) and in

fewer electrodes and shorter duration in the AV modality (e.g, FC2, FC6, C4). We will return to

this re-instantiation of the lexicality effect in the Discussion.

---

[4] As indicated in Figures 4 and 6, the *size* of the lexicality effect – defined by the difference between pseudoword and word ERPs – was also smaller in Experiment 2 than in Experiment 1. This modulates the outcome of cluster-based permutation analyses, and we therefore restricted the between-subject permutation analyses to the small time-windows where effects were observed in Experiment 1, but not in Experiment 2.

Lexical activation and response times

As in Experiment 1, we investigated whether ERP amplitudes could predict RT and thus shed light on the dynamics between lexical access and decision processes. The models we constructed were identical to the ones we used for the speeded task, with the results summarized in Figure 8. As can be seen, the general pattern was the same as in Experiment 1, in the sense that *negative* amplitudes predicted *slower* RTs. However, all but one of these significant correlation effects occurred *after* the lexicality effect. This indicates that the correlations between ERP amplitude and RT were unlikely to be related to lexical access per se, but instead, to LDR processes.
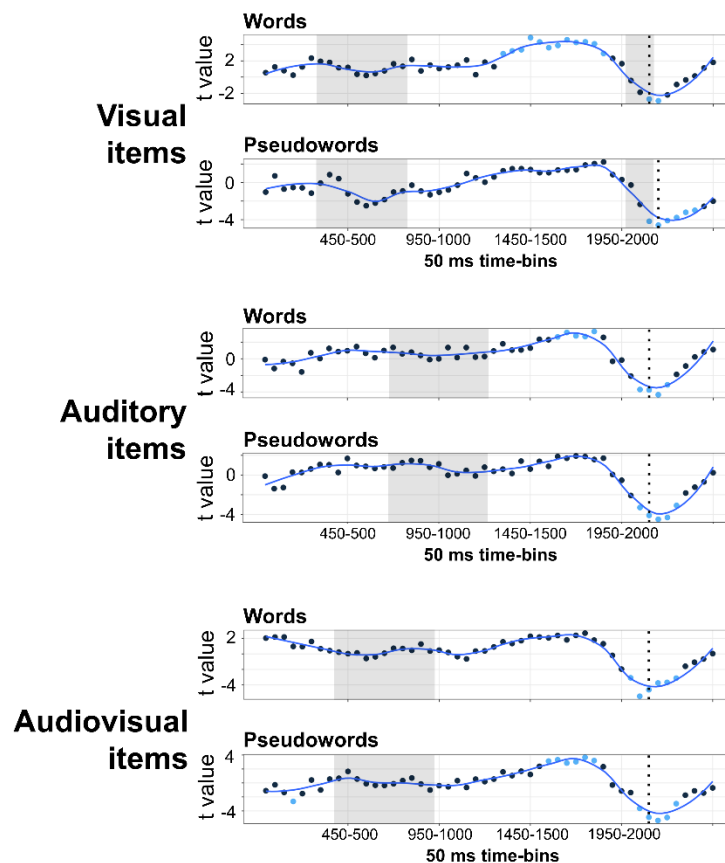
Figure 8. Effect of averaged ERP amplitude in each time bin on RTs for each modality. Dark blue dots represent non-significant effects, and light blue dots represent significant effects obtained after FDR correction. The gray shaded areas correspond to the overall lexicality effect (see Figure 5) and the vertical lines correspond to mean RTs.

For visual items, the negative relationship between ERP amplitudes after the response probe and RT (recall that the probe appeared 1750 ms after stimulus onset) was similar for words and pseudowords. However, pre-response probe, ERP amplitudes in the CNV time window positively predicted RT for words as well. This was also the case for auditory items, but again, for both words and pseudowords, negative ERP amplitudes after the response probe predicted slower RT. For audiovisual pseudowords there was a positive relation between amplitude and RT in the CNV time-window, but for both words and pseudowords, the relationship between ERP amplitudes after the response probe and RTs was negative.

Experiment 2, Discussion

In Experiment 2, we used a delayed version of the lexical decision task, in which participants were asked to withhold their responses until a response probe appeared. Based on Experiment 1, we generated the prediction that the observed ERP lexicality effect should become shorter in Experiment 2, not because lexical access is shorter in a delayed task, but because the temporally overlapping LDR processes in Experiment 1 had extended the apparent effect. The overall lexicality effect was similar to that observed in Experiment 1, featuring the typical more negative ERP amplitudes for pseudowords than for words. Critically, in all modalities, the lexicality effect was indeed shorter than in Experiment 1, but only significantly shorter for visual-only trials, in line with the predictions laid out in the Discussion of experiment 1.

