Early language dissociation in the bilingual mind: Sensor-level MEG evidence

Nicola Molinaro [1,2], Sanjeev Nara [1,3], & Manuel Carreiras [1,2,4]

[1] Basque center on Cognition, Brain and Language, Donostia/San Sebastian, Spain

[2] Ikerbasque, Basque Foundation for Science, Bilbao, Spain

[3] Mathematical Institute, Department of Mathematics and Computer Science, Physics, Geography, Justus-Liebig-Universität Gießen (University of Giessen), Gießen, Germany

[4] University of the Basque Country. UPV/EHU

Abstract

Is language selection decodable from neural activity in balanced bilinguals? Previous research employing various neuroimaging methods has not yielded a conclusive answer to this issue. However, direct brain stimulation studies in bilinguals have detected different brain regions related to language production in separate languages. In the present MEG study, we addressed this question in a group of proficient Spanish-Basque bilinguals (N=45), who performed two tasks. In a picture naming task, they were asked to name a picture either in Basque or in Spanish if the ink turned from black to green after one second (10% of trials). Similarly, in a word reading task, participants read printed words out loud only when the ink turned to green. Sensor-level evoked activity was similar for the two languages on both tasks. However, decoding analyses classified the language used in both tasks, starting ~100 ms after stimulus onset. Crucially, searchlight analyses revealed that the right occipital-temporal sensors contributed most to language decoding in the picture naming task, while the left occipital-temporal sensors contributed most to decoding in word reading. Temporal generalization analyses for the picture naming task showed a sustained effect, while for the word reading task two stages of processing (at ~100 ms and starting ~300 ms) were evident. Cross-task decoding analysis highlighted robust generalization effects from the picture naming to the word reading task in the later time interval. The present findings provide support for the possibility of exploring the trial-by-trail variability in MEG activity for dissociating languages in the bilingual brain. For the first time, we provide evidence for a differential role of the two hemispheres in language selection for picture naming and word reading. We also provide novel evidence on the shared language representations between naming and word reading.

Keywords: Bilingualism, Neural decoding, Magnetoencephalography, Lexical processing, Word Reading, Picture Naming

Introduction

Bilingual language processing implies dealing with different linguistic codes at the same time: the bilingual speaker must keep both languages ready for efficient use in real life, regardless of whether information from one language is segregated from the other or whether both languages are represented together. This astonishing ability has continually (Fabbro, 2001; Ojemann and Whitaker, 1978) attracted neuroscientists to the challenge of determining if different neural substrates are recruited by the use of different languages in the brain.

A number of recent meta-analyses have explored if there is supporting evidence for distinct neural correlates of the processing of different languages in the bilingual brain (for a recent overview, see Comstock and Oliver, 2021). Overall, the reviewed evidence highlights that the core cortical network activated by words in different languages is the same, with relatively minor differences that could be explained by differences in proficiency in the two languages, the use of those languages, or the relative age of acquisition. Nonetheless, when considering balanced bilinguals, highly proficient in both languages, no differences are typically observed across neuroimaging techniques: left hemisphere frontotemporal regions are activated across languages and tasks (Hernandez et al., 2000; Jones et al., 2012), as well as, to a lesser extent, homologous regions of the right hemisphere (e.g., Geng et al., 2022; Nichols et al., 2021), especially in lexical/semantic tasks (Sulpizio et al., 2020). This "null effect" supports the idea that the neural infrastructure supporting the two languages is not differentiated and, consequently, that the lexical repertoire of the bilingual mind is integrated and indistinct for the two languages. This "integrated view" is supported by evidence that the use of words in one language mentally activates the words of the other language even in cases where the two linguistic codes are highly different (Emmorey et al., 2008; Morford et al., 2011; Thierry and Yan, 2007; Villameriel et al., 2022).

In addition, ERP studies highlighted an early influence on target word processing from lexical neighbors of the other language (e.g., Midgley et al., 2011; Peeters et al., 2013). This indicates that, during reading, language selection is a relatively late process, preceded by earlier lexical activation dynamics that are not language specific. Along these lines, the Bilingual Interactive Activation + (BIA+) model (Dijkstra and van Heuven, 2002) assumes a high level of cross-language interaction in lexical processing for balanced bilinguals. Bilingual lexical processing does not seem to recruit highly segregated mental representation or tap into separate cognitive mechanisms for the two languages. These points support the observation that the neural

underpinnings for bilingual language processing do not necessarily need to rely on functionally and spatially separated networks.

However, evidence against the "integrated view" has been provided by direct brain stimulation research (Lucas et al., 2004; Ojemann and Whitaker, 1978). During awake brain stimulation, it is typically observed that stimulation of different areas can selectively impair the use of one language or the other. Based on this, the idea that different neural networks are recruited for separate languages should be reconsidered. One possibility is that the same macro-level brain structures are recruited for both languages, while differences could emerge at a more micro-level: highly overlapping circuits would be recruited for the neural computation of the two languages, and the functional language segregation could emerge by better scrutinizing the trial-by-trial variability in the neural correlates. Classical univariate analysis would not be suited for exploring such variability since the individual brain responses are averaged across multiple trials and the fine-grained trial-by-trial variability is lost. A similar rationale has been put forward by Xu and colleagues (2017), who explored the multivariate pattern of fMRI responses to different languages in a group of bilinguals. Instead of performing univariate analysis they employed a multivariate approach and observed functional separation for the two languages across the regions of the language brain network. This methodological approach inspired the present investigation, in which we focused on the functional dissociation between the time-courses of the neural responses to the two languages in balanced bilinguals.

In the present MEG (magnetoencephalography) study, we used a picture naming and a word reading task to assess covert production of lexical information within the same group of Basque-Spanish balanced bilinguals. We employed a classic evoked activity analysis approach (*ERFs, Event-Related Fields*) to compare the electrophysiological correlates of picture naming and word reading in the two languages. More importantly, we used a machine learning approach (Cichy et al., 2014; Grootswagers et al., 2017; King and Dehaene, 2014; Nara et al., 2021a) to explore the trial-by-trial variability in the two languages and evaluate a possible more fine-grained functional language classification in the neural responses. This analysis reveals, across time-points, if the functional response to the two languages can be accurately decoded, thus showing when the neural activity becomes language specific (*temporal decoding*). We then performed a searchlight analysis across the whole time window to investigate the sensor locations contributing maximally to the temporal decoding. We also analyzed if the decoder trained at a specific time-

point could generalize across the other time points (*temporal generalization*). This analysis can reveal if separate sources of variability explain decoding accuracy at the two time points - evidence of separate neurocognitive processes - or the same source of variability underlies decoding accuracy at both time points - evidence of the same neurocognitive processes. Finally, we explored if the decoder trained at a specific time-point for a specific task (either picture naming or word reading) could generalize across the time points of the other task (*cross-decoding*). This approach will reveal if language specific representations at a specific time in one task could explain brain activity across time points in the other task. Overall, we expect the decoding approach to provide a more precise information concerning bilingual lexical processing compared to classical evoked activity analyses.

