

The concise language paradigm (CLaP), a framework for studying the intersection of comprehension and production: Electrophysiological properties

Natascha Marie Roos^a, Julia Chauvet^{b,c}, Vitória Piai^{a,d}

Author names and affiliations:

Natascha Marie Roos | natascha.roos@donders.ru.nl

<https://orcid.org/0000-0002-1645-9110>

^a Radboud University, Donders Centre for Cognition, Nijmegen, Netherlands

Julia Chauvet | julia.chauvet@mpi.nl

<https://orcid.org/0009-0005-7250-0615>

^b Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands

^c Ernst Strüngmann Institute for Neuroscience in Cooperation with Max Planck Society, Frankfurt, Germany

Vitória Piai | vitoria.piai@donders.ru.nl

<https://orcid.org/0000-0002-4860-5952>

^a Radboud University, Donders Institute for Brain, Cognition, and Behavior, Nijmegen, Netherlands

^d Radboudumc, Donders Institute for Brain, Cognition, and Behavior, Department of Medical Psychology, Nijmegen, Netherlands

Corresponding author:

Natascha Marie Roos

natascha.roos@donders.ru.nl

Postbus 9104, 6500 HE NIJMEGEN, Netherlands

The concise language paradigm (CLaP), a framework for studying the intersection of comprehension and production: Electrophysiological properties

0. Abstract

Studies investigating language commonly isolate one language modality or process, focusing on comprehension or production. We aim to combine both in the new Concise Language Paradigm (CLaP), tapping into comprehension and production within one trial. The trial structure is identical across conditions, presenting an auditory sentence (constrained, unconstrained, time-reversed) followed by a picture to be named (normal, scrambled). We tested 21 young healthy speakers with EEG to examine several validations and novel contrasts afforded by the CLaP. Behavioral results revealed fastest naming times for pictures following constrained sentences, indicating word retrieval based on sentence constraint prior to picture onset. Naming following unconstrained sentences was equally fast as bare picture naming. Brain responses to normal speech after sentence onset (240-400ms) differed significantly from those to time-reversed speech. Picture-locked ERPs showed amplitude differences as a function of condition, especially in the P2 component (200-300ms), and were also modulated by preceding constrained sentences. Alpha-beta power was decreased for contextually guided picture naming relative to time-reversed or unconstrained sentences. These results indicate that the CLaP offers a promising framework to investigate the language system, affording different versions of the linguistic content and tasks, in combination with electrophysiology or other imaging methods.

Keywords: spoken word production, language comprehension, ERPs, alpha, beta, prediction

1. Introduction

A complete understanding of the language system in the brain requires a characterization of that system in its fullest range. However, studies investigating language usually isolate one modality, focusing on either comprehension or production. In fact, the language system in the brain has been mostly understood from the perspective of language comprehension, rather than production. While isolating a modality is convenient and enables highly controlled investigations of specific processes, it provides an incomplete picture of how language is used. In the present study, we propose a framework to study both aspects of language simultaneously, by tapping into comprehension and production within the same trial: The *concise language paradigm* (CLaP).

In the CLaP, all trials have exactly the same structure across conditions. This means they require the same instructions to participants and thereby reduce task-related confounds as much as possible. Irrespective of condition, all trials present an auditory stimulus followed by a visual stimulus. Specifically, participants listen to sentences that are either contextually constrained, unconstrained, or time-reversed. The last word of constrained and unconstrained sentences is shown as a picture, which participants have to name. Time-reversed sentences are either followed by a normal picture or a scrambled picture (see Figure 1). As such, the CLaP yields several low-level and high-level contrasts of language processes that are part of comprehension as well as production, such as sentence comprehension, contextually and visually guided word retrieval (Roos et al., 2023), object recognition, and naming.

So far, the language system in the brain has mostly been investigated using functional magnetic resonance imaging (fMRI) and its hemodynamic measure, the blood-oxygen-level-dependent (BOLD) signal. The BOLD signal is a metabolic measure rather than a direct reflection of neuronal activity (Logothetis & Wandell, 2004). Although it offers great spatial resolution with a common voxel size of 1 mm³, its temporal resolution in the range of multiple seconds is not sufficient to investigate the time-course of language processes, which happen in the range of milliseconds. Electrophysiology, by contrast, provides direct measures of net neuronal activity with excellent temporal resolution. Drawing

on these features of electrophysiology, here we combined the CLaP with electroencephalography (EEG), aiming to track processes of language comprehension and production within one framework.

In the present study we focused on two aims. First, as a form of validation, we inspected two low-level and one high-level contrast afforded by the current version of the paradigm for which clear predictions could be derived from the literature. For that, we derived a measure of speech perception/comprehension by comparing sentence-locked brain responses to meaningful speech sentences with those of unintelligible time-reversed speech (i.e., low-level auditory control condition). In addition, we examined object recognition and word production by comparing (early) picture-locked brain responses for normal objects with those of meaningless scrambled objects (i.e., low-level visual control condition). Finally, we also replicated a high-level pre-picture contrast of contextual constraint in the lead-in sentence for spoken word retrieval (e.g., Piai et al., 2014; Roos & Piai, 2020).

Second, we examined three novel contrasts that are of theoretical importance for understanding the neurophysiology of the language system. These novel, theoretically relevant contrasts focus on further understanding the nature of the context effect in constrained sentences, and the unfolding of contextually guided and visually guided picture naming and the relation between them. Below, we briefly review previous studies that are relevant for the contrasts reported here. Some of these contrasts are known for yielding phenomena at the level of event-related potentials (ERPs), and others at the level of oscillations (which include both phase-locked and non-phase-locked activity), quantified by time-frequency-resolved modulations (TFRs).

1.1. Validation contrasts

Time-reversed speech is an unintelligible counterpart of real speech with the same physical complexity and global acoustic characteristics. While real speech tends to have fast onsets and long decays, reversed speech results in the opposite, yielding sound sequences that do not occur in real speech. Reversed speech thus conveys less phonetic and lexical-semantic information and is commonly used as a non-semantic control condition for auditory speech processing (Binder et al., 2000; Narain et al., 2003; Stoppelman et al., 2013; for electrophysiology Brown et al., 2012; Forseth et al., 2018). When used in combination with fMRI, some studies report stronger BOLD responses in bilateral superior temporal regions to reversed speech (Binder et al., 2000; but see Brown et al., 2012), while other studies report that the BOLD responses of reversed speech mostly overlap with those of real speech in frontal and temporal language regions, arguing that it provides a less optimal baseline to isolate speech processing regions (Narain et al., 2003; Stoppelman et al., 2013). In terms of ERPs, it has been found that signal amplitude is increased by time-reversal of speech relative to intelligible speech in the first 300ms of stimulus presentation (Boulenger et al., 2011). In the present study, this could lead to a difference in auditory evoked ERPs (and their counterpart in the time-frequency domain) for reversed versus real speech (*validation contrast 1*).

Regarding brain responses to visual stimuli, previous studies have found differences in ERP amplitudes for scrambled versus normal objects. Visual evoked potentials (VEPs) for scrambled pictures showed higher amplitudes compared to normal objects over posterior sites (Gruber & Müller, 2005). Another study investigating brain network modularity for normal and scrambled pictures found higher interaction between brain modules for scrambled compared to normal pictures in visual processing brain regions (Rizkallah et al., 2018). They interpreted this finding as increasing communication in the brain while trying to match the characteristics of the scrambled pictures to already existing representations in visual memory. As the scrambled pictures are unknown and cannot be recognized as existing objects, all such attempts fail. With regard to the present study, these findings predict higher amplitudes in response to scrambled pictures compared to meaningful pictures (in the absence of a priming lead-in context) (*validation contrast 2*).

The high-level pre-picture context contrast has been investigated in several previous studies. Participants heard or read incomplete constrained and unconstrained sentences, followed by a picture concluding the sentence which they have to name. In the electrophysiological signal, the contrast between constrained and unconstrained picture naming yields power decreases in the alpha-beta frequency range (8-25 Hz) prior to picture onset (e.g., Piai et al., 2014, 2015, 2017, 2018). This effect is generated in left hemisphere (posterior) perisylvian areas and is not only replicable across studies, but also across sessions within-participants (Klaus et al., 2020; Roos & Piai, 2020, for fMRI see Roos et al., 2023). These power decreases have been interpreted as word retrieval processes taking place already before picture onset in constrained but not in unconstrained picture naming. In fact, we found that the amount of power decreases correlates with picture naming times, such that faster picture naming is associated with stronger power decreases in constrained picture naming (Roos & Piai, 2020). On the contrary, there was no correlation between these power decreases and cloze probabilities (i.e., percentage of people who complete the sentence with the by-us intended target word) for constrained contexts (Hustá et al., 2021). These results suggest that the alpha-beta power effects are more directly related to picture naming processes than being aspects of sentence comprehension. We expected to replicate the pre-picture alpha-beta power decreases for the context effect (*validation contrast 3*).

1.2. Novel contrasts

The interpretation of the context effect discussed above has remained unsubstantiated so far by the lack of a low-level control condition. That is, it remained unclear whether the relative power decreases in constrained versus unconstrained picture naming emerged from power *increases* in unconstrained picture naming, rather than power *decreases* in constrained picture naming. The addition of time-reversed speech trials as a low-level auditory control condition to the CLaP can thus further inform us on the underlying power dynamics of the already established context effect (*novel contrast 1*).

