
This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.

van Vliet, Marijn; Van Hulle, Marc; Salmelin, Riitta

Exploring the Organization of Semantic Memory through Unsupervised Analysis of Event-related Potentials

Published in:
Journal of Cognitive Neuroscience

DOI:
[10.1162/jocn_a_01211](https://doi.org/10.1162/jocn_a_01211)

Published: 01/03/2018

Document Version
Publisher's PDF, also known as Version of record

Please cite the original version:
van Vliet, M., Van Hulle, M., & Salmelin, R. (2018). Exploring the Organization of Semantic Memory through Unsupervised Analysis of Event-related Potentials. *Journal of Cognitive Neuroscience*, 30(3), 381-392.
https://doi.org/10.1162/jocn_a_01211

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

Exploring the Organization of Semantic Memory through Unsupervised Analysis of Event-related Potentials

Marijn van Vliet^{1,2}, Marc M. Van Hulle¹, and Riitta Salmelin²

Abstract

■ Modern multivariate methods have enabled the application of unsupervised techniques to analyze neurophysiological data without strict adherence to predefined experimental conditions. We demonstrate a multivariate method that leverages priming effects on the evoked potential to perform hierarchical clustering on a set of word stimuli. The current study focuses on the semantic relationships that play a key role in the organization of our mental lexicon of words and concepts. The N400 component of the event-related potential is considered a reliable neurophysiological response that is indicative of whether accessing one concept facilitates subsequent access to another (i.e., one “primes” the other). To further our understanding of the organization of the human mental lexicon, we propose to

utilize the N400 component to drive a clustering algorithm that can uncover, given a set of words, which particular subsets of words show mutual priming. Such a scheme requires a reliable measurement of the amplitude of the N400 component without averaging across many trials, which was here achieved using a recently developed multivariate analysis method based on beamforming. We validated our method by demonstrating that it can reliably detect, without any prior information about the nature of the stimuli, a well-known feature of the organization of our semantic memory: the distinction between animate and inanimate concepts. These results motivate further application of our method to data-driven exploration of disputed or unknown relationships between stimuli. ■

INTRODUCTION

Semantic priming experiments (McNamara & Holbrook, 2003; Neely, 1991) have revealed that accessing a word in our mental lexicon facilitates future access to semantically related words. Because words usually occur in a logical sequence, this “priming” behavior facilitates the processing of likely continuations of a sentence or story (Neely, 1976) and thereby contributes to our ability to exchange messages with others at high speed.

The semantic priming effect has been helpful for studying the organization of human semantic memory (e.g., Kutas & Federmeier, 2000; Collins & Loftus, 1975). For example, the exact nature of the relationships that causes one word to prime another word continues to be the focus of research (e.g., De Deyne, Navarro, Perfors, & Storms, 2016; Van Petten, 1993). In this article, we demonstrate how unsupervised techniques, such as hierarchical clustering, are a particularly useful tool in this case and develop a new technique to study the organization of semantic memory based on a neural correlate of the semantic priming effect.

The boost in signal-to-noise ratio (SNR) provided by multivariate data analysis (Norman, Polyn, Detre, & Haxby, 2006; Friston et al., 1996) enables an exciting paradigm shift in how new insights may be obtained from neurophysiological data. When the SNR is high enough, a re-

searcher can approach the data analysis in an unsupervised manner, instead of labeling data according to some predetermined division (e.g., words vs. pseudowords or tools vs. vegetables). Multivariate analysis reduces the need for averaging across trials, thus facilitating the generation of sufficiently many data points for learning the underlying structure in the data distribution, for example, via clustering techniques (Jain, Murty, & Flynn, 1999). This allows for a data-driven approach to complement theoretical work.

In the application of clustering techniques, the key component to consider is the (dis)similarity score employed by the algorithm. This score is a measure of the distance between two items and is used by the clustering algorithm to determine which items to group together in a cluster. Hence, the effectiveness and validity of clustering techniques in neuroscience depend a great deal on how the measured brain activity is translated into a similarity score.

In the context of semantic relationships, the similarity score corresponds to the concept of semantic distance (Rips, Shoben, & Smith, 1973). Such distance metrics are traditionally based on behavioral data, such as the co-occurrence of words in a large text corpus (Jones, Willits, & Dennis, 2015), the degree of overlap of semantic features (De Deyne et al., 2008; McRae et al., 2005; Hutchison, 2003), or the forward association strength (FAS) score, which is produced by performing an association study where participants, presented with a target word, are asked to write down which words come to

¹KU Leuven, ²Aalto University

mind (De Deyne, Navarro, & Storms, 2013; Nelson, McEvoy, & Schreiber, 2004). In this study, we develop a semantic distance metric that is based solely on a neurophysiological response.

Previous studies that have developed semantic distance metrics from brain activity did so by showing that concepts belonging to the same natural semantic category (e.g., tools, animals) produce similar brain activity. For example, fMRI studies have shown that stimuli from the same semantic category generate similar BOLD activity patterns (Huth, De Heer, Griffiths, Theunissen, & Jack, 2016; Huth, Nishimoto, Vu, & Gallant, 2012; Gerlach, 2007), and EEG and MEG studies have shown that they produce similar spatiotemporal time courses (Chan, Halgren, Marinkovic, & Cash, 2011; Simanova, van Gerven, Oostenveld, & Hagoort, 2010). However, although some semantic categories may activate unique brain activity patterns, there is currently no consensus that this should be the case for all categories (Pulvermüller, 2013) or, for that matter, other types of relationships that are important to the semantic systems in our brain. In this study, we explore an alternative route to obtain a semantic distance metric that is more closely tied to semantic priming.

The distance metric employed in this study is based on a component of the ERP as recorded through EEG, which has been shown to be reliably modulated by semantic priming. By contrasting different levels of priming, an effect can be seen that reaches its maximum around 400 msec post stimulus onset, and the component was hence named the N400 (Kutas & Federmeier, 2011; Kutas & Hillyard, 1984). Since its discovery, relative changes in the amplitude of the N400 component have been shown to correlate well with various behavioral metrics of the strength of the semantic relationship between words, such as word co-occurrence (Van Petten, 2014), FAS (van Vliet et al., 2016; Luka & Van Petten, 2014), and semantic feature overlap (Koivisto & Revonsuo, 2001).

In this study, we demonstrate how to find semantic clusters for a given set of words by measuring the amplitude of the N400 component that was evoked in a semantic priming experiment. Because the semantic priming effect and its relation to the N400 component have been thoroughly studied, the metric and the clustering result it produces are straightforward to interpret.

EEG was recorded while all pairwise combinations of the stimuli, a set of 14 written words, were presented sequentially to the participants. For the second word of each word pair (the target), the amplitude of the N400 component of the evoked EEG response was estimated using a linearly constrained minimum variance (LCMV) beamformer (Van Veen, Van Drongelen, Yuchtman, & Suzuki, 1997), modified to be suitable for ERP analysis (Treder, Porbadnigk, Shahbazi Avarvand, Müller, & Blankertz, 2016; van Vliet et al., 2016; Wittevrongel & Van Hulle, 2016). This approach breaks down the problem of finding proper weights into two steps. The first step is to

construct a template of the desired signal, in this case the spatial and temporal shapes of the N400, based on a traditional ERP analysis consisting of averaging many epochs across many participants. A novelty here is that, instead of doing this on the data obtained in the current study, we used the recordings of a previous semantic priming study (van Vliet et al., 2014). The second step is to obtain the set of weights that isolates this signal from the rest of the EEG, which entails estimating the inverse covariance matrix of the recording currently under consideration. The advantage of this approach is that it leverages a previous study for knowledge about the signal of interest, so predefined experimental conditions are not required for the target recording, that is, indicating beforehand which trials are assumed to have high and low N400 amplitudes.