For auditory and visual items, the onset of the lexicality effect was strikingly similar to Experiment 1, but for audiovisual items, the lexicality effect began 100 ms later than before. In Experiment 1 we suggested that the vision-dominated response in the AV condition is related to the fact that printed items are presented at once, whereas spoken items need to unfold over time. However, this is also true for the audiovisual stimuli in Experiment 2, but the onset of the lexicality effect was later for audiovisual than visual items. One possible explanation could be that under time pressure (in Experiment 1), the visual signal was prioritized because it provides the quickest route towards the lexicon. When response speed is no longer an issue (in Experiment 2), the spoken item was weighed more heavily in the process because it is presented in the primary language modality. This would account for the shift of the onset of the lexicality effect towards the auditory case (i.e., a later point in time), but since the shift was not significant, this interpretation is rather tentative at this point.

Another interesting finding is the re-instantiation of the lexicality effect (more negative amplitudes for pseudowords than words) after the response probe. This re-instantiation effect was observed in the visual and auditory modalities and to a lesser extent in the audiovisual modality. Possibly, the re-instantiation effect reflects a re-activation of the item in memory, which may be less needed when the signal is bi-modal rather than uni-modal. This explanation, however, is quite speculative, and requires further testing.

With respect to the relationship between ERP amplitudes and RT, we again found a systematic negative relationship, in all three modalities, near the time of the response probe. This pattern suggests that the correlations are related to LDR processes rather than to lexical access per se.

Finally, we found that ERP amplitudes in the CNV time window were positively associated with RT for words in the visual and auditory modalities and for pseudowords in the audiovisual modality. That is, the larger the CNV, the slower the RT. The CNV component has been associated with the anticipation of and preparation for an event. One interpretation is that the CNV is a memory representation of a time interval, with its peak reflecting the duration of a standard interval in working memory. More recent evidence suggests that both memory and decision processes contribute to the CNV (Kononowicz & Penney, 2016 for a review). Thus, perhaps it is not surprising to find that it predicted speed of responding if the CNV reflects both the duration of the time interval and the maintenance of the stimulus in memory. Such findings of integrated representations of the "evidence" supporting a response as well as the planning for that response have been found in single cell recordings in simple visual discrimination tasks in macaques (Kim & Shadlen, 1999). The CNV observed here might be attributable to a partially integrated representation of the memory for whether a word or a pseudoword stimulus was presented, combined with the representation of the response associated with that stimulus.

The fluctuating positive correlation between ERP amplitudes and RT may reflect changes in how different amounts of neural activity align with word and nonword responses as the representation of a stimulus is accessed over time. For example, Laszlo & Plaut (2012; Figure 7) observed that early in processing, there was (transiently) greater semantic activity for stimuli that had large orthographic neighborhood sizes, with only a small effect of lexicality per se. Thus, words and pseudowords generated similar high levels of semantic activity, whereas acronyms and illegal strings generated similar low levels of semantic activity. Early transient N400 activity may therefore not be a reliable index of lexicality per se, but rather of a variable that is correlated with lexicality. In contrast, as the model settled into stable state, words maintained higher

overall levels of activity than nonwords (although there were some residual effects of

neighborhood size). Thus, higher activity later on in processing is a more reliable indicator that

the stimulus was a word. Related to this, Joordens et al. (2009) have outlined how in a

connectionist model of word recognition, word representations should settle into a stable

representational state more quickly than nonwords. This is because as individual representational

features come to be in a correct activation state, they provide an increasingly consistent signal to

the other features that are not yet in a correct state, speeding the activation of a full, stable

representation. The same is not true for nonwords, however, which typically partially activate a

large set of features that are all partially consistent with one another and with the written (or

spoken) form. In the delayed response situation of Experiment 2, the response system may have

learned to rely on this later, less transient and more stable representation. Although potentially

generated by the same underlying system, this later activity may have substantially different

sensitivity and informativeness for making lexical decisions (either directly based on lexicality,

or indirectly based on related covariates) than the representations activated early in processing.