The next theoretically relevant question concerns how conceptual-lexical retrieval unfolds. Of the four different picture naming conditions of the CLaP we were first of all interested in comparing visually and contextually guided picture naming (both following meaningful speech sentences). In visually guided naming, the concept is presented to participants as a picture, whereas in contextually guided naming it is revealed via the context of constrained sentences. For the latter, once the picture is visually presented, a concept and its associated label (i.e., “lemma”) and potentially even a corresponding phonological form are already pre-activated (Piai et al., 2014, 2020). This could lead to a different brain response at picture presentation for constrained picture naming compared to all other conditions, where no concept is activated yet. However, the best-controlled contrast of this comparison is between picture naming following constrained versus unconstrained sentences, as both are preceded by meaningful speech stimuli (*novel contrast 2*).

The question remains to what extent this conceptual pre-activation proceeds. As participants are familiarized with the picture stimuli before the start of the experiment, the pre-activation could include having a visual representation of the concept in form of the exact picture used in the experiment. Along these lines, other picture naming studies have found repetition suppression in ERPs for repeated picture presentation leading to lower amplitudes for repeated compared to initial picture presentation (Gruber et al., 2004; Gruber & Müller, 2002, 2005; Rugg et al., 1995). If pre-activation of the concept based on the constraint of preceding sentences induces visual imagination of the familiarized stimuli, early brain responses to picture onset could behave similarly to repetition suppression and reveal lower amplitudes for constrained picture naming compared to all other conditions. Similarly, pre-activation of an abstract conceptual representation could also lead to modulations of early brain responses.

Other studies on picture naming have linked the P2 component (approx. 200ms post picture onset) to processes of lexical selection (Fargier & Laganaro, 2017, 2020; Indefrey, 2011), including linking it to the ease or difficulty of lexical selection. These studies report positive correlations between ERP

amplitudes and picture naming times, such that high P2 amplitudes are associated with less accessible words and slower naming, and lower P2 amplitudes with more easily accessible, high-frequency words and faster naming (Rabovsky et al., 2021; Strijkers et al., 2010). In line with the findings in repetition suppression above, this would again suggest lower amplitudes for constrained pictures around 200ms after picture onset, as lexical information is retrieved prior to picture onset in constrained picture naming.

A last theoretically relevant question regarding conceptual-/lexical retrieval is whether naming following unconstrained sentences is any different from commonly used bare picture naming (*novel contrast 3*). Both are visually guided naming conditions, but preceded by different auditory sentence stimuli (unconstrained: meaningful, bare: reversed). In a previous single case study of a person with aphasia due to extensive left hemisphere damage, the rate of anomia was the same for unconstrained and bare picture naming, whereas constrained sentences increased the rate of successful naming attempts (Chupina et al., 2022). This suggests no difference between these two different types of visually guided picture naming in the current study.

1.3. Summary

In sum, we anticipated a difference in brain responses to time-reversed speech compared to real speech after sentence onset (*validation contrast 1*). In terms of visual stimuli, we predicted higher amplitudes to scrambled pictures compared to meaningful ones (*validation contrast 2*). We further predicted that power in the alpha-beta frequency range shows a context effect before picture onset between constrained and unconstrained picture naming (*validation contrast 3*). For the novel contrasts, by virtue of the time-reversed speech sentences, we hoped to clarify whether these previously observed alpha-beta oscillatory dynamics are either driven by power decreases in constrained or power increases in unconstrained sentences (*novel contrast 1*). This clarification would add to our understanding of how oscillatory brain activity can be mapped onto retrieval processes. As word retrieval in constrained picture naming happens prior to picture presentation, we expected early brain responses after picture onset for this condition to diverge from the other conditions (*novel contrast 2*). Regarding visually guided naming, we expected no differences in brain responses between unconstrained and bare picture naming (*novel contrast 3*). Collectively, these hypotheses aim to provide an understanding of the brain responses underpinning the CLaP, and highlight the paradigm's approach to integrate measures of comprehension and production within a single trial.

2. Methods

This study was approved by the Ethics Committee of Social Sciences at Radboud University, following the Code of Ethics of the World Medical Association (Declaration of Helsinki). Informed consent was obtained from all individual participants included in the study. Neither the study nor any procedures or analyses were pre-registered prior to conducting the research. Data collection took place at the Donders Centre for Cognition (Radboud University) in Nijmegen in the Netherlands. The data and code are available via the Donders Repository (<https://doi.org/10.34973/19gn-7v46>).

2.1. Participants

We recruited 26 participants in total (20 females) between the ages of 18 and 28 ($M = 22$, $Mdn = 24$) to participate in the study for monetary compensation. One participant was a substitute for an unusable dataset due to missing EEG markers, another four datasets were excluded due to measurement mistakes (2) and noisy EEG data (2). Thus, the data presented here derives from a total of 21 subjects (15 females), for which we report naming accuracy and EEG results. For two subjects, there were no audio recordings available, so response time results are based on 19 subjects. All participants were right handed, native speakers of Dutch, with normal or corrected-to-normal vision and hearing, and without any neurological or language deficits.

2.2. Materials

The experimental stimuli consisted of 156 pictures to be named by the participants. While 126 of those were photographs of objects depicting the target noun, 30 pictures were scrambled pictures which are meaningless and unknown to the viewer, used as a low-level visual control condition. Photographs were taken from the BOSS database (Brodeur et al., 2010) and from the internet. Scrambled pictures were created based on photographs from the BOSS database and distorted to make them unrecognizable without majorly changing the basic visual properties, called diffeomorphic transformation (Stojanoski & Cusack, 2014). All object or scrambled pictures were depicted on black background with a height of 270 pixels. Seventy-eight of the photographs were preceded by an auditorily presented sentence that was either constrained or unconstrained towards the target noun. Sentence recordings were taken from a previous study (Chupina et al., 2022), recorded by a native speaker of Dutch at a slow pace. Sentences were 4 to 8 words long including the target word ($M = 6$ words) and auditory sentence duration varied from 1.79 to 3.59s ($M = 2.61$ s, $Mdn = 2.62$ s). The cloze probabilities for the target words were 0-39% in unconstrained sentences and 60-100% in constrained sentences ($t(77) = 51.236$, $p < 0.001$) (Chupina et al., 2022; Hustá et al., 2021). The other 48 photographs for bare picture naming and the 30 scrambled pictures were preceded by a reversed speech sentence. The reversed speech stimuli were created by time-reversing the 78 shortest constrained and unconstrained sentences using in-house MATLAB code.

2.3. Design

The experiment consisted of 234 trials in total. Sentences were divided into three different conditions: constrained ($n = 78$), unconstrained ($n = 78$), and reversed ($n = 78$). Constrained and unconstrained sentences preceded the same 78 pictures, such that each of these pictures appeared once after a constrained and once after an unconstrained sentence. The reversed sentences preceded either a normal photograph of an object for bare picture naming ($n = 48$), or a scrambled picture as control ($n = 30$). Thus, there were four production conditions: constrained naming, unconstrained naming, bare naming, and scrambled naming. Every trial was set up in the same manner: audio followed by visual stimuli to be named, irrespective of condition. Trials were pseudorandomized using Mix (van Casteren & Davis, 2006) with at least 20 trials between both appearances of the same target picture and no more than four consecutive trials of the same condition, yielding a unique list per subject. Figure 1 illustrates example trials of each condition.

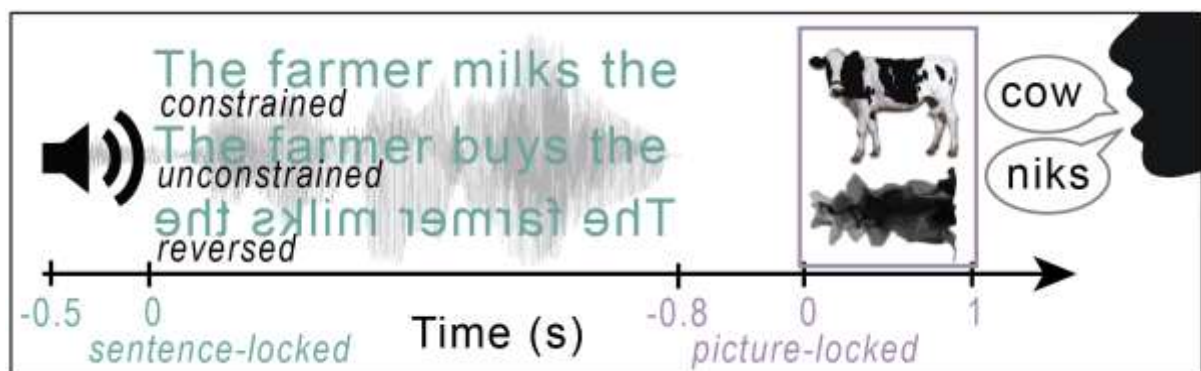


Figure 1. Schematic overview of example trials for sentence conditions (constrained, unconstrained, reversed) and picture types (normal, scrambled), yielding constrained naming, unconstrained naming, bare naming, and scrambled naming conditions. Note the different time-lockings to sentence or picture onset. Sentence time varies per trial. Reproduced with permission from the authors from <https://doi.org/10.17605/OSF.IO/W5Q7S>.