Methods

Participants. Forty-five Spanish-Basque bilinguals (29 females; age range: 19-42; mean age: 27.1 years; standard deviation: 5.2) took part in the present study. All of them were right-handed and had normal or corrected-to-normal vision. All participants acquired both languages (Spanish and Basque) before three years old. All participants went through a short semi-structured oral proficiency interview in each of their two languages (de Bruin et al., 2017) on which they all scored higher than 4 (on a 5-point scale) in both languages. In a picture naming test, across a set of 65 items per language, participants correctly named 64.6 pictures in Spanish (sd: 0.83; range 61-65) and 61.8 pictures in Basque (sd: 2.86; range 56-65), showing a high level of proficiency in both languages. All participants gave their written informed consent in accordance with ethical guidelines approved by the Research Committees of the Basque Center on Cognition, Brain and Language.

Stimuli. We used a set of 321 line-drawn pictures representing concrete nouns that were used in previous studies (e.g., Rueckl et al., 2015). From this initial set, we selected 288 pictures (mainly determined by the exclusion of pictures referring to phonologically similar words in the two languages, i.e., cognates) that unambiguously referred to a specific noun (name agreement > 80%, based on (Gisbert-Muñoz et al., 2021) These pictures were used to create six lists of 96 stimuli in which we balanced the overall lexical properties of the nouns referring to each picture. Across

lists, each item was present twice. Reference databases (Spanish: EsPal, Duchon et al., 2013; Basque E-Hitz, Perea et al., 2006) differed in corpus size, however these resources still provide useful information about the words referring to each picture in both languages. When considering all 288 items, Spanish words had higher frequency values (LogFreq: mean=1.0, sd=0.6) than Basque words (LogFreq: mean=0.9, sd=0.7), and were shorter (number of letters: mean=5.9, sd=1.8) than Basque words (mean=6.3, sd=2). In addition, words in Spanish had fewer syllables (mean=2.5, sd=0.8) compared to Basque (mean=2.7, sd=0.9). Importantly, when considering the separate lists, lexical parameters of the nouns referring to each picture did not statistically differ between languages within each list (all p-values > 0.1). Based on those lists, 96 items were assigned to each task and condition (naming or reading, either in Spanish or in Basque). In this process, 96 items (the word/picture pair) were used twice; however, if one item was used in the word reading task in one language (e.g., Spanish), we made sure it was used in the picture naming task in the other language (e.g., Basque). Order of presentation was pseudorandomized so that across participants (i) each picture was named in each language and (ii) each word was read in each language the same number of times.

Procedure. Stimuli were delivered using Psychtoolbox 3 (Brainard, 1997) in Matlab 2014b. Visual stimuli were presented in the MEG room on a back-projection screen with an external projector at 60 Hz. The screen was placed at 60 cm from the participant´s head, so that the stimuli covered a maximum visual angle of 5 degrees at the center of the screen. Participants performed the picture naming and word reading tasks separately in one session. Each task was divided into 16 blocks of 12 trials in one of the two languages. To ensure participants understood the instructions, they performed 20 practice trials under the experimenter's observation at the start of each task (10 in each language). Overall, including participant preparation, the whole experimental session lasted ~1.5 hours.

In the picture naming task, a fixation point was presented at the center of the screen for a jittered interval of 2500 ms on average (± 1500 ms) followed by a picture drawn in black ink for an additional 1000 ms. On 10% of trials the ink turned to green, prompting the participant to name the picture out loud. Otherwise, the fixation point reappeared on the screen for the next trial. Trials requiring overt speech production were not considered in the analysis. Each block of twelve trials was introduced by the instruction of the language that should be used for naming: either a Spanish

or a Basque flag preceded the block. At the end of each block a self-paced break was included. Naming accuracy was monitored by the experimenter in the control room (each participant's accuracy > 91%).

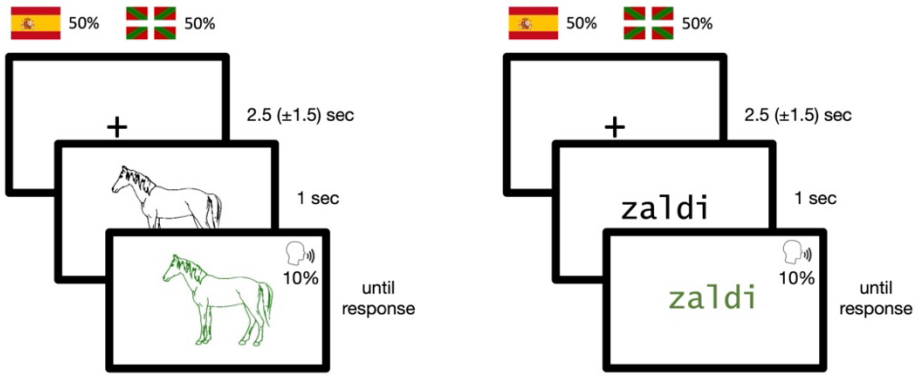

Figure 1: Example of one trial for the picture naming task (left) and the word reading task (right).

In the word reading task, we used the same experimental structure (jittered fixation point followed by the stimulus at the center of the screen). However, instead of pictures, words printed in black (turning to green in 10 % of cases) were presented (font: Lucida Console; size: 34). The same instruction at the beginning of each block informed the participant about the language of the next trials. Reading accuracy was monitored by the experimenter in the control room (each participant's accuracy > 94%).

Magnetoencephalographic recordings. MEG data were acquired in a magnetically shielded room using the whole head MEG system (MEGIN-Elekta Neuromag, Finland) installed at BCBL (https://www.bcbl.eu/en/infrastructure-equipment/meg). The MEG system consists of 306 sensors (102 magnetometers and 204 gradiometers arranged in a helmet configuration). The head position during the experiment was continuously monitored with five Head positioning coils (HPI). The location of HPI coils were defined relative to three fiducial points (nasion, left preauricular (LPA) and right preauricular point (RPA)) using 3D head digitizer (Fastrack Polhemus, Colchester, VA, USA). This process is critical for correction of movement artifacts (movement compensation) during data acquisition. The MEG data was recorded continuously with a bandpass filter (0.01 –

330 Hz) and a sampling rate of 1 kHz. Eye movements were monitored with two pairs of electrodes placed above and below each eye (horizontal electrooculogram - HEOG), and on the external canthus of both eyes (vertical electrooculogram - VEOG). Similarly, electrocardiogram (EKG) was also recorded using two electrodes (bipolar montage) placed on the right side of the abdomen and below the left clavicle of the participant. The continuous MEG data were pre-processed offline to suppress the external electromagnetic field using temporal Signal-Space-Separation (tSSS; Taulu and Simola, 2006) method implemented in Maxfilter software (MaxFilter 2.0). The MEG data were also corrected for movement compensation and bad channels were repaired within the MaxFilter algorithms. All the following analyses were carried out in Matlab version 2021B.

Data analysis

Event-Related Fields. Continuous MEG data were first segmented in time intervals starting 1 second before the onset of the visual stimuli until 2 seconds afterwards. Based on visual inspection, we detected epochs containing large artifacts (SQUID jumps, large head movements) and excluded them from the following analyses. On average 3% of trials were rejected, with no differences among conditions. A FastICA algorithm (Hyvärinen, 1999) was then used to detect independent components which can represent either eye-movements or heart-beats. Across participants we excluded 1 to 3 components based on visual inspection. We focused on planar gradiometers for the evoked analysis. After baseline correction (−500 to −10 ms), gradiometer pairs in the same location were combined using the root-mean square approach. Afterwards, we separated the trials based on conditions for averaging, independently for each participant. Statistical analyses were performed comparing the neural activity in the two languages (Spanish and Basque) independently for each task (picture naming and word reading) using the cluster-based approach (10000 permutations, cluster alpha: $p<0.05$, alpha: $p<0.05$) on the time interval of interest (0-1000 ms) across the whole set of sensors (Maris and Oostenveld, 2007).