The N400 amplitudes, as estimated by the beamformer filter, formed the elements of a word-to-word distance matrix that served as input to a hierarchical clustering algorithm, with the aim to discover clusters of semantically related words. Because the main focus of this study is to explore if such a scheme can work, the chosen stimuli in this study were either animals or furniture items, thus items that most semantic theories place in separate clusters (Martin, 2007). The validity of the method was assessed by determining whether the clustering algorithm reveals these clusters.

Importantly, although the stimuli in this study were designed with a clear dichotomy, the method will be agnostic to this fact. Accordingly, the proposed method should also be suitable for exploring data sets where the proper clustering is ambiguous or disputed. Furthermore, because of the unsupervised nature of the method, additional subclusters may also be revealed that were not an intentional part of the experimental design.

METHODS

The study was performed with 19 participants. The data of two participants were discarded because of poor sensor contact quality, and the data of one participant were discarded because of excessive eye blinks. Of the remaining 16 participants, 10 were male and 6 were female, with an age range of 20–58 years (mean = 38 years, $SD = 11$ years); all but one were right-handed; and six were native speakers of Walloon-French and the other 10 were native speakers of Flemish-Dutch.

This study was performed at KU Leuven, and ethical approval was obtained from its university hospital's medical ethics committee. All participants were unpaid volunteers who signed an informed consent form before the experiment.

Stimuli and Experimental Procedure

Word pairs were formed by using all possible prime–target combinations (182) of the 14 words listed in Table 1. The

Table 1. Words Used in the Unsupervised Clustering Study

<i>Dutch</i>	<i>French</i>	<i>English</i>
bed	lit	bed
bureau	bureau	desk
deur	porte	door
giraf	girafe	giraffe
kast	placard	closet
leeuw	lion	lion
neushoorn	rhinocéros	rhinoceros
nijlpaard	hippopotame	hippopotamus
olifant	éléphant	elephant
stoel	chaise	chair
tafel	table	table
tijger	tigre	tiger
zebra	zèbre	zebra
zetel	canapé	couch

The words were displayed in French or in Dutch, according to each participant's native language. The English translation is only for the sake of exposition and was not displayed to the participants. The stimuli consisted of all possible pairwise combinations of these words.

list contains category exemplars for African animals and common furniture items. The stimuli differ in length and frequency of usage, which are normally controlled for in linguistic experiments. However, our method is mostly insensitive to the influences of such word-specific properties, as will be further argued in the Discussion section. The stimuli were presented in the native language of the participant (Flemish-Dutch or Walloon-French). All possible word pairs were presented once, which means that each individual word was presented 26 times: 13 times as prime and 13 times as target. A word was never paired with itself (e.g., the pairs CHAIR–CHAIR or LION–LION were not included), which means there were altogether 84 (i.e., $2 \times 7 \times 6$) “within-category” pairs and 98 (i.e., $2 \times 7 \times 7$) “between-category” pairs.

Participants were seated in an upright position approximately 1 m from a computer screen. The hand used to give the button response rested upon a table with the index and middle fingers on the mouse buttons. A trial consisted of the sequential presentation of a single word pair. The first word of the word pair (the prime) was presented for 200 msec; and the second word (the target), for 1000 msec, with a SOA of 500 msec, after which a question mark appeared prompting a response.

Following the advice of Renoult and Debrulle (2011) for obtaining a semantic priming effect even when stimuli are shown multiple times during the experiment, the participants were asked to determine whether the prime and target words belonged to the same semantic category by pressing one of two mouse buttons. The mapping of the

yes/no response to the mouse buttons and the hand used to operate the mouse were counterbalanced independently across participants.

Data Recording and Preprocessing

EEG was recorded continuously using 32 active electrodes (extended 10–20 system) with a BioSemi Active II System (BioSemi, Amsterdam, The Netherlands), having a fifth-order frequency filter with a pass band of 0.16–100 Hz, and sampled at 2048 Hz. Two additional electrodes were placed on both mastoids, and their average signal was used as a reference for the other sensors. Furthermore, four additional electrodes were placed on the outer canthi of the eyes and above and below the left eye to record horizontal and vertical EOG.

The EEG and EOG signals were further bandpass-filtered offline between 0.3 and 30 Hz by a fourth-order zero-phase infinite impulse response filter to attenuate large drifts and irrelevant high-frequency noise. Electrodes with insufficient signal quality were detected based on visual inspection of the raw data and replaced by a virtual channel using spherical interpolation of the remaining electrodes (Perrin, Pernier, Bertrand, & Echallier, 1989). On average, 1.25 of 32 channels were replaced, with a maximum of four in one participant. The EOG signal was used to attenuate eye artifacts from the EEG signal using the aligned-artifact average regression method described in Croft and Barry (2000). Individual trials were obtained by cutting the continuous signal from 0.1 sec before the onset of each target stimulus to 1.0 sec after. All trials were used in the analysis. Baseline correction was performed using the average voltage in the 0.1-sec interval before the stimulus onset as baseline value. Finally, because any high-frequency content was removed by the bandpass filter, the signal was down-sampled to 50 Hz without losing much information. This step was included to reduce the dimensionality of the data matrices, which improves the numerical stability of the beamformer filter.

Beamformer Filter

After preprocessing the EEG signals, multivariate analysis was performed using a spatiotemporal LCMV beamformer filter. The filter takes a weighted sum of the data points from all EEG channels and all samples within an epoch. The result of this summation represents the estimated amplitude of the N400 component of the ERP within that epoch. For an in-depth explanation and implementation details of the method, see van Vliet et al. (2016).

The beamformer approach consists of two steps. The first step is to construct a template of the desired signal: in this case, the spatial and temporal shapes of the N400. The second step is to obtain the set of weights that isolates this signal from the rest of the EEG, which entails

estimating the inverse signal covariance matrix of the recording currently under consideration (the target recording).

To obtain a template of the N400 component and fine-tune the beamformer filter, we reused data that were collected in a previous semantic priming study (van Vliet et al., 2014). In that study, 10 native speakers of Flemish-Dutch were shown 800 word pairs with varying FAS, as determined from an association norm database compiled by De Deyne and Storms (De Deyne et al., 2013; De Deyne & Storms, 2008), covering the whole range of completely unrelated to the strongest related words in the database. The experimental procedure, recording setup, and data processing were identical to those used for the unsupervised clustering study as described above, with the exception that the responding hand was always the right hand and the mapping of yes/no responses to the mouse buttons was not counterbalanced. See van Vliet et al. (2014) for further details about the study.