General Discussion

We conducted two experiments with the central goal of separating lexical access

processes from processes related to lexical decision and/or generating a response (which we

labeled LDR processes). In Experiment 1, we used the standard lexical decision task in which

participants must respond to the stimuli as quickly and as accurately as possible whereas in

Experiment 2 we used a delayed task in which participants only responded after a response probe

appeared. In both experiments, we used the same large set of words and pseudowords presented

in counterbalanced blocks of different modalities: visual, auditory and audiovisual. We

employed RT and ERP measures at the single-trial level, analyzed our data with linear mixed effect models at the group level, and were indeed able to separate lexical access from LDR processes.

In Experiment 1, ERP amplitudes *within* the window of the lexicality effect significantly predicted response times at the single trial level. This finding led us to conclude that lexical access and LDR processes (quantified by the ERP $\times$ RT effects) overlapped in time. This inference was corroborated by two complementary results in Experiment 2. First, the ERP $\times$ RT correlations shifted to a later time-window, around the time of the response probe. Second, the time window for the ERP lexicality effect was shorter than in Experiment 1 (but only significantly so for visual items), consistent with the view that some of this separation of word and pseudoword amplitudes is driven by LDR processes.

In both experiments the lexicality effect was delayed in the auditory relative to the visual modality. Although this is exactly as predicted, it is in the opposite direction to what Holcomb and colleagues (1990, 1992) reported; they found earlier effects in the auditory than the visual modality. However, one crucial difference is that they investigated N400 *context effects* at the word and sentence level. More specifically, they focused on semantic priming processes rather than lexical access. In their first study, participants performed a primed lexical decision task where they were presented with a prime word followed by a target word or pseudoword (Holcomb & Neville, 1990). Thus, they looked at the relationship between the prime word and the target stimuli. In their second study, they compared the N400 effect between sentences ending with either a highly expected word or a semantically inappropriate word (Holcomb, Coffey, & Neville, 1992). Thus, rather than investigating the effect of recognizing isolated words and pseudowords, they investigated contextual effects. To the best of our knowledge, the

current study is the first to systematically investigate the word/pseudoword *lexicality effect* across sensory modalities. Another difference is that we used a language with transparent orthography, whereas Holcomb's studies were conducted in English, a language with a much less transparent mapping of print to sound.

For auditory and visual stimuli, the onset time of the lexicality effect was the same in both experiments, but this was not the case for audiovisual stimuli. In Experiment 2, the audiovisual lexicality effect shifted towards a later point in time. We can only offer a tentative explanation at this point, but it is nevertheless an intriguing one. When confronted with an audiovisual stimulus, the visual input provides the quickest route to the lexicon. In a speeded task, speed obviously matters, and the neural and behavioral response might therefore be driven by lexical evidence generated by the visual input. However, when there is no time pressure, processing shifts towards the auditory input – changing both the behavioral and the ERP patterns – because the auditory system is the biological default for processing linguistic input.

We also observed a re-instantiation of the lexicality effect in the visual and auditory modalities in Experiment 2. This effect occurred in time windows around the response probe, at a time that is clearly much later than lexical access. Post-N400 ERP components such as the Late Positive Complex and the P600 are thought to index memory retrieval (Danker et al., 2008; Mecklinger, 2000) and syntactic revision (Kutas, Federmeier, Staab, & Kluender, 2007 for a review), respectively. Perhaps the re-instantiation of the lexicality observed in our delayed LDT reflects the semantic memory system re-accessing information in order to make an accurate response at the required time. Why such re-instantiation was observed in fewer electrodes in the audiovisual modality is not clear, but one possibility is that when written and spoken stimuli are simultaneously presented, the semantic memory system is activated more strongly than when

stimuli are presented unimodally. If this produces a stronger "imprint", later reactivation may be less needed at the time of responding.

A widespread assumption in the language literature (e.g., Coltheart et al., 2001; Plaut et al., 1996) and in the decision making literature (e.g., Ratcliff et al., 2004; Usher & McClelland, 2001), is that the lexical system (the source for evidence regarding a response in lexical decision) and the decision system (which generates a response based on the available evidence) are essentially separate systems. From this perspective, the lexical system can generate simple monotonic predictions regarding the likelihood of a word response or speed of a response for one stimulus (e.g., a high frequency word vs. a low frequency word). The decision system is responsible for the variability in the distribution of these responses, reflecting response biases, the shape of the latency distributions, and phenomena of this sort. Indeed, it is this independence assumption that has allowed researchers studying the lexical and decisions systems to operate largely independently of one another.