Time-resolved decoding. Time-resolved decoding (also called temporal decoding) was used to decode the language used by participants (Spanish or Basque) from neural patterns recorded by MEG. The preprocessed MEG data in both tasks (picture naming and word reading) were segmented from 200 ms prior to and 1000 ms after the onset of the stimuli. The data were first low pass filtered to 25 Hz and then down sampled to 200 Hz. A covariance-based trial rejection was

performed to remove the outlier trials. The trials deviating by less than 2 percent and more than 95 percent from the mean covariance were excluded prior to the classification procedure (average trial rejection ratio 7%). The data were then fed to the classifier separately for each task to decode the language of use using a logistic regression classifier implemented in the MVPA Light toolbox (Treder, 2020). The Classification was performed separately at each time point in both tasks (Words and Pictures). We used k-fold cross-validation (k=5) to avoid overfitting and used Area (*auc*) under the receiver operating characteristic curve (ROC) for evaluating the classification performance. We also used an '*undersampling*' approach to balance the classes before feeding the data for classification. This process was repeated 10 times for each subject to yield stable decoding patterns. The statistical significance of decoding across time was established using cluster corrected sign permutation tests (one-tailed) (similar to Dima et al., 2018; Nara et al., 2022, 2021b) applied to the '*auc*' values obtained from the classifier with cluster-defining threshold (i.e., alpha: $p<0.05$; cluster-alpha: $p<0.05$; 10000 permutations).

Searchlight across sensor-space. A searchlight analysis was implemented to determine the sensors contributing most to decoding the language from sensor-space MEG data. In our experiment, the data preprocessing steps (i.e., low pass filtering at 25 Hz, down-sampling at 200 Hz, logistic regression classifier, 5-fold cross validation, *auc* as accuracy metrics) were the same as explained in the previous section. Unlike time-resolved decoding, the logistic classifier considered all the time points (i.e., −200 ms to 1000 ms after the onset of stimulus presentation) to decode the language in both tasks separately (word reading and picture naming). To yield stable decoding patterns, the classification procedure was repeated 10 times for each subject. Since a longer time window was used, a strict and stringent threshold was applied to establish the statistical significance compared to time-resolved decoding. Cluster corrected sign permutation tests (one-tailed) were applied with cluster defining thresholds (i.e., alpha: $p<0.001$; cluster-alpha: $p<0.001$; 20000 permutations).

Temporal Generalization. Time-resolved decoding helps to understand how cognitive processes unfold over time; however, this method does not reveal anything about the temporal organization of information processing stages in the human brain. The temporal generalization method measures the ability of a classifier to generalize across time (King et al., 2016; King and Dehaene,

2014). This method provides a novel way to understand how mental representations are manipulated and transformed for a given cognitive process. In our experiment, the data preprocessing steps (i.e., low pass filtering at 25 Hz, down-sampling at 200 Hz, logistic regression classifier, 5-fold cross validation) were the same as explained in the previous sections. Here a classifier is trained on each particular time point in the data and then tested over all other available time points separately for both tasks (i.e., picture naming and word reading). This results in a square matrix per block per participant where each point reflects the generalization strength of the classifier at a particular time (x axis) given the training at a given time point (y axis). To evaluate the presence of a group level effect, cluster-based permutation test was used with cluster-defining threshold (i.e., alpha: $p<0.05$; cluster-alpha: $p<0.05$; 10000 permutations).

Cross Decoding. Cross decoding is a variant of temporal generalization which generalizes across different experiments, tasks, or blocks. This method is used to understand how the mental representations in one cognitive scenario can predict brain activity in another scenario. In this study, the classifiers were first trained using the data from the picture naming task and then tested over the time points from the word reading task; this process was then repeated in reverse (classifiers trained on reading data and then tested on naming data). The parameters for classification were the same as explained in the previous three sections.

Results

Evoked activity. ERFs (reported in Figure 2) computed separately for the two languages in the picture naming task showed peaks of activity at around 130 ms in the occipital sensors, around 250 ms in occipito-temporal sensors bilaterally and around 450 ms in occipital and temporal sensors. Cluster-based permutations (Spanish vs. Basque) did not reveal any significant differences between the two conditions. However, based on visual inspection it could be noticed that the early peak around 130 ms appeared to be reduced for Spanish compared to Basque. We therefore ran a post-hoc analysis in a time interval between 100 and 150 ms. The cluster-based approach indicated a negative cluster but this was not statistically significant ($p=0.23$).

In the word reading task, two early peaks could be clearly observed: at around 130 ms in the occipital sensors and at around 150 ms in the posterior left occipito-temporal sensors. A later

peak was observed at around 400 ms, mainly evident in the left temporal sensors. Cluster-based permutations (Spanish vs. Basque) did not reveal any significant differences between the two conditions. Visual inspection showed a potential difference starting around 400 ms. A post-hoc analysis in a time interval between 400 and 500 ms indicated a negative cluster but this was not statistically significant (p=0.37).