The data of the previous study were reanalyzed by performing linear regression, using the logarithm of the FAS of the stimuli as predictor and the EEG as response variable, resulting in what Smith and Kutas (2015) refer to as a “slope” ERP. This slope ERP is a generalization of the difference wave and can be thought of as “the part of the ERP that changes when the FAS of the stimulus changes.” Next, we determined the time point when the global field power of the slope ERP reached its maximum, which was at 430 msec after stimulus onset. The distribution of the slope ERP across the sensors at that time point was taken to be the spatial pattern for the N400 component (Figure 1, left). The temporal template was constructed by using the spatial template to create a spatial LCMV beamformer (van Vliet et al., 2016), the output of which represents an estimation of the summed activity at the cortical source locations of the N400 (Figure 1, right, gray line). This time course was further refined by multiplying it with a Gaussian kernel

($\mu = 400$ msec, $\sigma = 10$ msec), which has the effect of limiting the nonzero values to a window of interest centered on the peak amplitude of the N400 (Figure 1, right, black line). Finally, the full spatiotemporal template was obtained by taking the outer product of the spatial and temporal templates.

To compute the filter weights that will isolate the signal component described by the template from the rest of the signal, the template must be multiplied with the spatiotemporal covariance matrix $\hat{\Sigma}$ of the target data. This matrix can be readily computed from the data of the current study, because it does not require contrasting different experimental conditions. Because of the high dimensionality of this matrix, it is recommended to employ heavy shrinkage during its estimation. In this study, we employed shrinkage toward the diagonal:

$$\hat{\Sigma} = \mathbf{X}^T \mathbf{X} \quad (1)$$

$$\hat{\Sigma} = (1 - a)\hat{\Sigma} + a\frac{\text{Tr}\hat{\Sigma}}{n}\mathbf{I} \quad (2)$$

where \mathbf{X} is a matrix where each row corresponds to one of the n epochs and contains a flattened version (i.e., all elements are placed on a single row) of the Channels \times Samples matrix. $\hat{\Sigma}$ is the empirical covariance matrix, “Tr $\hat{\Sigma}$ ” means the sum of the diagonal elements of $\hat{\Sigma}$, and \mathbf{I} is an identity matrix. The value of the shrinkage parameter a was optimized by designing beamformer filters with different values for a and applying them to the data of the previous study. The optimization criterion was to maximize the correlation between the output of the filter and the FAS of the stimuli that were used in that study. This resulted in an optimal a value of 0.9, which is the value we subsequently used to design the filter for this study.

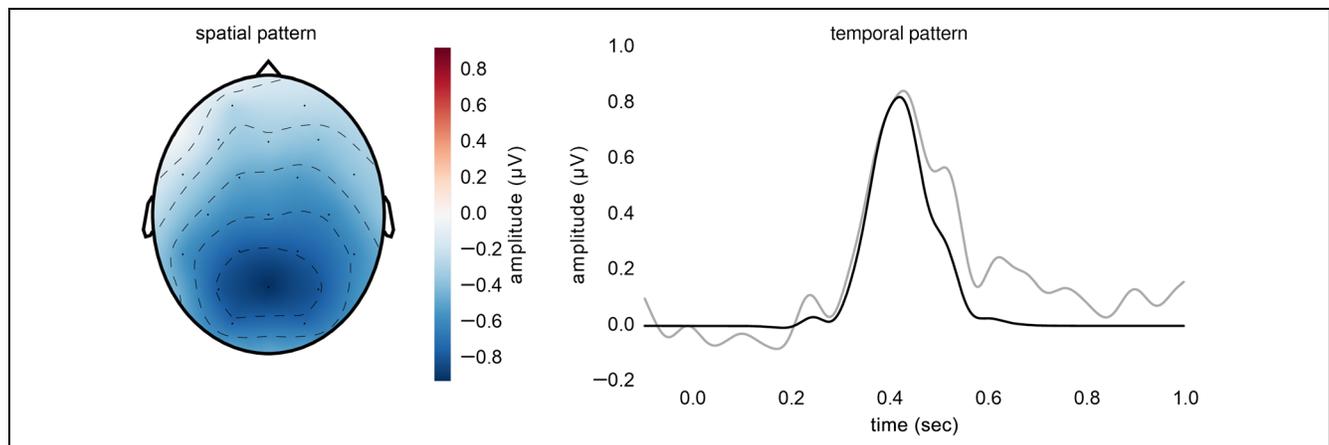


Figure 1. Spatial (left) and temporal (right) patterns of the N400 ERP component, evoked in a semantic priming experiment. In the figure depicting the temporal template, the gray line represents the result of the spatial beamformer, and the black line represents the result after multiplying with a Gaussian kernel.

Given the covariance matrix and the template of the N400 component, the estimation of the amplitude of this component for a given epoch is

$$\mathbf{w} = \frac{\boldsymbol{\Sigma}^{-1} \mathbf{a}}{\mathbf{a}^\top \boldsymbol{\Sigma}^{-1} \mathbf{a}} \quad (3)$$

$$\hat{y} = \mathbf{w}^\top \mathbf{x} \quad (4)$$

where $\boldsymbol{\Sigma}^{-1}$ is the inverse of the covariance matrix, \mathbf{a} is a flattened version of the Channels \times Samples matrix containing the N400 template, and \mathbf{x} is the flattened version of the Channels \times Samples matrix containing the EEG epoch.

Hierarchical Clustering

The amplitude of the N400 ERP component \hat{y} as quantified by the spatiotemporal LCMV beamformer filter was further processed to obtain a suitable metric for the semantic distance between the prime and target stimuli. First, for each participant, z scoring was performed across the \hat{y} 's to equalize the scaling. Then, a distance metric was derived from the z -scored N400 amplitude estimates.

As distance metric, we chose the relative change in the amplitude of the N400 component, as a target word is paired with different prime words. Given the set S of all words used in the study (here, we regard the Dutch and French translations as the same word), the amplitude of the N400 component evoked by the target word of word pair $a \in S$ and $b \in S$ is denoted $N_{400}(a, b)$ and the distance between the words, denoted $d(a, b)$, is quantified as

$$d(a, b) = N_{400}(a, b) - \frac{1}{n} \sum_{w \in S} N_{400}(w, b) \quad (5)$$

where n is the total number of words in S . Because a word was never paired with itself during our study, an actual measurement of the amplitude of the N400 component is missing for this case. We therefore assume $d(b, b) = 0$. The resulting distance values were subsequently organized in a Words \times Words matrix \mathbf{D} .

Matrix \mathbf{D} contains, for each pairwise combination of two words in Table 1, two responses for each participant: one response for the case where the first word was used as prime and the second word as target and another response for the reversed case. Because the hierarchical clustering algorithm operates on geometric distance, which is symmetric and positive, the distance matrix \mathbf{D} should be symmetric and positive as well. This was achieved by averaging \mathbf{D} with its transposed form and subtracting the lowest value:

$$\mathbf{D}_{\text{sym}} = \frac{\mathbf{D} + \mathbf{D}^\top}{2} \quad (6)$$

$$\mathbf{D}_{\text{pos}} = \mathbf{D}_{\text{sym}} - \min \mathbf{D}_{\text{sym}} \quad (7)$$

The final matrix, used as input for the hierarchical clustering algorithm, was obtained by averaging the distance

matrices across participants. Because the N400 amplitude estimates are noisy, it is beneficial to base the distance between two clusters on as many measurements as possible. Therefore, average linkage (also known as unweighted pair group method with arithmetic mean linkage) was chosen as the clustering algorithm (Jain et al., 1999). It determines the distance between two clusters by considering the average distance between all items in the clusters (Sokal & Michener, 1958). We present the output of the clustering algorithm in the form of a dendrogram.