2.4. Procedure

Before giving informed consent we instructed participants about the EEG measurement and the experimental task. The purpose was to listen to the auditory stimuli and accurately name the following picture with the respective target noun. We instructed participants to name scrambled pictures with the word “niks” (Dutch for *nothing*). During the EEG preparation, participants went through all photograph stimuli and the corresponding target nouns in a slide show, including some examples of scrambled pictures, to decrease naming variability and increase accuracy. Before the start of the EEG recording, we instructed participants to sit as still as possible, keep their back and shoulders relaxed and to keep blinking to the blinking intervals following picture offset. Participants sat on a chair in front of the computer screen on which the experiment was presented. We presented the stimuli by means of Presentation software (*Neurobehavioral Systems Inc.*, Berkeley, CA, www.neurobs.com), which enabled us to start audio recordings with picture onset to capture participants’ responses for later analyses. Each trial started with a fixation cross in the middle of the screen and the presentation of the auditory stimulus after 1500ms. Then, 800ms after auditory stimulus offset the respective picture appeared on screen for 1000ms, followed by a 2000ms blinking interval showing three asterisks (***), leading to the following trial. Before the start of the EEG recording we started with eight practice trials (three normal sentences and five reversed sentences) to ensure that participants had understood the task well. If not, the practice trials were repeated. The experiment was divided into nine blocks of 26 trials each and lasted around 40 minutes. After each block, participants could determine if and for how long they wanted to rest. A whole testing session including preparation and EEG recording took approximately 120 minutes.

2.5. Response time analysis

We coded participants’ picture naming responses online during the EEG recording for accuracy. If participants hesitated, uttered more than one word or nothing at all, responses were coded as incorrect. Synonyms for the target nouns were coded as correct, given that they made sense in the context of the preceding sentence, if applicable. Voice recordings started with picture onset and lasted for 3500ms. Based on these recordings we later manually marked the response times using the speech editor PRAAT (Boersma & Weenink, 2017), blinded for condition. Statistical analysis of the response times was done in R (R. Core Team, 2017) for correct trials only. To get all relevant comparisons between contrasts, we ran two linear mixed-effects regression models. Both models had fixed effects of condition and by-participant and by-item random intercepts and slopes for condition. The reference used for the first model was bare picture naming to which the other conditions were compared. For the second model, we used constrained picture naming as the reference, to have a direct comparison of naming times for constrained picture naming with unconstrained and scrambled picture naming.

2.6. EEG acquisition

The EEG was recorded continuously using BrainVisionRecorder from 64 active electrodes. Electrode positions were based on the international 10-20 convention using an Acticap system and a BrainAmps DC amplifier (500 Hz sampling, 0.016-100 Hz band-pass). The online reference electrode was placed on the left mastoid and the ground electrode in the position of ‘AFz’. Impedance for both was kept below 5 kOhm, and below 10 kOhm for all other electrodes. The vertical and horizontal electro-oculogram (EOG) were recorded with electrodes above and next to the eyes, and an additional electrode (‘T7’) removed from the cap to be placed below the left eye. Two other electrodes (‘T8’, ‘TP10’) were removed from the cap to be placed on the orbicularis muscle on the right side of the mouth, above and below the lips. To synchronize the presentation of the stimuli with the EEG data, we sent condition specific markers at the onset of every sentence and picture which were recorded with the EEG data.

2.7. EEG preprocessing and analyses

EEG preprocessing and analyses were performed in MATLAB using FieldTrip (Oostenveld et al., 2011). For data cleaning, we cut the trials from 800ms before sentence onset to 1400ms after picture onset. Trials were demeaned and low-pass filtered at 50 Hz. After rejecting incorrect trials, we visually checked the data for flat and noisy channels or trials to be removed due to excessive artifacts (except eye movements and blinks). Then we performed an Independent Component Analysis (ICA) to remove components from the data that were related to blinking and eye movements (Jung et al., 2000). After the ICA, we interpolated bad channels using a weighted average of all neighboring channels and then re-referenced the data to the common average of all channels. Finally, we visually inspected each trial and marked the remaining noisy segments.

All EEG analyses were conducted on the scalp level, locked either to sentence or picture onset. Sentence or picture segments that contained artifacts as marked during visual inspection of the preprocessing stages were discarded from the analyses. Segments for sentence analysis were cut from -800ms prior to sentence onset to 1800ms into the sentence. Note that this time point is based on the length of the shortest included sentence. Hence, for the longest meaningful sentences this segment only included approximately the first half of the sentence. For picture segments, the segment was -1000ms to 300ms relative to picture onset.

All contrasts were calculated based on the lowest number of available trials in the conditions included in the respective comparison. Abundant trials were removed by sentence length to meet this target number with the shortest sentences included, as reversed sentences were created based on the 78 shortest meaningful sentences. For the sentence-locked analyses, all conditions (constrained, unconstrained, reversed) included a maximum of 78 trials. The pre-picture context contrasts (TFRs) were based on a maximum of 48 trials (amount of bare picture trials); and picture-locked ERPs on a maximum of 30 trials (amount of scrambled picture trials).

All ERPs were low-pass filtered at 40 Hz and averaged across trials per participant and condition. Sentence-locked ERPs (78 trials) were baseline-corrected using the interval from -800ms to sentence onset, while picture-locked ERPs (30 trials) were not baseline-corrected, as the signal amplitude was modulated by the preceding sentences resulting in systematic differences in “baseline” periods. All TFRs were computed for frequencies from 3 to 40 Hz with a sliding time window of 3 cycles, advanced in steps of 50ms and 1 Hz. Each time window was multiplied with a Hanning taper with implemented time–frequency transformation based on multiplication in the frequency domain. TFRs and ERPs were then averaged across trials per participant and condition.

2.7.1. Statistical analyses

For the sentence segments we compared language comprehension during meaningful and reversed speech sentences (*validation contrast 1*). Here we looked at sentence-locked ERPs for all three conditions to investigate auditory evoked potentials after sentence onset. We also looked at TFRs for the three possible contrasts between conditions from -800ms before to 1800ms during the sentence.

For the high-level context effect prior to picture onset, we looked at TFRs by contrasting the pre-picture interval for constrained and unconstrained picture naming (*validation contrast 3*), and compared constrained versus reversed with unconstrained versus reversed sentences (*novel contrast 1*). Here we only used sentences preceding bare picture naming (48 trials), to ensure that this time interval is not affected by anything related to scrambled picture viewing. We also looked at the context effect based on all (max 78) constrained and unconstrained trials per participant. As a final sanity check, we compared the pre-picture intervals of bare and scrambled picture viewing (both following reversed sentences).

To investigate early brain responses to picture onset for the four different picture conditions we compared picture-locked ERPs across conditions. As the scrambled condition contained only 30 trials

per session, all picture-locked ERP comparisons are based on a maximum of 30 trials. Specifically we were interested in scrambled versus bare picture naming (*validation contrast 2*), as well as constrained versus unconstrained picture naming (*novel contrast 2*) and unconstrained versus bare picture naming (*novel contrast 3*).

All ERP as well as TFR contrasts were evaluated by means of non-parametric cluster-based permutation tests (Maris & Oostenveld, 2007) on the group-level. A dependent samples t-test was performed at every channel-time(-frequency) sample and those exceeding a threshold ($p < .05$, two-tailed) were identified for subsequent clustering (adjacent time and frequency samples, minimum number of neighboring channels = 2). A cluster-level statistic was defined as the sum of t values within each cluster. Then, based on 1,000 random permutations of the conditions being tested, the same clustering procedure was performed. The Monte Carlo p-value of the empirically observed clusters was computed as the amount of 1,000 random permutations yielding a more extreme cluster-level statistic than the observed one, again at an alpha level of 0.05 (two-tailed). All available time (and frequency) points and channels entered the comparisons.

3. Results

3.1. Response times

For the behavioral results, we first looked at the errors that participants made overall as well as per condition. On average, participants ($n = 21$) made between 0 and 5 errors ($M = 2$, $Mdn = 2$, $SD = 1.7$). The overall error rate was 0.9% (43 errors in total). Participants made no errors in naming scrambled pictures, 10 errors in bare picture naming (0.2%), 13 in constrained picture naming (0.3%), and 20 in unconstrained picture naming (0.4%).

The mean naming times per condition on the group level were 599ms for constrained picture naming ($Mdn = 602$, $SD = 57$), 693ms for scrambled picture naming ($Mdn = 698$, $SD = 75$), 768ms for unconstrained picture naming ($Mdn = 753$, $SD = 64$), and 790ms for bare picture naming ($Mdn = 793$, $SD = 8$), based on 19 participants. These naming times are shown in Figure 2, with a different color for every participant. The black dots represent the group mean.