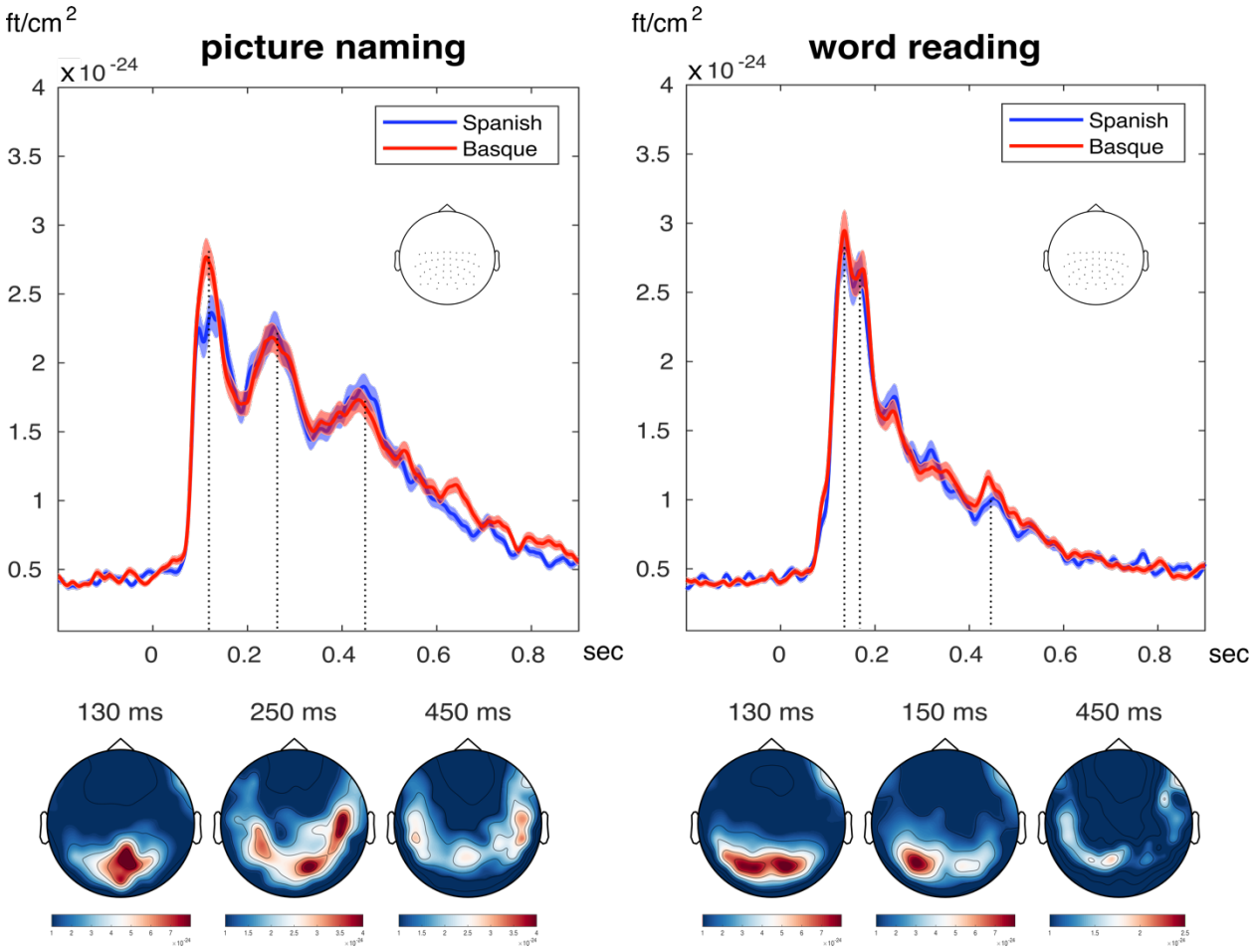

Figure 2: The top panels show event-related fields (combined gradiometers representation, averaged across a set of posterior sensors) elicited by Spanish (blue) and Basque (red) in the two experimental tasks. Shaded bars reflect the standard error from the mean. The bottom panels show topographies reflecting the sensors with maximum activity at representative time points marked with dotted lines in the time-course above.

Time-resolved decoding. Neural representations of language (Spanish or Basque) were successfully decoded in both picture naming and word reading tasks. We observed one significant cluster (p<0.001) for both blocks (Figure 3). The language specific representations dissociate

slightly earlier in picture naming (starting at 82 ms) compared to word reading (starting at 117 ms) and were sustained until the end of the time window considered. We also observed differences in peak decoding in the two tasks. The decoding for word reading peaked at 162 ms after the presentation of the stimulus whereas the peak decoding for the picture naming was at 432 ms. Thus, the decoding in the word reading task emerged slightly later but reached the peak earlier whereas in picture naming the decoding emerged earlier and slowly increased until ~300-400 ms.

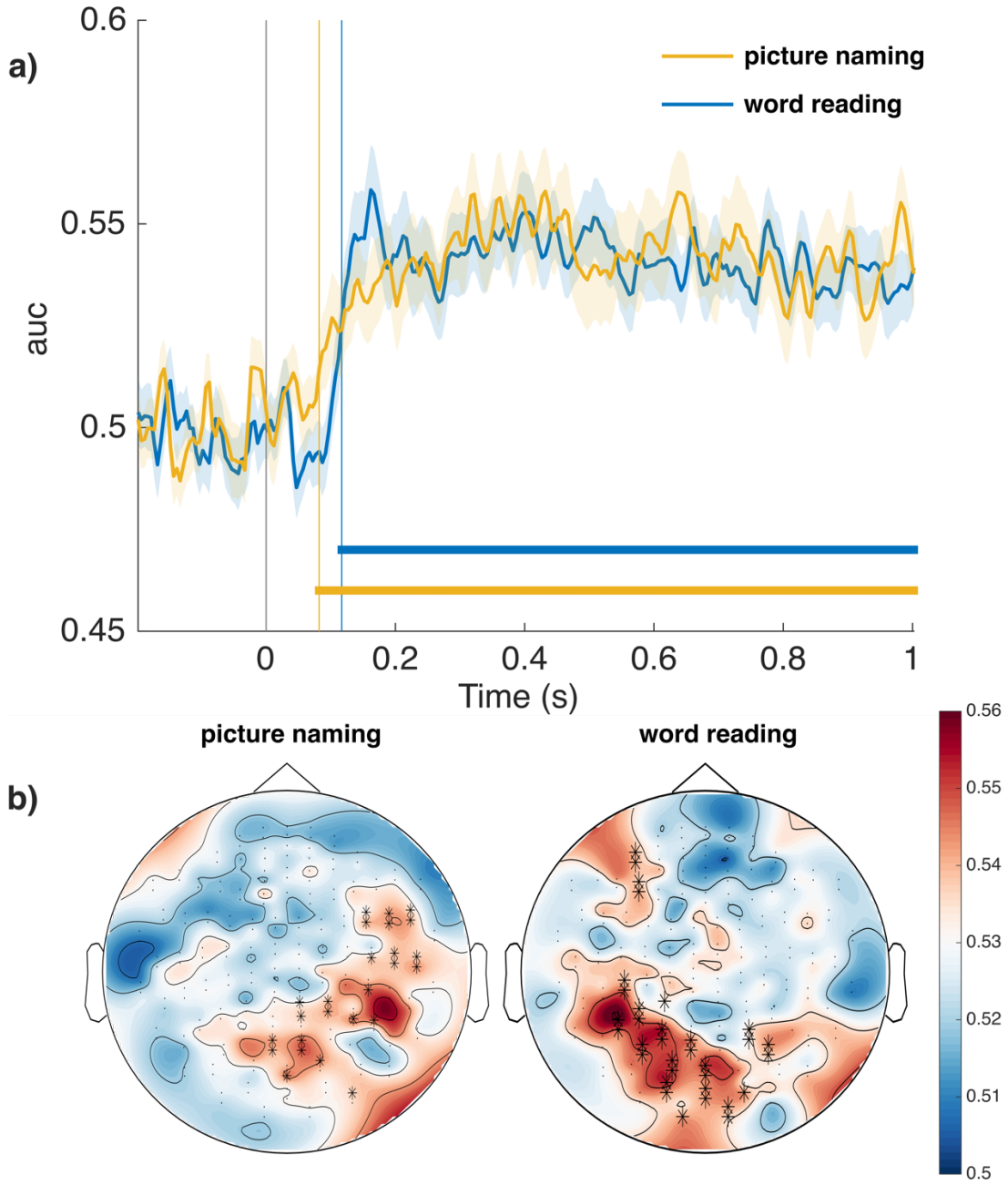

Figure 3: a) Time-resolved language decoding (Spanish vs. Basque) in word reading (blue) and picture naming (yellow). The horizontal lines in the figure indicate time points showing significant decoding above chance level ($p < 0.05$). b) Searchlight analysis of language decoding (Spanish vs. Basque) in word reading (left) and picture naming (right). The black asterisks (*) indicate the channels showing significant contribution to the decoding above chance ($p<0.05$).

Searchlight across sensor-space. Searchlight analysis (Figure 3b) successfully identified the group of sensors contributing to decoding the selected language in both word reading and picture naming. We observed a lateralization effect for word reading and picture naming in opposite hemispheres. The word reading task shows most successful language decoding in the left hemisphere and the effect was dominant in occipital-temporal sensors. A smaller cluster also emerged in left fronto-temporal sensors. The picture naming language decoding was more lateralized to the right hemisphere. The effect was dominant in right occipital and temporal sensors.