This assumption, however, is inconsistent with our observation of a re-activation of the lexicality effect prior to responding. If the two systems were, in fact, independent, the decision system should have had more than enough time to accumulate and select the intended response during the delay period, and simply initiate that response when the delay period ended. In that case, however, there would be no need to re-activate the stimulus and there would be no reason to predict a processing advantage for words over nonwords, as is typically observed in speeded lexical decision tasks (but not other tasks involving the same words, e.g., Hino, Pexman, & Lupker, 2006; for discussion of the source of the lexicality advantage, see; Joordens et al., 2009). Our data, however, indicate that such a re-activation occurs, and moreover, that this re-activation triggers a lexicality effect that is qualitatively consistent with the lexicality effect observed in the

speeded version of the task. This suggests that the observed pattern is not simply an epiphenomenal re-activation of the representation of the initial stimulus when executing a delayed response, but instead is an important contributor to the delayed response (and, correspondingly, to the lexicality effect in the speeded task). Similar effects have been observed in fMRI correlates and TMS manipulations of performance in delayed recognition tasks for words, faces, and direction of motion (Rose et al., 2016). The re-instantiation found here adds to a growing body of evidence supporting the integration of decision processes with perceptual processes.

In the current study, we leveraged the methodological power available by combining behavioral measures (reaction time) with neural measures (ERPs). The combination allowed us to assess whether the electrophysiological data was directly related to when a participant would respond to a particular stimulus. The resulting patterns offer new insights into the relationship between lexical access and processes that are specifically related to performing a lexical decision task. Our findings pave the way towards a better integration of theories of the lexical system with those of the decision system, enabling a more comprehensive understanding of how stimuli that engage the language system produce particular effects in particular tasks.

Conflict of interest

None.

Acknowledgments

References

Armstrong, B. C., Joordens, S., & Plaut, D. C. (2009). Yoked criteria shifts in decision system adaptation: Computational and behavioral investigations. In N. A. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 2130-2135). Austin, Texas: Cognitive Science Society.

Armstrong, B. C., & Plaut, D. C. (2011). Inducing homonymy effects via stimulus quality and (not) nonword difficulty: Implications for models of semantic ambiguity and word recognition. . In L. Carlson, C. Holscher & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 2223-2228). Austin, TX.: Cognitive Science Society.

Armstrong, B. C., & Plaut, D. C. (2016). Disparate semantic ambiguity effects from semantic processing dynamics rather than qualitative task differences. *Language, Cognition, & Neuroscience, 31*(7), 940-966. doi: 10.1080/23273798.2016.1171366

Baart, M., Armstrong, B. C., Martin, C. D., Frost, R., & Carreiras, M. (2017). Cross-modal noise compensation in audiovisual words. *Scientific Reports, 7*, 42055. doi: 10.1038/srep42055

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

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390-412.

Bates, D. M., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1-48. doi: 10.18637/jss.v067.i01

Benjamini, Y. , & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological), 57*(1), 289-300.

Bentin, S., Mouchetant-Rostaing, Y., Giard, M. H., Echallier, J. F., & Pernier, J. (1999). ERP manifestations of processing printed words at different psycholinguistic levels: time course and scalp distribution. *Journal of Cognitive Neuroscience, 11*(3), 235-260.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: a dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*(1), 204-256.

Danker, J. F., Hwang, G. M., Gauthier, L., Geller, A., Kahana, M. J., & Sekuler, R. (2008). Characterizing the ERP Old-New effect in a short-term memory task. *Psychophysiology, 45*(5), 784-793. doi: 10.1111/j.1469-8986.2008.00672.x

Deacon, D., Hewitt, S., Yang, C., & Nagata, M. (2000). Event-related potential indices of semantic priming using masked and unmasked words: evidence that the N400 does not reflect a post-lexical process. *Brain Research Cognitive Brain Research, 9*(2), 137-146.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134*(1), 9-21. doi: 10.1016/j.jneumeth.2003.10.009

Duchon, A., Perea, M., Sebastian-Galles, N., Marti, A., & Carreiras, M. (2013). EsPal: one-stop shopping for Spanish word properties. *Behavior Research Methods, 45*(4), 1246-1258. doi: 10.3758/s13428-013-0326-1

Hino, Y., Pexman, P. M. , & Lupker, S. J. (2006). Ambiguity and relatedness effects in semantic tasks: Are they due to semantic coding? . *Journal of Memory and Language* (55), 247-473.