The first linear mixed effects regression to compare response times (in ms) per condition with bare picture naming as the reference showed that bare picture naming was significantly slower than constrained picture naming (191ms avg. difference, Estimate = -192.07, SE = 11.69, $t = -16.43$, $p < .001$), as well as scrambled picture naming (97ms avg. difference, Estimate = -97.47, SE = 14.83, $t = -6.57$, $p < .001$). For bare and unconstrained picture naming there was no significant difference (22ms avg. difference, Estimate = -22.08, SE = 11.7, $t = -1.89$, $p = 0.06$). Note that both scrambled and bare naming conditions were preceded by reversed sentences, differently from unconstrained naming, yet naming times differed in a way that did not directly follow from the type of the preceding sentence. For the second model, we used constrained picture naming as the reference. Unsurprisingly, this yielded significant differences between constrained picture naming and all other conditions: unconstrained (169ms avg. difference, Estimate = 169.99, SE = 5.03, $t = 33.79$, $p < .001$), scrambled (94ms avg. difference, Estimate = 94.6, SE = 13.68, $t = 6.91$, $p < .001$), bare (Estimate = 192.07, SE = 11.69, $t = 16.43$, $p < .001$).

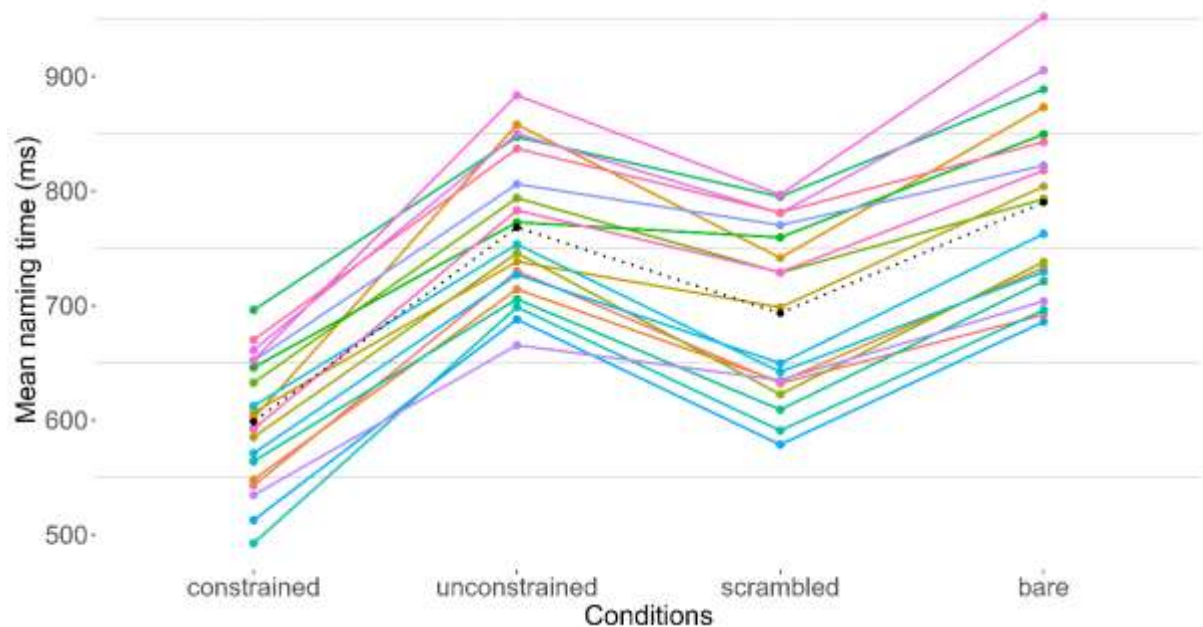


Figure 2. Mean naming times per condition. All values for one participant are connected by one line. Black dots connected by the dotted line represent the group mean per condition. Reproduced with permission from the authors from <https://doi.org/10.17605/OSF.IO/W5Q7S>.

3.2. Sentence comprehension (sentence-locked ERPs and TFRs)

As a measure of comprehension (*validation contrast 1*), we compared ERPs across sentence conditions (constrained, unconstrained, reversed). These results are shown in Figure 3A. The auditory responses were clearly visible for constrained and unconstrained (both meaningful speech) sentences from 240ms to 400ms after sentence onset with a peak amplitude between 3.5-4 μ V for electrode FCz (for electrode location see Figure 3A). These potentials both significantly differed from reversed speech ($p < .006$), which also showed a slight amplitude increase in the same time-frame, but with a much lower peak amplitude of 1.5 μ V. We found no significant differences between constrained and unconstrained sentences.

In Figure 3A, we also show the topographies for the auditory response peak from 240-400ms per condition. In constrained and unconstrained sentences, we see a strong posterior negativity and a central to bi-lateral positivity. In reversed sentences, the amplitude is smaller in general, and the positivity seems to be more focal and symmetrical, rather than central. This could potentially indicate the difference between left-lateralized processing of speech in meaningful sentences compared to low-level auditory, non-linguistic input in reversed sentences.

Following this initial evoked response, reversed sentences continued to diverge from meaningful speech with sustained amplitude differences throughout the course of the sentence, significantly differing from unconstrained sentences between 1.1 and 1.4s ($p = 0.0099$).

To investigate oscillatory dynamics during the sentence time-window (-800ms to 1800ms), we looked at all possible contrasts between the three sentence types. The most salient aspect of the TFRs of both constrained and unconstrained over reversed sentences was the power increase up to 10 Hz corresponding to the phase-locked responses we see in the ERPs, roughly 240-400ms after sentence onset as shown in Figure 3B and 3C. Constrained over unconstrained sentences did not yield any significant differences in this interval, neither did the contrast between unconstrained and reversed sentences. The only significant difference in TFR contrasts during the sentence was between constrained and reversed sentences, which yielded one cluster ($p = .032$) of power increases. More

trials or a different analytical approach (e.g., Zioga et al., 2023) would likely be needed for investigations of these auditory responses in the spectrotemporal domain.

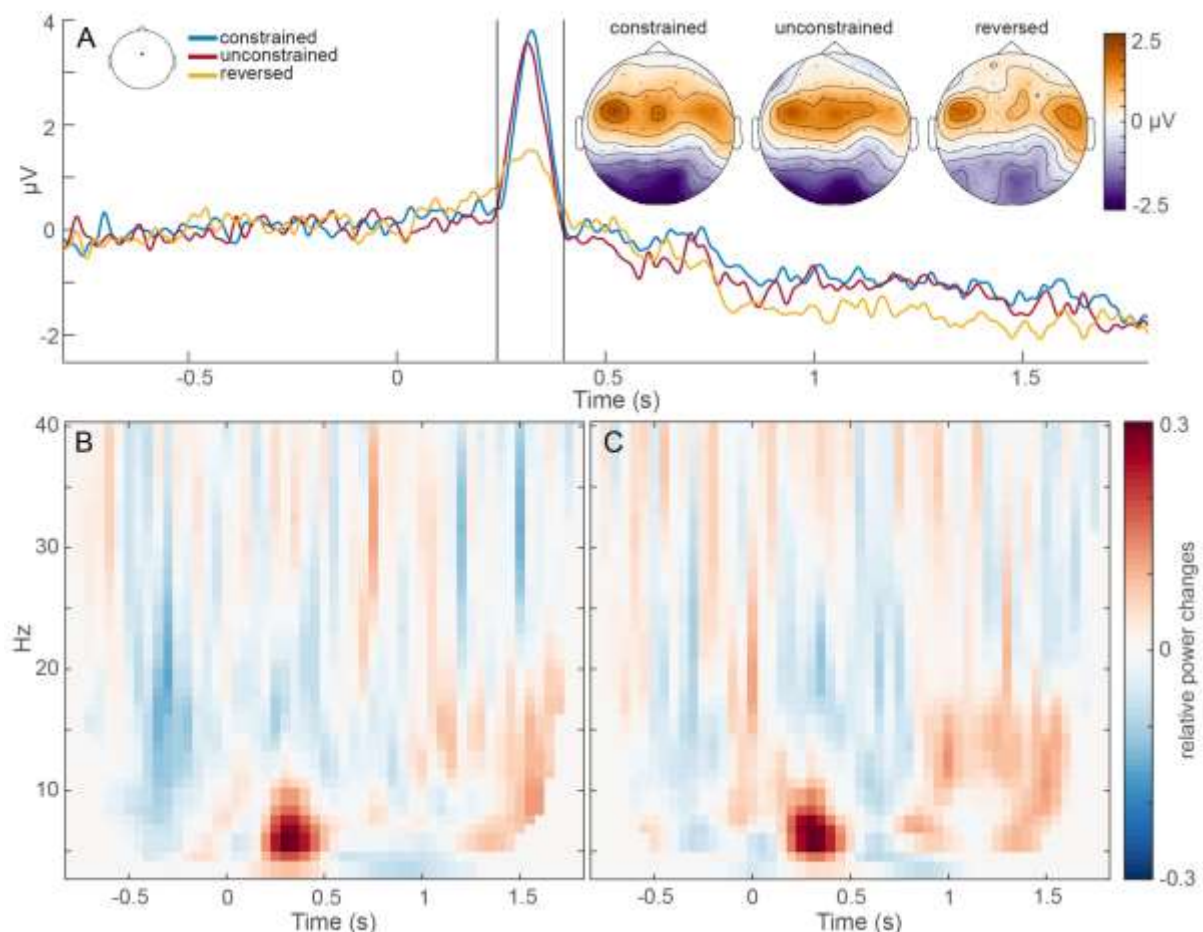


Figure 3. Event-related potentials (A) and time-frequency representations (B, C) during sentence onset (-0.8 to 1.8s) for electrode FCz as shown in the empty topographic schematics on top. Topographic plots show topographies in μV for the ERPs from 240-400ms after sentence onset as marked by the two vertical lines in the ERP plot per condition. Time-frequency plots in panels B and C display relative power changes (difference between conditions divided by their mean) for the contrasts of constrained over reversed (B) and unconstrained over reversed sentences (C). Time 0s represents sentence onset in all plots. Reproduced with permission from the authors from <https://doi.org/10.17605/OSF.IO/W5Q7S>.