Temporal generalization. Temporal generalization matrices for word reading and picture naming show the highest diagonal decoding as reflected by a long-lasting significant cluster. Similar to the time-resolved decoding, the picture naming data showed early reliable language decoding (172 ms) with a peak at ~560 ms, reflecting a slowly increasing dissociation between the languages across time. Here the generalization plot indicates that the language classifier can most reliably generalize across time points after ~300 ms.

In the word reading data the decoding starts a little bit later (~150 ms) but peaks much earlier at 173 ms post-stimulus onset. After this initial activity a second accuracy peak become evident at ~400 ms. In word reading, the classifiers trained at ~150 ms (close to the peak decoding effect) could generalize to the neural activity arising later, ~250-750 ms after the stimulus presentation.

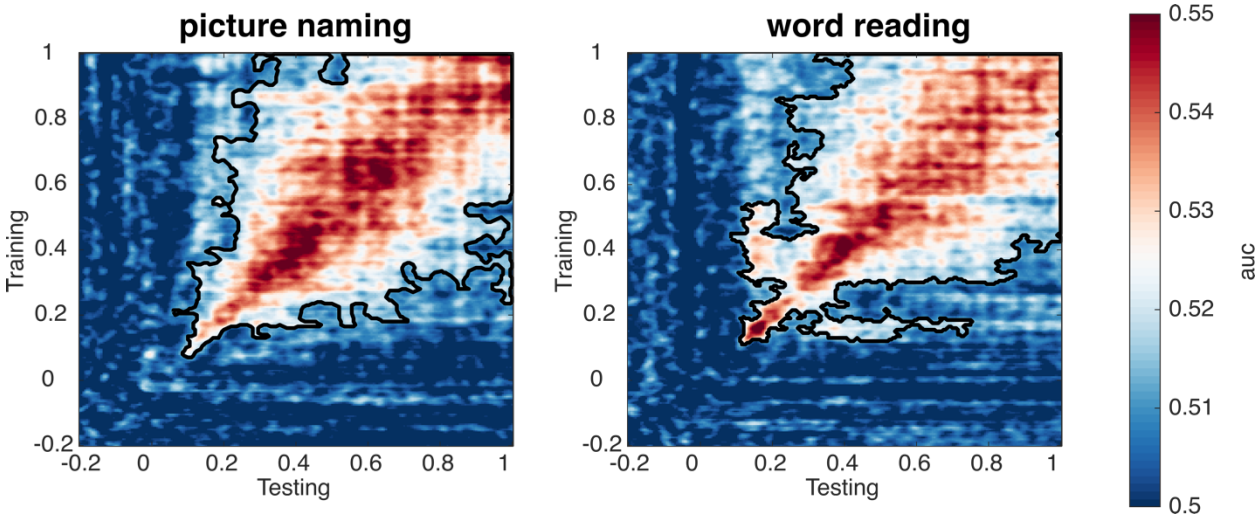

Figure 4: Temporal generalization matrices for language decoding (Spanish vs. Basque) in picture naming (left) and word reading (right). The black line in the figure indicates the boundary of significant time points emerging from cluster-based permutation test ($p < 0.05$).

Cross-task decoding. In the previous section we observed that the language representations recruited for word reading and picture naming follows a slightly different generalization pattern. It would be interesting to estimate if there is any overlap between the language representations recruited by the two tasks. To this aim, we computed cross decoding across picture naming and word reading.

In Figure 5 (left), we report how the classifier trained on picture naming generalizes to the word reading task. An early short-lasting effect is visible, where the picture naming classifier trained at around 77-90 ms generalizes to the neural activity starting from 250 post-stimulus onset. A more robust effect emerges later, with the picture naming data from the 200-700 ms time interval eliciting reliable classification for the word reading data in the ~250-500 ms time interval with an accuracy peak at ~300 ms.

In Figure 5 (right), we report how the classifier trained on word reading can generalize to the picture naming task. Here it is possible to appreciate how the reading data around 230-300 ms can trigger reliable language classification in picture naming task in a long-lasting ~100-800 ms time interval.

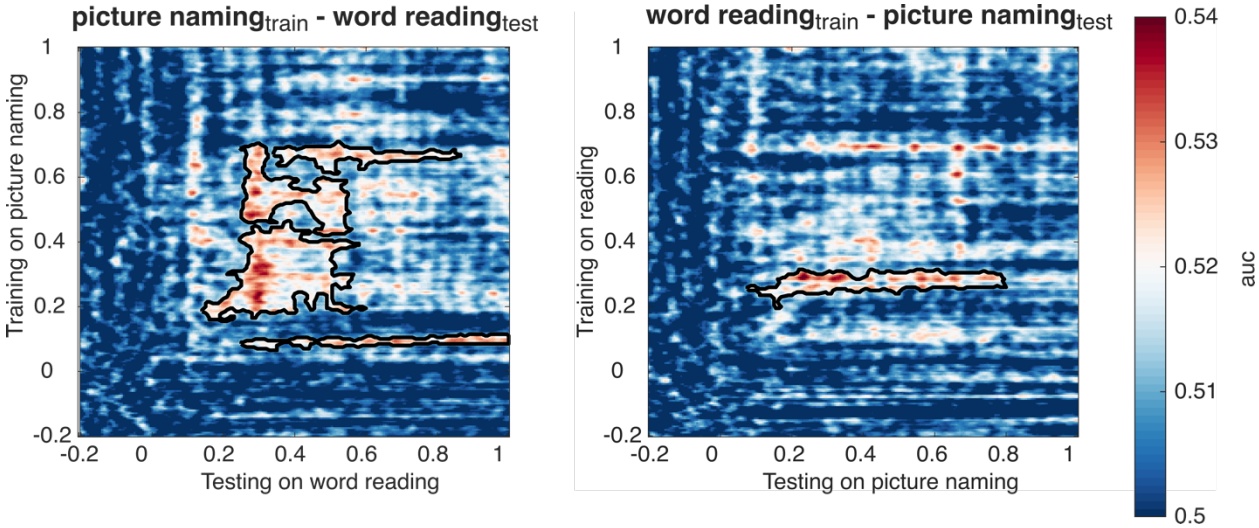

Figure 5: Cross decoding results for language (Spanish vs. Basque). The figure on the left was trained on picture naming and tested on the word reading data whereas the figure on the right was trained on word reading data and tested on picture naming. The dotted black line in the figure indicates the boundary of significant time points emerging from cluster-based permutation test ($p < 0.05$).

Discussion

The present study tested the potential of non-invasive neuroimaging for decoding the language produced by a bilingual language user. While invasive brain imaging has highlighted that distinct brain regions could causally modulate production of one of the two languages in bilingual speakers, noninvasive brain imaging has been elusive in providing experimental evidence for such dissociation. Importantly, this problem has theoretical implication for models of bilingual lexical processing. In their BIA+ model, Dijkstra and Van Heuven (2002) proposed a fully integrated and undifferentiated mental space for the lexical representations of two languages. Finding neural evidence for their dissociation would cast some doubts on this theoretical account.