Holcomb, P. J. (1993). Semantic priming and stimulus degradation: implications for the role of the N400 in language processing. *Psychophysiology, 30*(1), 47-61.

Holcomb, P. J., Coffey, S. A., & Neville, H. J. . (1992). Visual and auditory sentence processing: A developmental analysis using event-related brain potentials. *Developmental Neuropsychology, 8*(2 & 3), 203-241.

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. doi: 10.1162/089892902760191153

Holcomb, P. J., & Neville, H. J. (1990). Auditory and visual semantic priming in lexical decision: A comparison using event-related brain potentials. *Language and Cognitive Processes, 5*, 281-312.

Joordens, S. , Piercey, C. D., & Azarbehi, R. (2009). Modeling performance at the trial level within a diffusion framework: a simple yet powerful method for increasing efficiency via error detection and correction. *Canadian Journal of Experimental Psychology, 63*(2), 81-93.

Jung, T. P., Makeig, S., Humphries, C., Lee, T. W., McKeown, M. J., Iragui, V., & Sejnowski, T. J. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology, 37*(2), 163-178.

Keuleers, E., & Brysbaert, M. (2010). Wuggy: a multilingual pseudoword generator. *Behavior Research Methods, 42*(3), 627-633. doi: 10.3758/BRM.42.3.627

Kim, J. N., & Shadlen, M. N. (1999). Neural correlates of a decision in the dorsolateral prefrontal cortex of the macaque. *Nature Neuroscience, 2*(2), 176-185. doi: 10.1038/5739

Kononowicz, T. W., & Penney, T. B. (2016). The contingent negative variation (CNV): timing isn't everything. *Current Opinion in Behavioral Sciences, 8*.

Kounios, J., & Holcomb, P. J. (1994). Concreteness effects in semantic processing: ERP evidence supporting dual-coding theory. *Journal of Experimental Psychology. Learning, Memory and Cognition, 20*(4), 804-823.

Kutas, M., Federmeier, K. D., Staab, J., & Kluender, R. (2007). Language. In J. T. Cacioppo, L. G. Tassinary & G. G. Berntson (Eds.), *Handbook of psychophysiology* (pp. 555-580). New York: Cambridge University Press.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science, 207*(4427), 203-205.

Kutas, M., & Hillyard, S. A. (1983). Event-related brain potentials to grammatical errors and semantic anomalies. *Memory and Cognition, 11*(5), 539-550.

Kutas, M., & Hillyard, S. A. (1984). Event-related brain potentials (ERPs) elicited by novel stimuli during sentence processing. *Ann N Y Acad Sci, 425*, 236-241.

Kutas, M., & Van Petten, C. (1994). Psycholinguistics electrified: event-related brain potential investigations. In M. A. Gernsbacher (Ed.), *Handbook of Psycholinguistics* (pp. 83-143). San Diego, CA: Academic

Kuznetsova, A., Brockhoff, P. B. , & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software, 82*(13), 1-26. doi: 10.18637/jss.v082.i13

Laszlo, S., & Federmeier, K. D. (2007). Better the DVL you know: acronyms reveal the contribution of familiarity to single-word reading. *Psychological Science, 18*(2), 122-126. doi: 10.1111/j.1467-9280.2007.01859.x

Laszlo, S., & Federmeier, K. D. (2012). The N400 as a snapshot of interactive processing: Evidence from regression analyses of orthographic neighbor and lexical associate effects. *Psychophysiology, 48*, 176-186. doi: 10.1111/j.1469-8986.2010.01058.x

Laszlo, S., & Federmeier, K. D. (2014). Never Seem to Find the Time: Evaluating the Physiological Time Course of Visual Word Recognition with Regression Analysis of Single Item ERPs. *Language and Cognitive Processes, 29*(5), 642-661. doi: 10.1080/01690965.2013.866259

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: an open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience, 8*, 213. doi: 10.3389/fnhum.2014.00213

López Zunini, R. A., Renoult, L., & Taler, V. (2017). N400 Effects of Semantic Richness can be Modulated by Task Demands. *Neuropsychology, 31*(3), 277-291.