Psycholinguistic Variables

We investigated the extent to which the distance metric (Equation 5) is influenced by properties of the prime and target words that are independent of their semantic relationship. Estimations of word frequencies on a log scale (denoted "log freq") were taken from the SUBTLEX-NL project (Keuleers, Brysbaert, & New, 2010) for Dutch words and from the French Lexicon project (Ferrand et al., 2010) for French words. Age of acquisition (AoA) estimates were provided by Brysbaert, Van Wijnendaele, and De Deyne (2000) for Dutch words and Rijn et al. (2008) for French words. Finally, the mean family-wise error of participants performing a lexical decision task for a word, presented in isolation, was determined in the large-scale Dutch and French lexicon projects (Ferrand et al., 2010; Keuleers, Diependaele, & Brysbaert, 2010).

Statistics

At each "node" in the dendrogram, where two subclusters were joined to form a new cluster, a statistical test was performed to provide an indication of the reliability of the distinction presented by the two subclusters. To this end, a linear mixed-effects (LME) model was used to analyze the difference between the distance values (Equation 5) for within-cluster word pairs versus between-cluster word pairs. Note that this test can only be performed if both clusters consist of at least two words; otherwise, there are no within-cluster word pairs. The distance values were used as the dependent variable, with a dummy encoding of the labels "within-cluster" = 1 versus "between-cluster" = 0 as fixed effects. Because the model needs to generalize beyond the participants included in the study, participants were modeled as a random effect (random slopes and random intercepts). However, because the model does not need to generalize beyond the words in the clusters, words were not included as a random effect. The model was fitted using restricted maximum likelihood, with degrees of freedom and the resulting p values estimated using Satterthwaite's approximation (Satterthwaite, 1946). To control for the FWE rate, the p values were Bonferroni corrected by multiplying them by the number of tests performed. When this resulted in $p > 1$, we report $p = 1$.

When testing the effect of a psycholinguistic variable on the amplitude of the N400 component or the distance metric, the fact that the values for Dutch and French words originate from different norm studies must be taken into account. Therefore, in these cases, the LME model was used with both participants and language (Dutch or French) as random effects (random slopes and random intercepts), which means that the model will use different regression weights for each language.

Software

Stimulus presentation was performed using MATLAB (The MathWorks, Natick, MA) in combination with the Psychophysics toolbox (Brainard, 1997). Data analysis was performed using Python in combination with the Psychic, NumPy, and SciPy packages (Oliphant, 2007). Covariance estimation with shrinkage was performed using the *Scikit-learn* package (Pedregosa et al., 2012). Plots were created using the *Matplotlib* package (Hunter, 2007). Statistical analysis was performed using R (R Core Development Team, 2015) in combination with the *LME4* (Bates, Maechler, Bolker, & Walker, 2015) and *lmerTest* (Kuznetsova, Brockhoff, & Christensen, 2015) packages.

Data and Code Availability

A software implementation of the N400 template estimation procedures and spatiotemporal LCMV beamformer can be found at github.com/wmvanvliet/ERP-beamformer. The raw EEG data and the N400 template constructed from the data collected in van Vliet et al. (2014) can be acquired upon request from the corresponding author. In addition, for use in future studies that employ a methodology that is similar to that in the current work, a template that is based on the data collected during the current study is also available upon request. The processed data and source code pertaining to the subsequent computation of the distance matrix, hierarchical clustering, and all statistics performed in this study are available at github.com/wmvanvliet/jocn2017.

RESULTS

As expected, the button responses collected during the experiment showed that the participants very consistently marked word pairs as “related” and “unrelated” according to a classification of animal versus furniture item. Furniture–furniture pairs received a “related” response 89.0% of the time; animal–animal pairs, 93.6%; furniture–animal pairs, 1.1%; and animal–furniture pairs, 0.6%. It is likely that, after a few trials, the participants noticed the pattern and started to perform a classification task (do the two words belong to the same animate/inanimate category?) instead of a judgment of association task (are the two words associatively related?). The distance matrix, based on estimations of the amplitude of

the N400 component (Figure 2), also shows as overall trend a dichotomy between animal versus furniture items. Although single-item measurements can be unreliable (e.g., CHAIR–HIPPO shows up as relatively related, which is probably a measurement error), hierarchical clustering can reveal the underlying patterns.

The dendrogram produced by the hierarchical clustering algorithm (Figure 3) has as the topmost two clusters all animal stimuli versus all furniture stimuli. The fact that these clusters could be reliably reconstructed shows that the multivariate analysis of the EEG data yielded a measurement with a high enough SNR to perform this type of unsupervised clustering. As these clusters are themselves divided into subclusters, the results are based on less data and therefore less reliable. Statistical tests at each “node” of the dendrogram are an indication of this reliability and show whether there is a significant difference in N400 amplitude between within-cluster and between-cluster trials.

The only explicit distinction in the experimental design was a distinction between animals and furniture items. However, the dendrogram suggests that there may be a dichotomy in the chosen furniture stimuli (DESK/BED/CLOSET vs. DOOR/TABLE/CHAIR/COUCH). The cluster containing the animal stimuli did not show any reliable further subclustering.

Given the experimental design, it is likely that the amplitude of the N400 component is influenced by conscious decision-making processes. To exclude the possibility that the N400 effect solely reflects the upcoming behavioral response, we reanalyzed the [DESK, BED, CLOSET] versus [DOOR, TABLE, CHAIR, COUCH] subclusters. These subclusters only contain furniture–furniture

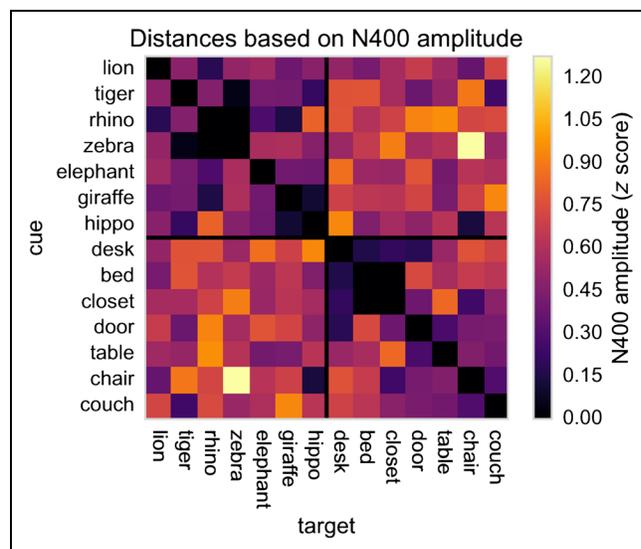


Figure 2. Distance matrix based on the amplitude of the N400 component, averaged across participants. The order of the words mirrors the order in which they appear in the dendrogram (Figure 3). Black lines mark the boundary between the top clusters in the dendrogram.