Our experimental approach exploited two methodological advantages regarding our participant sample and data analysis. First, the proficiency of our participants in their two languages was balanced. As highlighted by several recent reviews (for a summary, see Comstock and Oliver, 2021), proficiency is a critical factor in accounting for language processing neural specificity in bilinguals. In the case of balanced bilinguals, the literature leans towards the conclusion that the two languages recruit the same neural circuitry (for a discussion, see Xu et al., 2017). Importantly, we took advantage of a sample of bilingual speakers who have equal

proficiency in two languages (Spanish and Basque) that present very different linguistic (typological) properties, especially at the lexical level. Focusing on this bilingual population provides a perfect test bench to evaluate if two languages are associated with different neural correlates.

Second, the present study took inspiration from recent advances in fMRI research (e.g., Xu et al., 2017) that employed a fine-grained machine learning approach in accounting for the variability of the neural data. This approach showed high sensitivity in differentiating linguistic representations that recruit highly overlapping neural networks. In our study, the standard analysis of the evoked activity did not show reliable sensitivity to the language used by our participants in either of the two tasks. However, we adopted a novel time-resolved decoding approach, in order to take advantage of the item-level variability in the neural responses. This latter analysis returned positive results, showing that it was possible to decode the language employed by our balanced bilinguals in two different lexical processing tasks. To our knowledge, this is one of the first successful attempts to dissociate language-specific lexical processing in balanced bilinguals with a non-invasive neuroimaging method.

One important aspect of the present study was the use of two complementary linguistic tasks, picture naming and word reading. Interestingly we observed very early language dissociation (~100 ms) in both tasks. This early timing emerges much earlier compared to classical language-related ERP components, which have been reported starting around ~200 ms (for single word reading: Kutas and Federmeier, 2011; Swaab et al., 2012; for picture naming: Ganushchak and Schiller, 2009; Indefrey and Levelt, 2004). In the reading task, one could argue that an early visual distinction of the word stimuli in the two languages could have triggered such an early effect (although Casaponsa and Duñabeitia, 2016, reported orthographic effects later in time, ~250 ms). However, in the picture naming task we used the same images for both languages, meaning such an early effect cannot be ascribed to visual dissimilarities in the stimulus materials. Overall, this result has relevant implications for theoretical accounts of the neural dynamics of language processing. Our data suggest that the divergence of language-specific processing in the bilingual brain occurs prior to early hierarchical linguistic processes such as orthographic and lexical processes in word reading or lexical and phonological processes in picture naming. The early language selection observed in the present study could be due to our participants knowing which language they were supposed to use before the presentation of the actual visual stimulus (language

was indicated before each block of 12 items). However, this would imply that very slow developing neural activity lasting for a whole block (and not strictly time-locked to each stimulus) is different for the two languages, thus triggering the long-lasting effects we found here with our decoding approach. Still, it is worth noting that such "differential drifting" was not evident in the standard analysis of the evoked activity, nor in the decoding analysis in the time window prior to stimulus presentation. Thus, we do not think that this early effect was driven by a difference already building up during the baseline period. We maintain our interpretation of the present findings: the item-level representations showed different patterns of variability in the two languages, and these separate sources of variance dissociate early in time.

From a linguistic point of view, determining which level of linguistic representation contributed more to the decoding of the language of use is not straightforward based on the present experimental design. In fact, the timings and the generalization patterns (Figures 3 and 4) were not very dissimilar in the two tasks. One difference that could be observed is that in picture naming the decoding accuracy grew slowly, reaching its peak late (> 500 ms), while in word reading the accuracy reached its peak very early (~170 ms). In previous research we already showed this asymmetry between the processing of words and pictures in a language task (Giannelli and Molinaro, 2018). Word production models (e.g., Indefrey and Levelt, 2004) assume that pictures first activate lexical-semantic representations that cascade information to sub-lexical (phonological and articulatory) processing later. On the other hand, reading models assume early sub-lexical (visual and orthographic) processing supporting lexical-semantic activation (e.g., Carreiras et al., 2014). Based on this asymmetry, both lexical/semantic and phonological/orthographic representations contribute to successful language decoding, perhaps with slightly better decoding accuracy triggered by sub-lexical features. However, this last claim cannot be fully supported based on the present results since we did not observe any statistical difference in the accuracy level in the two tasks. Nonetheless, we think it is a hypothesis that requires further scrutiny.

A clearer dissociation between the two tasks was observed in the searchlight analyses. While in the word reading task, the posterior and left-lateralized sensors contributed more to language classification, in picture naming the pattern was almost the opposite, with more contribution from right-lateralized sensors. First of all, it is interesting to observe such a strong posterior distribution of such effects, similar to previous MEG studies on picture naming (Geng et

al., 2022; Salmelin et al., 1994). Second, while the left-lateralized effect reported for word reading could be predicted based on reading models (e.g., Rueckl et al., 2015), the right-lateralized effect observed for picture naming is less expected. In reading, the involvement of the right hemisphere appears to be mainly related to proficiency levels (Cargnelutti et al., 2019). Since we focused on a group of balanced bilinguals, proficiency should not affect our decoding analysis. On the other hand, activation of different neural representations involved in reading (Dehaene et al., 2015), could be the critical factor boosting decoding accuracy. For picture naming, a critical involvement of the right hemisphere has been discussed in previous fMRI metanalysis (see Comstock and Oliver, 2021; Sulpizio et al., 2020). It is possible that, even if picture naming recruits both hemispheres (Geng et al., 2022), the language neural dissociation is stronger in the right-lateralized network. In previous studies, the right temporal regions have been associated with language control (Luk et al., 2011), while the right supramarginal grey matter density was modulated by the vocabulary size of bilinguals (Grogan et al., 2012). Importantly though, the present speculations on why word reading and picture naming recruit different hemispheres during language selection in bilinguals are mainly based on functional and structural MRI data. Evidence from time-resolved decoding of neurophysiological data is scarce. Future research should determine if the task-related hemispheric dissociation we report here reflects a functional reorganization of the two hemispheres based on bilingual experience.

Another clear asymmetry between the two tasks emerged in the cross-decoding analysis, with a more extended effect emerging from the generalization of the picture naming to the word reading task. The generalization pattern indicates that brain activity elicited by picture naming between 200 and 700 ms could generalize to brain activity emerging in word reading at ~300 ms. Interestingly, the opposite was detected in the word reading-to-picture naming cross-decoding: word reading related brain activity at ~300 ms generalized to brain activity elicited by picture naming in the 200-800 ms time window. This indicates that language-specific representations that are commonly recruited by both tasks are active for a longer period in picture naming compared to word reading. This cross-task effect can be explained by the different time course in the recruitment of lexical-sematic information. On the one hand, in picture naming, such representations are recruited early and maintained for a long time for the selection of the correct sub-lexical units, even triggering a continuous self-monitoring loop that plays a critical role in language production (Levelt et al., 1999). On the other hand, in word reading, such lexical-

semantic representations are activated for a shorter amount of time during the classical N400 time interval. Critically, lexical-semantic information is not critical for reading aloud, since "surface reading" can be successfully achieved without access to the meaning of words (as implemented in the dual route cascaded (DRC) model by Coltheart et al., (2001). This explanation fits with the idea that common language-specific representations in the two tasks are lexical-semantic in nature and are differently recruited depending on the complexity of the language production processes required.