3.3. Picture Viewing and Naming (picture-locked ERPs)

In order to compare ERPs after picture onset across conditions, and specifically visual evoked potentials (VEPs), we investigated the potentials locked to picture onset. Based on the data as plotted in Figure 4 for occipital channels, we divided the three different components of the VEP as follows: first positive component (P1) at 100-160ms, first negative component (N1) at 160-200ms, and second positive component (P2) at 200-300ms.

The pattern of condition-specific VEPs remains the same, also when plotting different channel groups than occipital. In the right half of Figure 4 we show the corresponding topographies to the P1, N1, and P2 (rows) per condition (columns). These clearly show strong occipital responses to the visual stimuli, in line with visual evoked potential topographies in other studies (cf. Gruber et al., 2004; Gruber &

Müller, 2005). As for the ERPs, the topographies of constrained sentences diverge most from the other three conditions with lower amplitudes across all three components.

3.3.1 Object Recognition and Naming (validation contrast 2)

To evaluate object recognition and naming, we compared bare with scrambled picture conditions. As both are preceded by reversed sentences, they only differ in the aspect of showing a normal picture (bare) versus a scrambled picture. During the pre-picture interval, the ERPs of bare and scrambled pictures do not show any difference. Only after picture onset their VEPs diverge. While the P1 still looks the same for both, bare picture naming has a significantly lower amplitude in the N1 component ($p = 0.002$), as well as the P2 component ($p = 0.002$).

3.3.2 Conceptual-Lexical Retrieval (novel contrast 2)

As a measure of conceptual-lexical retrieval between constrained and unconstrained picture naming, we compared the ERPs of these two conditions. As demonstrated in Figure 4, the ERPs of constrained and unconstrained pictures already significantly differ during the end of the sentence and first half of the pre-picture interval, from -1000 to -500ms ($ps < 0.034$), indicating that the ERPs are already modulated by the degree of sentence constraint. With respect to the VEPs of constrained and unconstrained picture conditions, they differ in all three components (100-300ms), with constrained sentences showing significantly lower amplitudes at all three potentials ($ps < 0.004$).

3.3.3 Visually Guided Naming (novel contrast 3)

Finally, we also investigated visually guided naming, for which we compared bare and unconstrained picture naming (preceded by reversed versus meaningful sentences, respectively). The ERPs of these two conditions behaved very similarly throughout the VEPs, as well as the rest of the segment, and only significantly differed from -540 to -440ms during the pre-picture interval ($p = 0.034$). The amplitudes of their VEP components are highly similar throughout, and only start to diverge towards the end of the segment available for analysis, after the P2.

3.3.4. Summary

In sum, during the P1, only constrained picture viewing diverged from the other three conditions, with a $1\mu V$ lower amplitude. During the N1, the constrained condition remains the one with the lowest amplitude, with bare and scrambled picture conditions also differing. At the P2, the amplitude of constrained picture viewing is still significantly lower than that of unconstrained (and bare) picture viewing, and bare picture viewing has a lower amplitude than scrambled picture viewing.

Taken together, these results show that the ordering of conditions in terms of ERP amplitude does not follow from the type of sentence preceding the picture (constrained, unconstrained, reversed). This is, in fact, the same as the behavioral results of naming times already revealed. The key factors modulating signal amplitude are, firstly, that of a real object being presented, and secondly, the semantic constraint in constrained sentences leading to conceptual pre-activation and word retrieval prior to picture onset, which is absent in all other conditions, regardless of whether their lead-in sentence was reversed or meaningful (unconstrained).

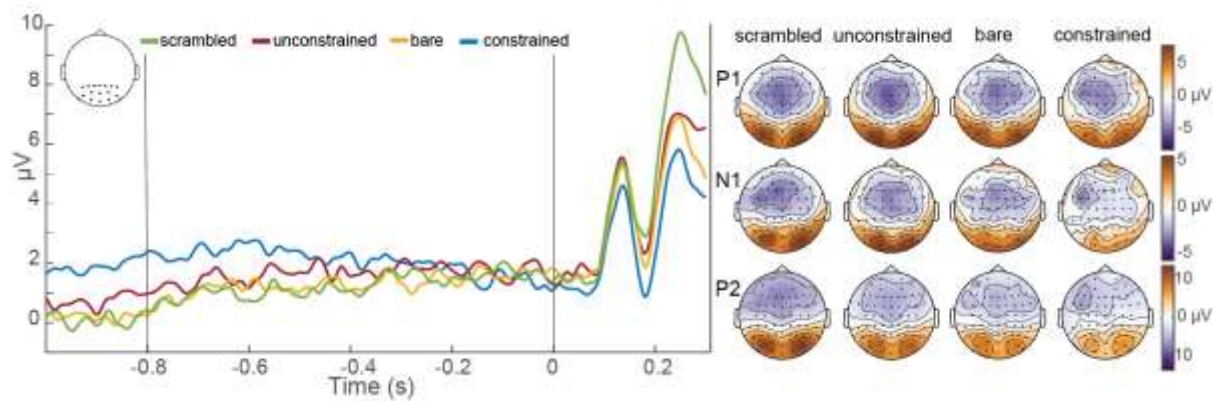


Figure 4. Event-related potentials during pre-picture interval and picture presentation per condition over selected occipital channels (shown in the empty topographic plot in the top left corner). The first vertical line at -0.8s represents sentence offset and the second vertical line at 0s represents picture onset. For each peak of the visual evoked potential (P1: 100-160ms, N1: 160-200ms, P2: 200-300ms) we show the respective topography in columns per condition on the right. The color scales are the same per ERP component, but different between components. Reproduced with permission from the authors from <https://doi.org/10.17605/OSF.IO/W5Q7S>.

3.4. Context Effect (pre-picture TFRs)

3.4.1 Contextually Guided Naming (*validation contrast 3*)

As a validation step, contrasting constrained and unconstrained pre-picture intervals yielded the expected context effect as found in previous studies. The most prominent cluster of power decreases was found in the alpha-beta frequency range, approximately 8-25 Hz, already prior to picture onset, spanning the entire time-window from -800ms pre to 200ms post picture onset ($p = .002$). The power decreases initially have a left bias in the topographies, but become more widespread over time. These results, together with time-frequency plots of the t -values above cluster-level threshold for two different channels are depicted in Figure 5A. The same pattern was found when all available trials per participant (78 instead of 48) were used in the comparison.

3.4.2 Nature of the Context Effect (novel contrast 1)

So far, the nature of the power decreases of the context effect has remained underdetermined. In theory, the decreases could derive from power increases in unconstrained picture naming, rather than power decreases in constrained picture naming. To investigate this, we contrasted the pre-picture intervals following both meaningful sentence types with those following reversed sentences (a low-level control condition). Just as for constrained versus unconstrained contexts, this analysis revealed significant clusters of power decreases ($ps \leq .03$) for constrained versus reversed pre-picture intervals. The respective topographies show initial power increases around sentence offset over right electrode sites, as well as left-hemisphere power decreases becoming more dominant and widespread before picture onset (see also the lower two plots with t -values above cluster-level threshold before as well as after picture onset, Figure 5B).

The comparison of unconstrained over reversed pre-picture intervals resulted in a significant cluster of power increases ($p = .04$) around sentence offset, which can be seen in Figure 5C. Here, the topographies almost exclusively reveal central to bilateral power increases just after sentence offset that become weaker towards picture onset with a slight left lateralized decrease prior to picture onset. For this contrast (unconstrained versus reversed pre-picture intervals), we do not get any t -values above threshold at the same two electrodes that yielded above-threshold t -values for the previous two comparisons. For our sanity check of comparing the pre-picture intervals of bare versus scrambled picture viewing, both preceded by reversed sentences, the results revealed no significant differences between these two conditions.

In sum, we find similar alpha-beta power decreases for both contrasts including constrained picture naming, opposed to power increases for unconstrained picture naming at sentence offset. This indicates that the context effect is characterized by power decreases linked to constrained picture naming, rather than (only) power increases in unconstrained picture naming.