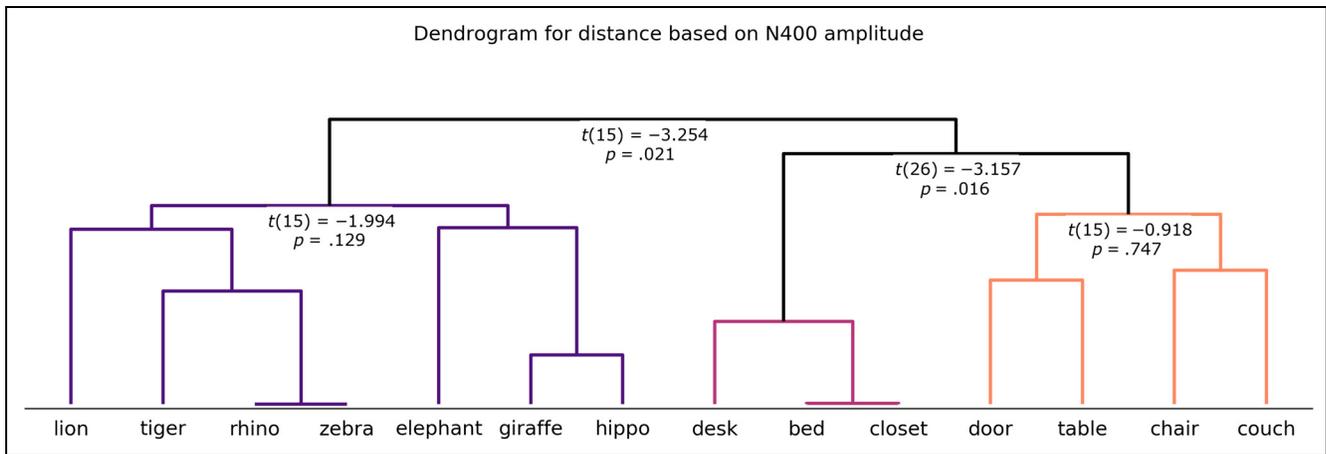


Figure 3. Dendrogram resulting from the hierarchical clustering algorithm applied to the distance matrix based on the amplitude of the N400 component. Statistical tests were performed to test for differences in N400 amplitude in response to between-(sub)cluster versus within-(sub)cluster word pairs. Clusters that could be significantly distinguished from each other at $p < .05$ have been assigned different colors. The reported p values are Bonferroni-corrected.

pairs. A total of 672 trials (7×6 word pairs \times 16 participants) are available for this analysis, in which a “words are related” button response was given 598 times and a “words are unrelated” response was given 74 times. This reflects the tendency of the participants to make the relatedness judgments based on semantic category membership. Discarding the 74 trials with an “unrelated” response, there remained 254 “within-subcluster” pairs and 344 “between-subcluster” pairs for which the behavioral response was the same. For these pairs, a significant difference in the distance values persisted for within- versus between-cluster word pairs, $t(12.78) = -2.65, p = .020$. This suggests that the distance metric we employ in this study is not driven solely by the behavioral responses.

The grand-averaged ERPs, obtained by assigning the labels “within-cluster” and “between-cluster” based on the topmost clustering in the dendrogram, are presented in Figure 4. Two components can be observed in the ERP, with the first being the N400 component with a posterior distribution, present during both the within- and between-cluster conditions. The second component is only observed in the between-cluster condition and has a more frontal distribution, which can be possibly classified as a P600 component, commonly observed when stimuli are repeated (Van Strien, Hagenbeek, Stam, Rombouts, & Barkhof, 2005).

There are many factors that influence the amplitude of the N400 component. In our study, we are only interested in capturing effects that are due to the relationship between the prime and target words. Therefore, we wish to ensure that effects that cannot be attributed to this relationship do not affect our results. Table 2 shows the results of statistical tests for various psycholinguistic variables on both the amplitude of the N400 component and the distance metric that was derived from this amplitude (Equation 5). In the experimental paradigm used by

our method, the psycholinguistic variables that were tested only had a small effect on the amplitude of the N400 component, none of which passed the significance threshold. The distance metric employed in this study corresponds to the change in amplitude of the N400 component as the target word is presented in combination with different prime words, relative to the mean N400 amplitude for the word. It is therefore insensitive to effects that pertain to the target word alone.

DISCUSSION

The main result is that the distinction between animals and furniture items could be reliably extracted, based purely on EEG responses. This could be done without supplying any information about the nature of the clusters to the algorithm (i.e., no experimental conditions, no information about the clusters having an equal number of members), thus giving confidence that the method can produce trustworthy results for data sets where the optimal clustering is not known beforehand, provided that the distance (in our case, semantic distance) between the clusters is large enough.

It is worth noting that, although p values are provided in the dendrogram, the clustering result goes beyond the statistical statement that these p values make. Although there are many possible ways to cluster the stimuli in such a manner that there is a significant difference in N400 amplitude between the within- and between-cluster pairs, the dendrogram reveals, of all possible ways to arrange the items, the strongest hierarchical clustering (according to the linkage metric). When this clustering corresponds to the clustering predicted by a hypothesis (as it does in this case) and the accompanying p value is small, the evidence that the hypothesis is correct is much stronger than is provided by a p value alone.