MacGregor, L. J., Pulvermüller, F. , van Casteren, M. , & Shtyrov, Y. (2012). Ultra-rapid access to words in the brain. *Nature Communications, 3*, 711.

Mazzoni, D. (2013). Audacity (Version 2.0.5) [Computer software]. Available from http://audacity.sourceforge.net/download/.

Mecklinger, A. (2000). Interfacing mind and brain: a neurocognitive model of recognition memory. *Psychophysiology, 37*(5), 565-582.

Muller, O., Duñabeitia, J. A., & Carreiras, M. (2010). Orthographic and associative neighborhood density effects: what is shared, what is different? *Psychophysiology, 47*(3), 455-466. doi: 10.1111/j.1469-8986.2009.00960.x

Pierce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods, 162*(1-2), 8-13.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: computational principles in quasi-regular domains. *Psychological Review, 103*(1), 56-115.

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. Clinical Neurophysiology, 118(10), 2128-2148. doi: 10.1016/j.clinph.2007.04.019

Rabovsky, M., Sommer, W., & Abdel Rahman, R. (2012). The time course of semantic richness effects in visual word recognition. *Frontiers in Human Neuroscience, 6*, 11. doi: 10.3389/fnhum.2012.00011

Ratcliff, R., Gomez, P., & McKoon, G. (2004). A diffusion model account of the lexical decision task. *Psychological Review, 111*(1), 159-182. doi: 10.1037/0033-295X.111.1.159

Ratcliff, R., Van Zandt, T., & McKoon, G. (1999). Connectionist and diffusion models of reaction time. *Psychological Review, 106*(2), 261-300.

Reicher, G. M. (1969). Perceptual recognition as a function of meaninfulness of stimulus material. *Journal of Experimental Psychology, 81*(2), 275-280.

Rogers, T. T., Lambon Ralph, M. A., Garrard, P., Bozeat, S., McClelland, J. L., Hodges, J. R., & Patterson, K. (2004). Structure and deterioration of semantic memory: a neuropsychological and computational investigation. *Psychological Review, 111*(1), 205-235. doi: 10.1037/0033-295X.111.1.205

Rose, N. S., LaRocque, J. J., Riggall, A. C., Gosseries, O., Starrett, M. J., Meyering, E. E., & Postle, B. R. (2016). Reactivation of latent working memories with transcranial magnetic stimulation. *Science, 354*(6316), 1136-1139. doi: 10.1126/science.aah7011

Rugg, M. D. (1990). Event-related brain potentials dissociate repetition effects of high- and low-frequency words. *Memory and Cognition, 18*(4), 367-379.

Rugg, M. D., & Nagy, M. E. . (1987). Lexical contributions to nonword-repetition effects: Evidence from event-related potentials. *Memory and Cognition, 15*, 473-481.

Taft, M. . (1986). Lexical access codes in visual and auditory word recognition. *Language and Cognitive Processes, 1*(4), 297-308.

Taler, V., Kousaie, S., & López Zunini, R. (2013). ERP measures of semantic richness: the case of multiple senses. *Frontiers in Human Neuroscience, 7*, 5. doi: 10.3389/fnhum.2013.00005

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: the leaky, competing accumulator model. *Psychological Review, 108*(3), 550-592.

van den Brink, D., Brown, C. M., & Hagoort, P. (2001). Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 effects. *Journal of Cognitive Neuroscience, 13*(7), 967-985. doi: 10.1162/089892901753165872

van den Brink, D., & Hagoort, P. (2004). The influence of semantic and syntactic context constraints on lexical selection and integration in spoken-word comprehension as revealed by ERPs. *Journal of Cognitive Neuroscience, 16*(6), 1068-1084. doi: 10.1162/0898929041502670

Van Petten, C., & Kutas, M. . (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Memory and Cognition, 18*(4), 380-393.

Wessel, J. R., & Aron, A. R. (2015). It's not too late: the onset of the frontocentral P3 indexes successful response inhibition in the stop-signal paradigm. Psychophysiology, 52(4), 472-480. doi: 10.1111/psyp.12374

Wheeler, D. D. (1970). Processes in word recognition. *Cognitive Psychology, 1*(1), 59-85. doi: 0.1016/0010-0285(70)90005-8