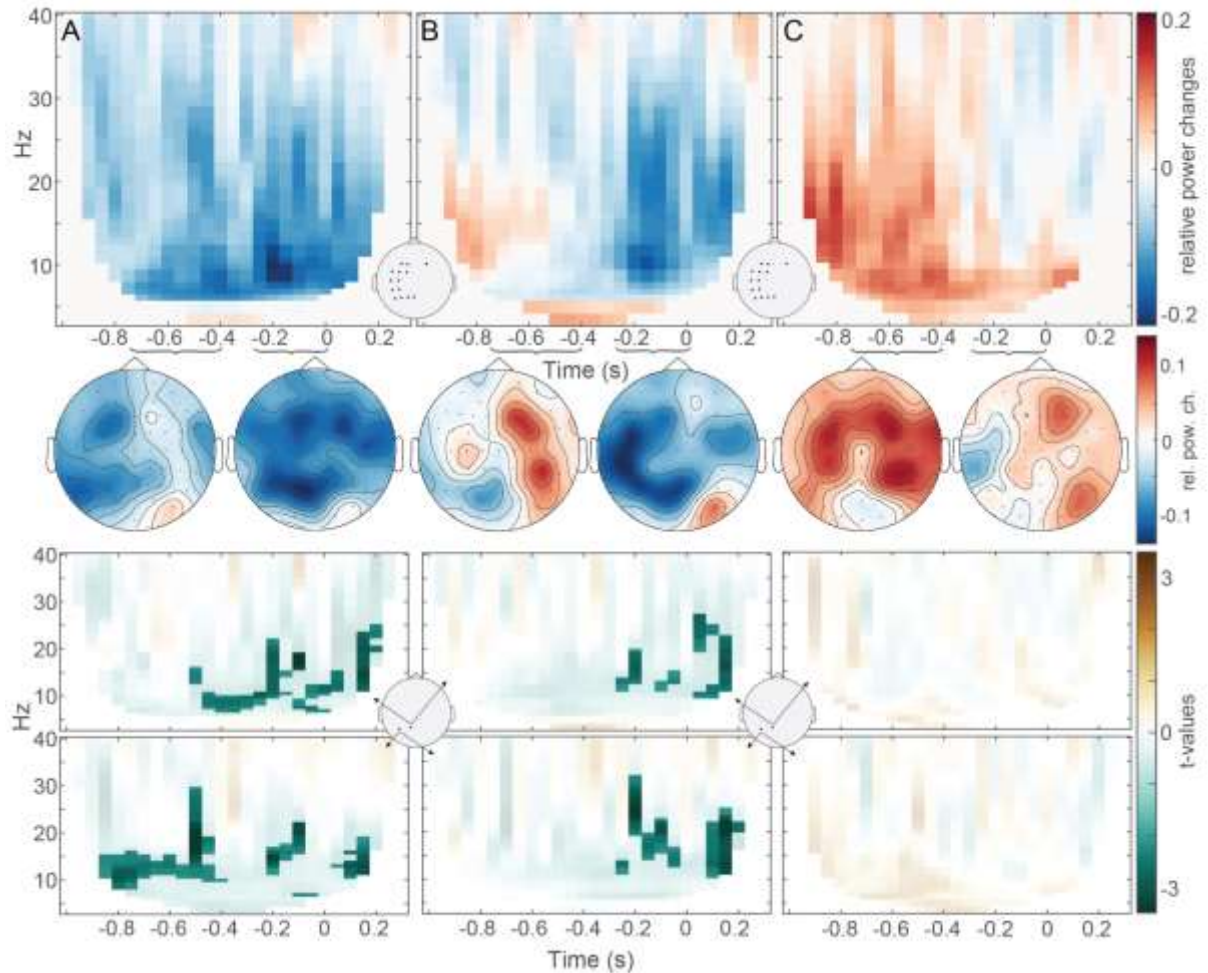


Figure 5. Time-frequency representations during the pre-picture interval per contrast between conditions. All TFR effects were calculated by taking the difference between two conditions of interest divided by their mean. Time 0s represents picture onset, -0.8s is sentence offset. A) constrained over unconstrained sentences, B) constrained over reversed sentences, C) unconstrained over reversed sentences. Selected channels are shown in the empty topographic plots in between. The large topographic plots show topographies from 8-25 Hz for -750ms to -400ms (left topographic plot) and -250ms to 0ms (right topographic plot) per contrast. The lower TFR plots show t-values masked at the cluster-level threshold for channels P1 (top) and P5 (bottom). The contrast of unconstrained over reversed sentences (C) does not yield any significant clusters during the pre-picture interval for the selected channels. Note that the upper two colorbars show relative power changes and the lower one shows t-values. Reproduced with permission from the authors from <https://doi.org/10.17605/OSF.IO/W5Q7S>.

4. Discussion

The present study was conducted to test the new concise language paradigm, or CLaP, a framework designed to tap into contrasts of comprehension as well as production within a single trial. The CLaP combines measures of sentence comprehension, contextually and visually guided word retrieval (Roos et al., 2023), object recognition, and picture naming. Here, we focused on the validation of three different established contrasts and investigated three novel contrasts with the current version of the CLaP, for which we tested 21 participants in combination with 64-electrode EEG recordings.

Picture naming was fastest following constrained sentences, but equally fast if the task was bare picture naming or naming following unconstrained sentences. For the validation aim, in terms of comprehension, ERPs around sentence onset differed between reversed and meaningful speech, but not between constrained and unconstrained sentences. Comparing bare and scrambled picture conditions to get a measure of object recognition and naming (both following reversed-speech sentences), we found differences in the N1 and P2 components. As a final validation we replicated the well-established alpha-beta power decreases prior to picture onset in constrained versus unconstrained picture naming (e.g., Piai et al., 2014; Roos & Piai, 2020). The nature of this alpha-beta oscillatory effect was clarified with the first novel contrast: the effect arises from power decreases for constrained picture naming, rather than power increases for unconstrained picture naming. The second and third novel contrasts of (the unfolding of) conceptual-lexical retrieval revealed that contextually and visually guided picture naming differ in terms of their ERPs. By contrast, within visually guided naming, whether the task is bare picture naming or picture naming following unconstrained sentences does not modulate the ERPs. Below we discuss all validation and novel contrasts of interest in more detail.

4.1 Sentence comprehension (validation contrast 1)

The current measure of sentence comprehension was based on a comparison of brain responses to the auditory sentence stimuli (constrained, unconstrained, reversed). The evoked potentials after sentence onset occurred in the same time frame across conditions, namely from 240 to 400ms (cf, Figure 3). The peaks of constrained and unconstrained sentences hardly differed, because both are meaningful speech and there is no difference between conditions at that point. The peak of reversed sentences, however, was relatively lower. This is in line with previous findings describing an early decay of the brain signal due to top-down attenuation once the auditory input is classified as non-speech (Stoppelman et al., 2013).

But not only the first evoked potentials differ between meaningful and unintelligible speech. After the initial evoked response, reversed sentences also diverged from meaningful speech with sustained lower amplitude than constrained and unconstrained sentences. However, the analysis approach we adopted here was likely not optimized for studying these differences in more depth. Given our proposal of the CLaP as a framework to study language comprehension and production, one could easily opt for a different analysis strategy, a point to which we turn below.

4.2 Picture Viewing and Naming

Our results for the comparison of ERPs locked to picture onset clearly showed that VEPs differed as a function of condition. While all conditions showed the same peak timings of the VEP (P1: 100-160ms, N1: 160-200ms, P2: 200-300ms), the amplitudes differed. However, the way in which they diverged from each other does not follow from the sentence type preceding picture onset (constrained, unconstrained, reversed). If we take bare and unconstrained picture naming as the norm, one could argue that the only divergence appears in scrambled and constrained picture naming, due to aspects of visual appearance and sentential semantic constraint respectively. We discuss these in line with our validation and novel contrasts below.

4.2.1 Object Recognition (validation contrast 2)

The VEP amplitude during the P2 component had a much higher amplitude for scrambled pictures compared to normal pictures. This is in line with previous studies investigating ERPs in response to scrambled pictures (Gruber & Müller, 2005), where it has been interpreted as the integration of an unknown visual stimulus trying to be matched to an existing object in visual memory (Rizkallah et al., 2018).

Although these previous findings would explain the current VEP results for scrambled pictures, we think that picture naming times for scrambled objects also need to be considered. These were significantly

shorter than those of bare picture naming and, in fact, the shortest after constrained picture naming. Participants did not have to comprehend/process a sentence preceding the picture and only had to identify the stimuli as scrambled/recognize that the picture is not depicting a real object. Then they could directly name the scrambled picture with the Dutch high-frequency word for 'nothing'.

Potential integration processes of matching unknown visual stimuli to visual memory as suggested in the literature have not kept our participants from naming scrambled pictures faster than normal pictures without context. We thus argue that the high amplitude in VEPs of scrambled pictures might also be driven by seeing an unknown and therefore highly unexpected stimulus. Since only 30 out of 234 trials (<13%) presented scrambled pictures, participants likely did not familiarize themselves with this type of stimuli. To compare, a standard oddball paradigm usually uses an oddball stimuli appearance of about 20%.

Previous research has linked the P2 component of VEPs to the ease of lexical selection (Rabovsky et al., 2021; Strijkers et al., 2010). This explanation alone, however, cannot directly account for our results. Scrambled pictures evoked the highest P2 amplitude, but revealed the second shortest picture naming times across conditions, being named with the high-frequency word 'nothing', which should be easily accessible. Therefore, the high P2 amplitudes in response to scrambled pictures in our study might simply be due more to visual than lexical aspects, whereas the modulations of the P2 component for the real objects could be explained by ease of lexical selection (see below). It should not be disregarded that the P2 component in picture naming studies is first of all induced by a visual stimulus.