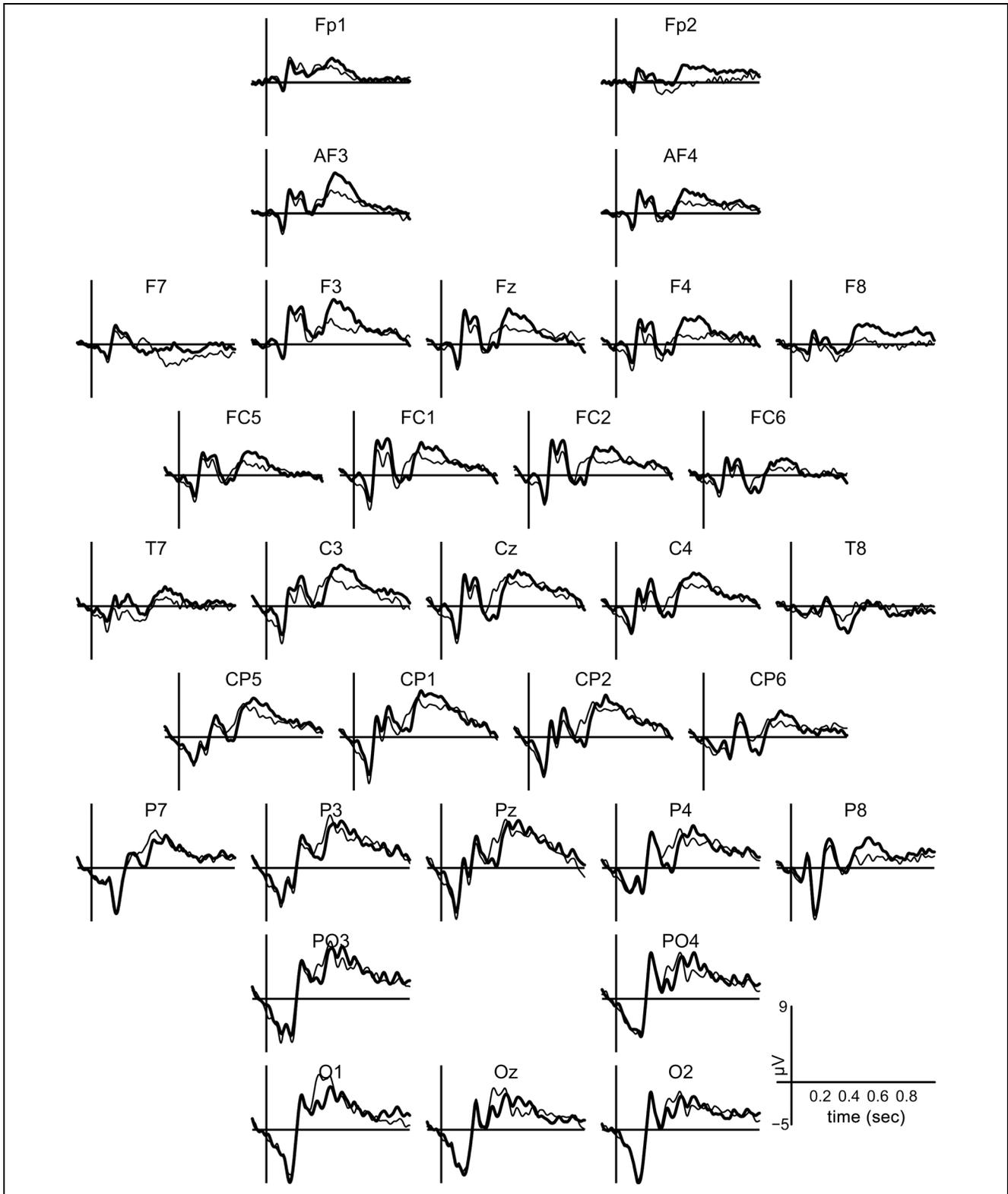


Figure 4. Grand-averaged ERPs in response to within-cluster (thin line) and between-cluster (thick line) word pairs, corresponding to the topmost clusters in the dendrogram: animals versus furniture items.

We employed a semantic distance metric that is based on the amplitude of the N400 component of the ERP, evoked using a semantic priming paradigm. This metric may capture different semantic relationships than earlier

work that analyzed the full spatiotemporal activity pattern evoked by single words (Huth et al., 2016; Chan et al., 2011; Simanova et al., 2010; Gerlach, 2007). Furthermore, because the proposed metric does not require the user to

Table 2. Effect of Various Psycholinguistic Variables on the Amplitude of the N400 Component and the Distance Metric Derived From This Amplitude

Word ^a	Variable ^b	N400 Amplitude				Distance Metric			
		Effect Size	<i>t</i>	Estimated <i>df</i>	<i>p</i>	Effect Size	<i>t</i>	Estimated <i>df</i>	<i>p</i>
Prime	Length	0.0169	0.280	2909.95	.779	0.0244	0.419	2839.53	.675
	Log freq	-0.0190	-0.437	2909.97	.662	-0.0199	-0.476	2910.00	.634
	AoA	-0.0155	-0.154	2363.00	.877	-0.0100	-0.103	16.11	.919
	RT	-0.1687	-0.266	16.00	.397	-0.130	-0.166	2.89	.879
Target	Length	0.0970	1.577	19.06	.131	0.000	0.000	2909.88	1.000
	Log freq	-0.0130	-0.285	1.45	.811	0.000	0.000	2910.00	1.000
	AoA	0.0717	0.714	2361.92	.475	0.000	0.000	2364.00	1.000
	RT	0.4729	0.671	16.00	.744	0.000	0.000	14.50	1.000

Note that the psycholinguistic variables for the target word have no discernible effect on the distance metric.

^aThis indicates whether the variable pertains to the first (prime) or second (target) word of the word pair.

^bSee the Psycholinguistic Variables section for a description of the variables.

distinguish brain activity between different spatial locations, the measurement can also be performed using techniques that have a relatively poor spatial resolution, such as EEG.

It is likely that our method could be applicable to study ERP components other than the N400, when the amplitude of such components is affected by the relationship between stimuli. Examples include the P300, mismatch negativity and the N2 component, all of which have been used to study aspects of memory (Folstein & Van Petten, 2008; Novak, Ritter, Vaughan, & Witznitzer, 1990; Johnson & Donchin, 1980).

Considerations Regarding the Interpretation of the Results

The method requires the detection of differences in N400 amplitude when a target word is presented in combination with different prime words. How large these differences need to be in order for clusters to be differentiated depends on the SNR that can be achieved in estimating the amplitudes. In this study, we employed a spatiotemporal LCMV beamformer, which has been shown to produce more reliable estimates of the N400 amplitude than more traditional approaches, such as measuring the mean voltage in a fixed time window (van Vliet et al., 2016).

Because stimuli need to be repeated to construct a full word-to-word distance matrix, the N400 effect is degraded somewhat because of semantic facilitation through STM (e.g., due to the old/new effect; Rugg & Curran, 2007). Nevertheless, our results reproduce the earlier finding that the N400 effect persists even when the stimuli are repeated (Renoult & Debruille, 2011; Debruille & Renoult, 2009), as long as the target word cannot be predicted

from the prime word and an explicit task is given to the participant (Renoult, Wang, Mortimer, & Debruille, 2012).

It is likely that there are small differences between the N400 template and the actual N400 observed in this study, due to the repetition of stimuli, which can cause small shifts in the timing of the component (Renoult, Wang, Calcagno, Prévost, & Debruille, 2012). Furthermore, the earlier study that provided the N400 template for the current study (van Vliet et al., 2014) explores some possible motor-related and P300 confounds when using an explicit decision task.

The fact that good results were obtained using a template based on an independent data set (Figure 1) provides some validation that the component reaching a maximum around 400 msec (Figure 4) is similar to the N400 component observed in classical priming experiments. The ability of the beamformer algorithm to accurately estimate N400 amplitudes depends greatly on the accuracy of the supplied template (Treder et al., 2016; van Vliet et al., 2016). If the component evoked in this study would deviate too much from the template (in either spatial distribution or timing), it would fall outside the passband of the filter.

Considerations Regarding the Experimental Design

In this study, our primary research question is whether the amplitude of the N400 component could be estimated with a high enough SNR in order for the unsupervised clustering to produce the expected result. To this end, the experimental paradigm was chosen to maximize the measured N400 effect. For example, a relatively long SOA of 500 msec was chosen, and no masking of the prime stimulus was performed.

Whereas the current study focuses on a well-known animate–inanimate dichotomy to provide an initial validation of the method, further studies are needed to explore the sensitivity of the method to more intricate aspects of memory organization. For example, one may attempt to disentangle the influence of conscious processes on the N400-based distance metric. To this end, a very short SOA may be used (Hill, Strube, Roesch-Ely, & Weisbrod, 2002), as well as masking of the prime word (Deacon, Hewitt, Yang, & Nagata, 2000). In addition, the task for the participants may be modified such that they no longer perform a conscious categorization task while still requiring deep processing of the stimuli (e.g., Heyman, De Deyne, Hutchinson, & Storms, 2015), to reduce conscious decision-making effects and confounds of the P3 component (van Vliet et al., 2014; Roehm, Bornkessel-Schlesewsky, Rösler, & Schlewsky, 2007).

The construction of a full word-to-word distance matrix of n items requires the presentation of $n^2 - n$ stimuli, hence the number of items that can be included in the analysis is restricted. Because the method can more reliably reveal patterns in semantic relationships when there are clearly distinguishable clusters in the stimulus set, the items that are included should be carefully chosen.