Overall, neural decoding analyses can contribute to our knowledge on language specific processing in the bilingual brain. By focusing on a population of Spanish-Basque balanced bilinguals and exploiting the item-specific variability in the neural correlates of two different language tasks, we provide novel evidence for the possibility of detecting language-specific neural correlates in the bilingual mind, and show that the two hemispheres are differentially recruited for word reading and picture naming. This bridges the gap between non-invasive and invasive experimental evidence on bilingual lexical processing. These results could pave the way for future studies that could focus on different experimental manipulations and the decoding of more fine-grained linguistic information during bilingual language processing.

Acknowledgments:

This research was supported by the Basque Government through the BERC 2018–2021 program and by the Spanish State Research Agency through BCBL's Severo Ochoa excellence accreditation CEX2020-001010-S and the project BES-2016-077560 funded by the Spanish Ministry of Economy and Competitiveness (MINECO). SN acknowledges the support from "The Adaptive Mind," funded by the Excellence Program of the Hessian Ministry of Higher Education, Science, Research and Art. MC was supported by "la Caixa" Foundation (ID 100010434), under the agreement HR18-00178-DYSTHAL, and by the Agencia Estatal de Investigación PID2021-122918OB-I00. NM was supported by the Spanish Ministry of Science, Innovation and University (grants PSI2015-65694-P, RTI2018-096311-B-I00), the Agencia Estatal de Investigación (AEI), the Fondo Europeo de Desarrollo Regional (FEDER) and by the Basque government (grant PI_2016_1_0014).

SN thanks the BCBL lab research staff for their valuable support.

Author contributions:

NM: Conceptualization, Methodology, Validation, Supervision, Funding acquisition, Project administration, Writing – review & editing. **SN**: Conceptualization, Methodology, Software, Data curation, Writing – original draft. **MC**: Conceptualization, Methodology, Supervision, Writing – review & editing.

Data availability statement:

The preprocessed dataset is available upon requests directed to Dr. Nicola Molinaro (n.molinaro@bcbl.eu) or Dr. Sanjeev Nara (sanjeev.nara@math.uni-giessen.de). The data set could then be shared through the private BCBL-secured institutional servers temporarily available for big data transfer.

References:

Brainard, D.H., 1997. The Psychophysics Toolbox. Spat Vis 10, 433–436. https://doi.org/10.1163/156856897X00357

Cargnelutti, E., Tomasino, B., Fabbro, F., 2019. Language brain representation in bilinguals with different age of appropriation and proficiency of the second language: A meta-analysis of functional imaging studies. Front Hum Neurosci 13, 154. https://doi.org/10.3389/FNHUM.2019.00154/BIBTEX

Casaponsa, A., Duñabeitia, J.A., 2016. Lexical organization of language-ambiguous and language-specific words in bilinguals. Quarterly Journal of Experimental Psychology 69, 589–604. https://doi.org/10.1080/17470218.2015.1064977/ASSET/IMAGES/LARGE/10.1080_17470218.2015.1064977-FIG1.JPEG

Cichy, R.M., Pantazis, D., Oliva, A., 2014. Resolving human object recognition in space and time. Nat Neurosci 17, 455–462. https://doi.org/10.1038/nn.3635

Coltheart, M., Rastle, K., Perry, C., Langdon, R., Ziegler, J., 2001. DRC: a dual route cascaded model of visual word recognition and reading aloud. Psychol Rev 108, 204–256. https://doi.org/10.1037/0033-295X.108.1.204

Comstock, L., Oliver, B., 2021. A meta-analysis of task-based differences in bilingual L1 and L2 language networks. biorxiv. https://doi.org/10.1101/2021.12.28.474335

Dehaene, S., Cohen, L., Morais, J., Kolinsky, R., 2015. Illiterate to literate: behavioural and cerebral changes induced by reading acquisition. Nature Reviews Neuroscience 2015 16:4 16, 234–244. https://doi.org/10.1038/nrn3924

Dehaene, S., Cohen, L., Sigman, M., Vinckier, F., 2005. The neural code for written words: a proposal. Trends Cogn Sci 9, 335–341. https://doi.org/10.1016/J.TICS.2005.05.004

Dijkstra, T., van Heuven, W.J.B., 2002. The architecture of the bilingual word recognition system: From identification to decision. Bilingualism: Language and Cognition 5, 175–197. https://doi.org/10.1017/S1366728902003012

Dima, D.C., Perry, G., Messaritaki, E., Zhang, J., Singh, K.D., 2018. Spatiotemporal dynamics in human visual cortex rapidly encode the emotional content of faces. Hum Brain Mapp 39, 3993–4006. https://doi.org/10.1002/hbm.24226

Duchon, A., Perea, M., Sebastián-Gallés, N., Martí, A., Carreiras, M., 2013. EsPal: One-stop shopping for Spanish word properties. Behav Res Methods 45, 1246–1258. https://doi.org/10.3758/S13428-013-0326-1/TABLES/8

Emmorey, K., Borinstein, H.B., Thompson, R., Gollan, T.H., 2008. Bimodal bilingualism. Biling (Camb Engl) 11, 43. https://doi.org/10.1017/S1366728907003203

Fabbro, F., 2001. The Bilingual Brain: Bilingual Aphasia. Brain Lang 79, 201–210. https://doi.org/10.1006/BRLN.2001.2480

Ganushchak, L.Y., Schiller, N.O., 2009. Speaking one's second language under time pressure: an ERP study on verbal self-monitoring in German-Dutch bilinguals. Psychophysiology 46, 410–419. https://doi.org/10.1111/J.1469-8986.2008.00774.X

Geng, S., Molinaro, N., Timofeeva, P., Quiñones, I., Carreiras, M., Amoruso, L., 2022. Oscillatory dynamics underlying noun and verb production in highly proficient bilinguals. Scientific Reports 2022 12:1 12, 1–10. https://doi.org/10.1038/s41598-021-04737-z

Giannelli, F., Molinaro, N., 2018. Reanalyzing language expectations: Native language knowledge modulates the sensitivity to intervening cues during anticipatory processing. Psychophysiology 55, e13196. https://doi.org/10.1111/PSYP.13196

Gisbert-Muñoz, S., Quiñones, I., Amoruso, L., Timofeeva, P., Geng, S., Boudelaa, S., Pomposo, I., Gil-Robles, S., Carreiras, M., 2021. MULTIMAP: Multilingual picture naming test for mapping eloquent areas during awake surgeries. Behav Res Methods 53, 918–927. https://doi.org/10.3758/s13428-020-01467-4

Grogan, A., 'O. Parker Jones, T., Ali, N., Crinion, J., Orabona, S., Mechias, M.L., Ramsden, S., Green, D.W., Price, C.J., 2012. Structural correlates for lexical efficiency and number of languages in non-native speakers of English. Neuropsychologia 50, 1347–1352. https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2012.02.019

Grootswagers, T., Wardle, S.G., Carlson, T.A., 2017. Decoding dynamic brain patterns from evoked responses: A tutorial on multivariate pattern analysis applied to time series neuroimaging data. J Cogn Neurosci 29, 677–697. https://doi.org/10.1162/jocn_a_01068

Hernandez, A.E., Martinez, A., Kohnert, K., 2000. In search of the language switch: An fMRI study of picture naming in Spanish-English bilinguals. Brain Lang 73, 421–431. https://doi.org/10.1006/BRLN.1999.2278