Attention has previously been mentioned as a potential confound of the P2 effects in word production (Strijkers et al., 2010). Larger P2 amplitudes for rare or salient over normal stimuli would again be in line with our argumentation for the high P2 amplitudes in response to scrambled pictures in our study. In another study including similar control conditions to ours, scrambled pictures were also named significantly faster than normal pictures (Forseth et al., 2018). Altogether, these findings are an important reminder that it may sometimes be difficult to draw a one-to-one mapping between an ERP component and a cognitive process (e.g., the P2 component in word production uniquely reflecting lexical selection).

In sum, object recognition (and stereotypical naming) for scrambled pictures is faster than for real pictures. However, the visual stimuli are new to the viewer and thus likely elicit high VEP amplitudes, especially during the P2 component.

4.2.2 Conceptual-Lexical Retrieval (novel contrast 2)

Another interesting aspect of the VEP differences across conditions is the picture appearance after constrained sentences. Here, the amplitude is lowest of all four conditions for all three components. This is in line with previous research on repetition suppression, showing that amplitudes for repeated pictures are lower compared to the first time of seeing a picture (Gruber et al., 2004; Gruber & Müller, 2002, 2005; Rugg et al., 1995). Brain responses to constrained picture appearance thus mimic a behavior of repetition suppression, acting as if the picture had been seen before, as aspects of it (abstract, visually specific, etc.) are pre-activated. In other words, as the constrained sentences funnel onto more specific representations, participants might not only retrieve a word and prepare to say it, but also imagine the target picture already before it appears. One might argue that this amplitude reduction could also be a normal form of repetition suppression still from the picture familiarization before the experiment. However, every picture appeared once as a constrained picture and once as an unconstrained picture. Thus, if it was a repetition suppression from the picture familiarization, we should find the same effect for unconstrained pictures, which we do not.

These findings also match those of other studies linking the amplitude of the P2 to lexical retrieval (Fargier & Laganaro, 2017, 2020; Indefrey, 2011) and the ease thereof (Rabovsky et al., 2021; Strijkers et al., 2010). From the current and previous studies on contextually guided naming (Klaus et al., 2020;

Piai et al., 2014; Piai, Roelofs, Rommers, & Maris, 2015; Piai et al., 2017, 2018, 2020; Roos et al., 2023; Roos & Piai, 2020), we know that participants initiate word retrieval processes before the picture appears. This is also reflected in the picture naming times of the present study across conditions, showing that pictures following constrained sentences were named significantly faster than scrambled pictures as a low-level visual control condition. If participants first had to recognize the depicted object and retrieve the respective word for it, they would need more time to name a meaningful picture than a scrambled picture. This should lead to an easily accessible target word at picture onset (as it has already been retrieved previously), and thus to a lower P2 amplitude. Unlike the P2 results for scrambled pictures discussed above, those for constrained pictures are in line with research linking the P2 to lexical selection and its effort.

4.2.3 Visually Guided Naming (novel contrast 3)

Finally, the VEPs of bare and unconstrained picture naming conditions did not significantly differ after picture onset. This is in line with the behavioral results for these two conditions. As both conditions show pictures of normal objects, it does not seem to make a difference whether the lead-in sentence is reversed or meaningful; neither for the VEP, nor for the naming times.

In fact, the same was observed in a previous case study of a 23-year old person with aphasia due to extensive left hemisphere damage (Chupina et al., 2022). Falling exactly into the age range of the participants of the present study, the naming results of this participant showed no difference between unconstrained and bare picture naming (without preceding reversed speech). This reinforces the idea that picture naming in unconstrained sentences is highly comparable to bare picture naming following reversed sentences, strongly suggesting that unconstrained naming is a well controlled and valid baseline contrast for constrained naming. Given the same behavioral as well as ERP outcomes for these two conditions, we discuss the potential omission of bare picture naming in future versions of the CLaP in section 4.4.

4.3 Context Effect (validation contrast 3, novel contrast 1)

Our results replicated the strong behavioral context effect in picture naming times between constrained and unconstrained sentences (Klaus et al., 2020; Roos et al., 2023; Roos & Piai, 2020). We also replicated the power decreases between constrained and unconstrained sentences in the alpha-beta frequency range before picture onset over left-hemisphere channels associated with word retrieval (Piai et al., 2014; Piai, Roelofs, Rommers, & Maris, 2015; Piai et al., 2017, 2018, 2020). This effect is already visible just after sentence offset, if not earlier, and intensifies towards picture onset (*validation contrast 3*)

Adding reversed speech as a low-level baseline condition to the paradigm enabled us to further characterize the dynamics of the power decreases during the pre picture interval of the context effect. Our previous studies without such a baseline could not inform us about the direction of alpha-beta power changes for constrained and unconstrained sentences. Here, we demonstrated that power in the alpha-beta range decreases towards picture onset when preceded by constrained sentences. While power did seem to increase after unconstrained sentences, these power increases happened more around sentence offset rather than towards picture onset as in constrained sentences (*novel contrast 1*). Even without contrasting with meaningful unconstrained sentences as in previous studies, we can capture the alpha-beta power context effect by contrasting constrained sentences with a simple low-level auditory baseline in the form of reversed speech sentences.

4.4 Adaptations of the CLaP

The CLaP yields a lot of flexibility and can be modified to adapt to one's preference or provide different contrasts of interest while the overall framework of conditions and trials stays the same. In the introduction we already mentioned studies arguing that time-reversed speech provides a less-optimal baseline for speech processing (Narain et al., 2003; Stoppelman et al., 2013). To counteract this issue,

one could readily substitute the time-reversed speech condition with a different baseline for speech processing, such as noise vocoded speech, for example.

Another adaptation could be replacing the auditory comprehension part with reading comprehension. For this one could use sentences presented word-by-word as we have previously done (e.g., Roos et al., 2023; Roos & Piai, 2020). Reversed speech sentences in this case could be replaced by scrambled word sentences. This version of the CLaP would then yield measures of word reading contrasted with non-word reading.

Due to the similar outcomes of the bare and unconstrained picture naming conditions in behavioral as well as ERP results, there is the possibility of omitting bare picture naming from the paradigm. In this case, however, one might lose the real effect of scrambled picture naming, as reversed speech sentences would then always be followed by scrambled pictures. Thus, participants would not have to wait for the picture after reversed sentences to appear and recognize whether it is a real or a scrambled image. Instead they could already prepare for a scrambled picture to appear and retrieve the word necessary to name these.

At the same time, if participants are explicitly instructed that reversed sentences are *always* followed by a scrambled picture, this could lead to an expectation effect. Similarly as constrained sentences pre-activate the concept completing the sentence, reversed sentences could induce the early preparation of the stereotypical utterance for *nothing* (or the word they are instructed to use for scrambled pictures). Thus, this omission could yield a potentially interesting control contrast to constrained picture naming (see also Piai, Roelofs, Rommers, Dahlsätt, et al., 2015).

Last but not least the CLaP can also readily be used with different imaging modalities than EEG. We have previously done so using a contextually driven naming task in combination with magnetoencephalography (e.g., Roos & Piai, 2020) and fMRI (Roos et al., 2023). Regarding a combination of the CLaP with hemodynamic imaging methods, one might have to extend and jitter the duration of the pre-picture interval (see Roos et al., 2023). For the comprehension aspect, one could integrate over a larger time window for auditory or written sentence comprehension, instead of looking at millisecond time resolution during the sentence.

4.5 Strengths and limitations

The CLaP presents a well-controlled framework to investigate processes of language comprehension as well as production. As all trials follow the same task instructions with a simple trial structure across conditions, we minimize the risk to capture task- or condition-specific confounds between trials. Our paradigm also includes a more naturalistic setting of language in terms of context-driven picture naming. However, as with many language studies, the group effects we present here might not be present in all tested subjects, and future studies could focus on examining prevalence measures instead (e.g., Ince et al., 2021). Also, for the purpose of the present study we exclusively tested this paradigm with a group of young healthy participants aged between 18 and 28 (as is commonly done in the literature), which is not a sample representative enough for generally benchmarking the effects of a new paradigm.

Regarding the comprehension contrast of our paradigm, we would like to note that our analyses for this were rather simplistic, serving as a first step. In our own future studies using the CLaP, we are planning on investigating the comprehension part of the paradigm using source localization. As the present study did not include participant-specific MRI scans, we were not able to perform this type of analysis. If one is interested in a more detailed investigation of the auditory sentence contrast, this could possibly be done with analysis approaches as used in naturalistic spoken language comprehension (Zioga et al., 2023).

5. Conclusion

The present study introduces the CLaP as a new concise language paradigm to investigate aspects of language comprehension and production within the same trial. We provided a first validation of the CLaP for measures of comprehension and production, and reported on novel, theoretically relevant findings characterizing the language system, here with a particular focus on spoken word production. We propose several optional adaptations to further improve the CLaP and potentially use it with different neurotypical and neurological populations. The conciseness and richness, as well as the flexibility in adapting the paradigm to different versions, yield promising outlooks to further investigate the relationship between language comprehension and production in neurotypical as well as neurological populations using one's own desired version of the CLaP framework.