An advantage of the distance metric we used in this study is that the outcome is quite robust against word-specific properties, thus possible confounding factors such as length, frequency of usage, AoA, and so forth. This is achieved by setting the mean across all the prime–target pairs, where the item was used as target, to zero (Equation 5). The remaining values only reflect the change in the N400 response when a word is preceded by different prime words. Furthermore, because the average linkage algorithm determines the distance between two clusters by computing the ratio between the mean within-cluster distance and the mean distance to every other cluster, the word pairs relevant to the computation always cover the complete set of words. Specifically, because the distance matrix is made symmetric, the choice of cluster to which a word is assigned is influenced by how the N400 amplitude changes when the word is paired with all other words, regardless of whether the word was used as a prime or target. This approach will not eliminate all possible confounding effects, but it leaves the experimenter with considerable freedom in how to select the stimuli for the experiment.

In addition to answering a predefined research question, post hoc analysis of the dendrogram may be used as a starting point for future exploration. Of course, proper consideration must be given to the level at which to “cut” the dendrogram; in this study, we compute p values for each node and cut at $p < .05$. In addition to the top level clusters, we find that the dendrogram also hints at a dichotomy among the selected furniture stimuli. Indeed, strong semantic clusters may well exist within this category of words, for example, based on the room that the furniture pieces are commonly found in.

Although this study does not include enough data to confirm such a hypothesis, the method suggests that this line of inquiry may be fruitful.

Although the proposed method is unsupervised and will always produce some clustering solution, a careful experimental design is needed to ensure that the result is interpretable. We show how measurement of the amplitude of the N400 component may be used to drive a clustering algorithm. Precisely what aspects of semantic memory are reflected in these amplitudes (e.g., Cheyette & Plaut, 2017), and the role of the experimental design therein (e.g., Roehm et al., 2007), is an ongoing debate for which our proposed method may yield new insights.

Conclusion

We have demonstrated a way to employ amplitude measurements of the N400 ERP component as a semantic distance metric between words. To obtain a reliable measurement, a multivariate analysis procedure based on the LCMV beamformer was successfully employed to overcome the low SNR of EEG signals. The resulting distance metric allows for successful application of unsupervised techniques, such as hierarchical clustering, on EEG priming data, to analyze how a chosen set of stimuli cluster together.

Our results illustrate how unsupervised techniques can be leveraged to analyze EEG data without strict adherence to predefined labels. This can be particularly useful when validating theories concerning the organization of memory systems in the brain.

Acknowledgments

M. v. V. was supported by the Interuniversity Attraction Poles Programme–Belgian Science Policy (IUAP P7/11) and is currently supported by a grant from the Aalto Brain Centre. M. M. V. H. is supported by research grants received from the financing program (PFV/10/008), an interdisciplinary research project (IDO/12/007), and an industrial research fund project (IOF/HB/12/021) of the KU Leuven; the Belgian Fund for Scientific Research–Flanders (G088314N and G0A0914N); the Interuniversity Attraction Poles Programme–Belgian Science Policy (IUAP P7/11); the Flemish Regional Ministry of Education (Belgium; GOA 10/019); and the Hercules Foundation (AKUL 043). R. S. is supported by the Academy of Finland (255349, 256459, and 283071; LASTU programme 256887) and the Sigrid Jusélius Foundation.

Reprint requests should be sent to Marijn van Vliet, Department of Neuroscience and Biomedical Engineering, Aalto University, Otakaari 3, 02150 Espoo, Finland, or via e-mail: marijn.vanvliet@aalto.fi, w.m.vanvliet@gmail.com.

REFERENCES

- Bates, D. M., Maechler, M., Bolker, B. M., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10, 433–436.

- Brysbaert, M., Van Wijnendaele, I., & De Deyne, S. (2000). Age-of-acquisition effects in semantic processing tasks. *Acta Psychologica, 104*, 215–226.
- Chan, A. M., Halgren, E., Marinkovic, K., & Cash, S. S. (2011). Decoding word and category-specific spatiotemporal representations from MEG and EEG. *Neuroimage, 54*, 3028–3039.
- Cheyette, S. J., & Plaut, D. C. (2017). Modeling the N400 ERP component as transient semantic over-activation within a neural network model of word comprehension. *Cognition, 162*, 153–166.
- Collins, A. M., & Loftus, E. F. (1975). Spreading-activation theory of semantic memory. *Psychological Review, 82*, 407–428.
- Croft, R. J., & Barry, R. J. (2000). Removal of ocular artifact from the EEG: A review. *Neurophysiologie Clinique, 30*, 5–19.
- De Deyne, S., Navarro, D. J., Perfors, A., & Storms, G. (2016). Structure at every scale: A semantic network account of the similarities between unrelated concepts. *Journal of Experimental Psychology: General, 145*, 1228–1254.
- De Deyne, S., Navarro, D. J., & Storms, G. (2013). Better explanations of lexical and semantic cognition using networks derived from continued rather than single-word associations. *Behavior Research Methods, 45*, 480–498.
- De Deyne, S., & Storms, G. (2008). Word associations: Network and semantic properties. *Behavior Research Methods, 40*, 213–231.
- De Deyne, S., Verheyen, S., Ameel, E., Vanpaemel, W., Dry, M. J., Voorspoels, W., et al. (2008). Exemplar by feature applicability matrices and other Dutch normative data for semantic concepts. *Behavior Research Methods, 40*, 1030–1048.
- Deacon, D., Hewitt, S., Yang, C. M., & 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. *Cognitive Brain Research, 9*, 137–146.
- Debrulle, J. B., & Renoult, L. (2009). Effects of semantic matching and of semantic category on reaction time and N400 that resist numerous repetitions. *Neuropsychologia, 47*, 506–517.
- Ferrand, L., New, B., Brysbaert, M., Keuleers, E., Bonin, P., Méot, A., et al. (2010). The French Lexicon Project: Lexical decision data for 38,840 French words and 38,840 pseudowords. *Behavior Research Methods, 42*, 488–496.
- Folstein, J. R., & Van Petten, C. (2008). Influence of cognitive control and mismatch on the N2 component of the ERP: A review. *Psychophysiology, 45*, 152–170.
- Friston, K. J., Stephan, K. M., Heather, J. D., Frith, C. D., Ioannides, A. A., Liu, L. C., et al. (1996). A multivariate analysis of evoked responses in EEG and MEG data. *Neuroimage, 3*, 167–174.
- Gerlach, C. (2007). A review of functional imaging studies on category specificity. *Journal of Cognitive Neuroscience, 19*, 296–314.
- Heyman, T., De Deyne, S., Hutchison, K. A., & Storms, G. (2015). Using the speeded word fragment completion task to examine semantic priming. *Behavior Research Methods, 47*, 580–606.
- Hill, H., Strube, M., Roesch-Ely, D., & Weisbrod, M. (2002). Automatic vs. controlled processes in semantic priming—Differentiation by event-related potentials. *International Journal of Psychophysiology, 44*, 197–218.
- Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science and Engineering, 9*, 99–104.
- Hutchison, K. A. (2003). Is semantic priming due to association strength or feature overlap? A microanalytic review. *Psychonomic Bulletin & Review, 10*, 785–813.
- Huth, A. G., De Heer, W. A., Griffiths, T. L., Theunissen, F. E., & Jack, L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature, 532*, 453–458.
- Huth, A. G., Nishimoto, S., Vu, A. T., & Gallant, J. L. (2012). A continuous semantic space describes the representation of thousands of object and action categories across the human brain. *Neuron, 76*, 1210–1224.
- Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: A review. *ACM Computing Surveys, 31*, 264–323.
- Johnson, R., & Donchin, E. (1980). P300 and stimulus categorization: Two plus one is not so different from one plus one. *Psychophysiology, 17*, 167–178.
- Jones, M. N., Willits, J., & Dennis, S. (2015). Models of semantic memory. In *Oxford handbook of mathematical and computational psychology* (pp. 232–254). New York: Oxford University Press.
- Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. *Behavior Research Methods, 42*, 643–650.
- Keuleers, E., Diependaele, K., & Brysbaert, M. (2010). Practice effects in large-scale visual word recognition studies: A lexical decision study on 14,000 Dutch mono- and disyllabic words and nonwords. *Frontiers in Psychology, 1*, 174.
- Koivisto, M., & Revonsuo, A. (2001). Cognitive representations underlying the N400 priming effect. *Cognitive Brain Research, 12*, 487–490.
- Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences, 4*, 463–470.
- Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). *Annual Review of Psychology, 62*, 621.
- Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307*, 161–163.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). *lmerTest: Tests for random and fixed effects for linear mixed effect models*. R package, version 2.0-29. Retrieved from <http://cran.r-project.org/package=lmerTest>.
- Luka, B. J., & Van Petten, C. K. (2014). Prospective and retrospective semantic processing: Prediction, time, and relationship strength in event-related potentials. *Brain and Language, 135*, 115–129.
- Martin, A. (2007). The representation of object concepts in the brain. *Annual Review of Psychology, 58*, 25–45.
- McNamara, T. P., & Holbrook, J. B. (2003). Semantic memory and priming. In A. F. Healy & R. W. Proctor (Eds.), *Handbook of psychology: Experimental psychology* (Vol. 4, pp. 447–474). New York: Wiley.
- McRae, K., Cree, G. S., Seidenberg, M. S., McNorgan, C., Siedenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods, 37*, 547–559.
- Neely, J. H. (1976). Semantic priming and retrieval from lexical memory: Evidence for facilitatory and inhibitory processes. *Memory & Cognition, 4*, 648–654.
- Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in visual word recognition* (pp. 264–323). Hillsdale, NJ: Erlbaum.
- Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers, 36*, 402–407.
- Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: Multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences, 10*, 424–430.