Hyvärinen, A., 1999. Fast and robust fixed-point algorithms for independent component analysis. IEEE Trans Neural Netw 10, 626–634. https://doi.org/10.1109/72.761722

Indefrey, P., Levelt, W.J.M., 2004. The spatial and temporal signatures of word production components. Cognition 92, 101–144. https://doi.org/10.1016/j.cognition.2002.06.001

Jones, O.P., Green, D.W., Grogan, A., Pliatsikas, C., Filippopolitis, K., Ali, N., Lee, H.L., Ramsden, S., Gazarian, K., Prejawa, S., Seghier, M.L., Price, C.J., 2012. Where, when and why brain activation differs for bilinguals and monolinguals during picture naming and reading aloud. Cereb Cortex 22, 892–902. https://doi.org/10.1093/CERCOR/BHR161

King, J.R., Dehaene, S., 2014. Characterizing the dynamics of mental representations: The temporal generalization method. Trends Cogn Sci. https://doi.org/10.1016/j.tics.2014.01.002

King, J.R., Pescetelli, N., Dehaene, S., 2016. Brain Mechanisms Underlying the Brief Maintenance of Seen and Unseen Sensory Information. Neuron 92, 1122–1134. https://doi.org/10.1016/j.neuron.2016.10.051

Kutas, M., Federmeier, K.D., 2011. Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). Annu Rev Psychol 62, 621. https://doi.org/10.1146/ANNUREV.PSYCH.093008.131123

Levelt, W.J.M., Roelofs, A., Meyer, A.S., 1999. A theory of lexical access in speech production. Behav Brain Sci 22, 1–75. https://doi.org/10.1017/S0140525X99001776

Lucas, T.H., McKhann, G.M., Ojemann, G.A., 2004. Functional separation of languages in the bilingual brain: a comparison of electrical stimulation language mapping in 25 bilingual patients and 117 monolingual control patients. J Neurosurg 101, 449–457. https://doi.org/10.3171/JNS.2004.101.3.0449

Luk, G., Bialystok, E., Craik, F.I.M., Grady, C.L., 2011. Lifelong Bilingualism Maintains White Matter Integrity in Older Adults. The Journal of Neuroscience 31, 16808. https://doi.org/10.1523/JNEUROSCI.4563-11.2011

Maris, E., Oostenveld, R., 2007. Nonparametric statistical testing of EEG- and MEG-data. J Neurosci Methods 164, 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Midgley, K.J., Holcomb, P.J., Grainger, J., 2011. Effects of Cognate Status on Word Comprehension in Second Language Learners: An ERP Investigation. J Cogn Neurosci 23, 1634. https://doi.org/10.1162/JOCN.2010.21463

Morford, J.P., Wilkinson, E., Villwock, A., Piñar, P., Kroll, J.F., 2011. When deaf signers read English: do written words activate their sign translations? Cognition 118, 286–292. https://doi.org/10.1016/J.COGNITION.2010.11.006

Nara, S., Lizarazu, M., Richter, C.G., Dima, D.C., Cichy, R.M., Bourguignon, M., Molinaro, N., 2021a. Temporal uncertainty enhances suppression of neural responses to predictable visual stimuli. Neuroimage 239, 118314. https://doi.org/10.1016/j.neuroimage.2021.118314

Nara, S., Rathee, D., Molinaro, N., Bois, N. du, Bhushan, B., Prasad, G., 2021b. Temporal Dynamics of Neural Processing of Facial Expressions and Emotions. bioRxiv 2021.05.12.443280. https://doi.org/10.1101/2021.05.12.443280

Nara, S., Raza, H., Carreiras, M., Molinaro, N., 2022. Decoding Numeracy and Literacy in the Human Brain: Insights from MEG and MVPA. bioRxiv 2022.12.29.522254. https://doi.org/10.1101/2022.12.29.522254

Nichols, E.S., Gao, Y., Fregni, S., Liu, L., Joanisse, M.F., 2021. Individual differences in representational similarity of first and second languages in the bilingual brain. Hum Brain Mapp 42, 5433–5445. https://doi.org/10.1002/HBM.25633

Ojemann, G.A., Whitaker, H.A., 1978. The Bilingual Brain. Arch Neurol 35, 409–412. https://doi.org/10.1001/ARCHNEUR.1978.00500310011002

Peeters, D., Dijkstra, T., Grainger, J., 2013. The representation and processing of identical cognates by late bilinguals: RT and ERP effects. J Mem Lang 68, 315–332. https://doi.org/10.1016/j.jml.2012.12.003

Perea, M., Urkia, M., Davis, C.J., Agirre, A., Laseka, E., Carreiras, M., 2006. E-Hitz: A word frequency list and a program for deriving psycholinguistic statistics in an agglutinative language (Basque). Behav Res Methods 38, 610–615. https://doi.org/10.3758/BF03193893

Rueckl, J.G., Paz-Alonso, P.M., Molfese, P.J., Kuo, W.J., Bick, A., Frost, S.J., Hancock, R., Wu, D.H., Einar Mencl, W., Duñabeitia, J.A., Lee, J.R., Oliver, M., Zevin, J.D., Hoeft, F., Carreiras, M., Tzeng, O.J.L., Pugh, K.R., Frost, R., 2015. Universal brain signature of proficient reading: Evidence from four contrasting languages. Proc Natl Acad Sci U S A 112, 15510–15515. https://doi.org/10.1073/PNAS.1509321112/-/DCSUPPLEMENTAL

Salmelin, R., Hari, R., Lounasmaa, O. v., Sams, M., 1994. Dynamics of brain activation during picture naming. Nature 1994 368:6470 368, 463–465. https://doi.org/10.1038/368463a0

Sulpizio, S., del Maschio, N., del Mauro, G., Fedeli, D., Abutalebi, J., 2020. Bilingualism as a gradient measure modulates functional connectivity of language and control networks. Neuroimage 205, 116306. https://doi.org/10.1016/J.NEUROIMAGE.2019.116306

Swaab, T.Y., Ledoux, K., Camblin, C.C., Boudewyn, M.A., 2012. Language-Related ERP Components. The Oxford Handbook of Event-Related Potential Components. https://doi.org/10.1093/OXFORDHB/9780195374148.013.0197

Taulu, S., Simola, J., 2006. Spatiotemporal signal space separation method for rejecting nearby interference in MEG measurements. Phys Med Biol 51, 1759–1768. https://doi.org/10.1088/0031-9155/51/7/008

Thierry, G., Yan, J.W., 2007. Brain potentials reveal unconscious translation during foreign-language comprehension. Proc Natl Acad Sci U S A 104, 12530–12535. https://doi.org/10.1073/PNAS.0609927104/ASSET/21DED4E2-6087-49BC-B182-B65BF77AD05D/ASSETS/GRAPHIC/ZPQ0300771100004.JPEG

Treder, M.S., 2020. MVPA-Light: A Classification and Regression Toolbox for Multi-Dimensional Data. Front Neurosci 14. https://doi.org/10.3389/fnins.2020.00289

Xu, M., Baldauf, D., Chang, C.Q., Desimone, R., Tan, L.H., 2017. Distinct distributed patterns of neural activity are associated with two languages in the bilingual brain. Sci Adv 3. https://doi.org/10.1126/SCIADV.1603309