- Novak, G. P., Ritter, W., Vaughan, H. G., & Wiznitzer, M. L. (1990). Differentiation of negative event-related potentials in an auditory discrimination task. *Electroencephalography and Clinical Neurophysiology*, *75*, 255–275.
- Oliphant, T. E. (2007). Python for scientific computing. *Computing in Science and Engineering*, *9*, 10–20.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2012). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.
- Perrin, F., Pernier, J., Bertrand, O., & Echallier, J. F. (1989). Spherical splines for scalp potential and current density mapping. *Electroencephalography and Clinical Neurophysiology*, *72*, 184–187.
- Pulvermüller, F. (2013). How neurons make meaning: Brain mechanisms for embodied and abstract-symbolic semantics. *Trends in Cognitive Sciences*, *17*, 458–470.
- R Core Development Team. (2015). R: A language and environment for statistical computing, 3.2.1. Vienna: R Foundation for Statistical Computing. <http://www.R-project.org/>.
- Renoult, L., & Debruille, J. B. (2011). N400-like potentials and reaction times index semantic relations between highly repeated individual words. *Journal of Cognitive Neuroscience*, *23*, 905–922.
- Renoult, L., Wang, X., Calcagno, V., Prévost, M., & Debruille, J. B. (2012). From N400 to N300: Variations in the timing of semantic processing with repetition. *Neuroimage*, *61*, 206–215.
- Renoult, L., Wang, X., Mortimer, J., & Debruille, J. B. (2012). Explicit semantic tasks are necessary to study semantic priming effects with high rates of repetition. *Clinical Neurophysiology*, *123*, 741–754.
- Rijn, V., Moor, D., French, I., Ferrand, L., Bonin, P., Méot, A., et al. (2008). Age-of-acquisition and subjective frequency estimates for all generally known monosyllabic French words and their relation with other psycholinguistic variables. *Behavior Research Methods*, *40*, 1049–1054.
- Rips, L. J., Shoben, E. J., & Smith, E. E. (1973). Semantic distance and the verification of semantic relations. *Journal of Verbal Learning and Verbal Behavior*, *12*, 1–20.
- Roehm, D., Bornkessel-Schlesewsky, I., Rösler, F., & Schlesewsky, M. (2007). To predict or not to predict: Influences of task and strategy on the processing of semantic relations. *Journal of Cognitive Neuroscience*, *19*, 1259–1274.
- Rugg, M. D., & Curran, T. (2007). Event-related potentials and recognition memory. *Trends in Cognitive Sciences*, *11*, 251–257.
- Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics*, *2*, 110–114.
- Simanova, I., van Gerven, M. A. J., Oostenveld, R., & Hagoort, P. (2010). Identifying object categories from event-related EEG: Toward decoding of conceptual representations. *PLoS One*, *5*, e14465.
- Smith, N. J., & Kutas, M. (2015). Regression-based estimation of ERP waveforms: I. The rERP framework. *Psychophysiology*, *52*, 157–168.
- Sokal, R. R., & Michener, C. D. (1958). *A statistical method for evaluating systematic relationships*. University of Kansas Science Bulletin (Vol. 38). <https://doi.org/citeulike-article-id:1327877>.
- Treder, M. S., Porbadnigk, A. K., Shahbazi Avarvand, F., Müller, K.-R., & Blankertz, B. (2016). The LDA beamformer: Optimal estimation of ERP source time series using linear discriminant analysis. *Neuroimage*, *129*, 279–291.
- Van Petten, C. K. (1993). A comparison of lexical and sentence-level context effects in event-related potentials. *Language and Cognitive Processes*, *8*, 485–531.
- Van Petten, C. K. (2014). Examining the N400 semantic context effect item-by-item: Relationship to corpus-based measures of word co-occurrence. *International Journal of Psychophysiology*, *94*, 407–419.
- Van Strien, J. W., Hagenbeek, R. E., Stam, C. J., Rombouts, S. A. R. B., & Barkhof, F. (2005). Changes in brain electrical activity during extended continuous word recognition. *Neuroimage*, *26*, 952–959.
- Van Veen, B. D., Van Drongelen, W., Yuchtman, M., & Suzuki, A. (1997). Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Transactions on Biomedical Engineering*, *44*, 867–880.
- van Vliet, M., Chumerin, N., De Deyne, S., Wiersema, J. R., Fias, W., Storms, G., et al. (2016). Single-trial ERP component analysis using a spatiotemporal LCMV beamformer. *IEEE Transactions on Biomedical Engineering*, *63*, 55–66.
- van Vliet, M., Manyakov, N. V., Storms, G., Fias, W., Wiersema, J. R., & Van Hulle, M. M. (2014). Response-related potentials during semantic priming: The effect of a speeded button response task on ERPs. *PLoS One*, *9*, e87650.
- Wittevrongel, B., & Van Hulle, M. M. (2016). Faster P300 classifier training using spatiotemporal beamforming. *International Journal of Neural Systems*, *26*, 1650014.