6. Statements & Declarations

6.1 Funding

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

6.2 Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

6.3. Author Contributions

All listed authors made substantial contributions to the conception or acquisition of the work, the analysis and interpretation of data; revised the work critically for important intellectual content, approved the version to be published, and agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. See below for detailed author contributions.

Natascha M Roos: Conceptualization, Methodology, Software, Formal analysis, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization, Project administration

Julia Chauvet: Data Collection, Data Cleaning, Writing - Review & Editing

Vitória Piai: Conceptualization, Methodology, Writing - Review & Editing, Supervision

6.4 Data Availability

All data and code are available via the Donders Repository (<https://doi.org/10.34973/19gn-7v46>).

6.5 Ethics Approval

This study was approved by the Ethics Committee of Social Sciences at Radboud University, following the Code of Ethics of the World Medical Association (Declaration of Helsinki).

6.6 Consent

Informed consent was obtained from all individual participants included in the study.

Acknowledgements

We would like to thank Nikita Boers and Rosemarije Weterings for their assistance during EEG testing sessions, as well as the M/EEG community at the Donders Institute, especially Robert Oostenveld, for helpful discussions and feedback regarding this work. The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

7. References

- Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S. F., Springer, J. A., Kaufman, J. N., & Possing, E. T. (2000). Human Temporal Lobe Activation by Speech and Nonspeech Sounds. *Cerebral Cortex*, 10(5), 512–528.
<https://doi.org/10.1093/cercor/10.5.512>
- Boersma, P., & Weenink, D. (2017). Praat, software for speech analysis and synthesis. *University of Amsterdam*.
- Boulenger, V., Hoen, M., Jacquier, C., & Meunier, F. (2011). Interplay between acoustic/phonetic and semantic processes during spoken sentence comprehension: An ERP study. *Brain and Language*, 116(2), 51–63.
<https://doi.org/10.1016/j.bandl.2010.09.011>
- Brodeur, M. B., Dionne-Dostie, E., Montreuil, T., & Lepage, M. (2010). The Bank of Standardized Stimuli (BOSS), a New Set of 480 Normative Photos of Objects to Be Used as Visual Stimuli in Cognitive Research. *PLoS ONE*, 5(5).
<https://doi.org/10.1371/journal.pone.0010773>
- Brown, E. C., Muzik, O., Rothermel, R., Matsuzaki, N., Juhász, C., Shah, A. K., Atkinson, M. D., Fuerst, D., Mittal, S., Sood, S., Diwadkar, V. A., & Asano, E. (2012). Evaluating reverse speech as a control task with language-related gamma activity on electrocorticography. *NeuroImage*, 60(4), 2335–2345.
<https://doi.org/10.1016/j.neuroimage.2012.02.040>
- Chupina, I., Sierpowska, J., Zheng, X. Y., Dewenter, A., Piastra, M.-C., & Piai, V. (2022). Time course of right-hemisphere recruitment during word production following left-hemisphere damage: A single case of young stroke. *European Journal of Neuroscience*, 56(8), 5235–5259. <https://doi.org/10.1111/ejn.15813>
- Fargier, R., & Laganaro, M. (2017). Spatio-temporal Dynamics of Referential and Inferential Naming: Different Brain and Cognitive Operations to Lexical Selection. *Brain Topography*, 30(2), 182–197. <https://doi.org/10.1007/s10548-016-0504-4>
- Fargier, R., & Laganaro, M. (2020). Neural dynamics of the production of newly acquired words relative to well-known words. *Brain Research*, 1727, 146557.
<https://doi.org/10.1016/j.brainres.2019.146557>
- Forseth, K. J., Kadipasaoglu, C. M., Conner, C. R., Hickok, G., Knight, R. T., & Tandon, N. (2018). A lexical semantic hub for heteromodal naming in middle fusiform gyrus. *Brain*, 141(7), 2112–2126. <https://doi.org/10.1093/brain/awy120>
- Gruber, T., Malinowski, P., & Müller, M. M. (2004). Modulation of oscillatory brain activity and evoked potentials in a repetition priming task in the human EEG. *European Journal of Neuroscience*, 19(4), 1073–1082. <https://doi.org/10.1111/j.0953-816X.2004.03176.x>

- Gruber, T., & Müller, M. M. (2002). Effects of picture repetition on induced gamma band responses, evoked potentials, and phase synchrony in the human EEG. *Cognitive Brain Research*, 13(3), 377–392. [https://doi.org/10.1016/S0926-6410\(01\)00130-6](https://doi.org/10.1016/S0926-6410(01)00130-6)
- Gruber, T., & Müller, M. M. (2005). Oscillatory Brain Activity Dissociates between Associative Stimulus Content in a Repetition Priming Task in the Human EEG. *Cerebral Cortex*, 15(1), 109–116. <https://doi.org/10.1093/cercor/bhh113>
- Hustá, C., Zheng, X., Papoutsis, C., & Piai, V. (2021). Electrophysiological signatures of conceptual and lexical retrieval from semantic memory. *Neuropsychologia*, 161, 107988. <https://doi.org/10.1016/j.neuropsychologia.2021.107988>
- Ince, R. A., Paton, A. T., Kay, J. W., & Schyns, P. G. (2021). Bayesian inference of population prevalence. *ELife*, 10, e62461. <https://doi.org/10.7554/eLife.62461>
- Indefrey, P. (2011). The Spatial and Temporal Signatures of Word Production Components: A Critical Update. *Frontiers in Psychology*, 2. <https://www.frontiersin.org/articles/10.3389/fpsyg.2011.00255>
- Jung, T.-P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., & Sejnowski, T. J. (2000). Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. *Clinical Neurophysiology*, 111(10), 1745–1758. [https://doi.org/10.1016/S1388-2457\(00\)00386-2](https://doi.org/10.1016/S1388-2457(00)00386-2)
- Klaus, J., Schutter, D. J. L. G., & Piai, V. (2020). Transient perturbation of the left temporal cortex evokes plasticity-related reconfiguration of the lexical network. *Human Brain Mapping*, 41(4), 1061–1071. <https://doi.org/10.1002/hbm.24860>
- Logothetis, N. K., & Wandell, B. A. (2004). Interpreting the BOLD signal. *Annual Review of Physiology*, 66, 735–769. <https://doi.org/10.1146/annurev.physiol.66.082602.092845>
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164(1), 177–190. <https://doi.org/10.1016/j.jneumeth.2007.03.024>
- Narain, C., Scott, S. K., Wise, R. J. S., Rosen, S., Leff, A., Iversen, S. D., & Matthews, P. M. (2003). Defining a Left-lateralized Response Specific to Intelligible Speech Using fMRI. *Cerebral Cortex*, 13(12), 1362–1368. <https://doi.org/10.1093/cercor/bhg083>
- Neurobehavioral Systems. (n.d.). Retrieved March 27, 2020, from https://www.neurobs.com/menu_support/menu_archives/citation_guide
- Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience*, 2011, 1–9. <https://doi.org/10.1155/2011/156869>
- Piai, V., Klaus, J., & Rossetto, E. (2020). The lexical nature of alpha-beta oscillations in context-driven word production. *Journal of Neurolinguistics*, 55, 100905.

- <https://doi.org/10.1016/j.jneuroling.2020.100905>
- Piai, V., Meyer, L., Dronkers, N. F., & Knight, R. T. (2017). Neuroplasticity of language in left-hemisphere stroke: Evidence linking subsecond electrophysiology and structural connections. *Human Brain Mapping, 38*(6), 3151–3162.
<https://doi.org/10.1002/hbm.23581>
- Piai, V., Roelofs, A., & Maris, E. (2014). Oscillatory brain responses in spoken word production reflect lexical frequency and sentential constraint. *Neuropsychologia, 53*, 146–156. <https://doi.org/10.1016/j.neuropsychologia.2013.11.014>
- Piai, V., Roelofs, A., Rommers, J., Dahlsätt, K., & Maris, E. (2015). Withholding planned speech is reflected in synchronized beta-band oscillations. *Frontiers in Human Neuroscience, 9*. <https://www.frontiersin.org/articles/10.3389/fnhum.2015.00549>
- Piai, V., Roelofs, A., Rommers, J., & Maris, E. (2015). Beta oscillations reflect memory and motor aspects of spoken word production. *Human Brain Mapping, 36*(7), 2767–2780. <https://doi.org/10.1002/hbm.22806>
- Piai, V., Rommers, J., & Knight, R. T. (2018). Lesion evidence for a critical role of left posterior but not frontal areas in alpha–beta power decreases during context-driven word production. *European Journal of Neuroscience, 48*(7), 2622–2629. <https://doi.org/10.1111/ejn.13695>
- R. Core Team. (2017). *R: A language and environment for statistical computing*.
- Rabovsky, M., Schad, D. J., & Abdel Rahman, R. (2021). Semantic richness and density effects on language production: Electrophysiological and behavioral evidence. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 47*(3), 508–517. <https://doi.org/10.1037/xlm0000940>
- Rizkallah, J., Benquet, P., Kabbara, A., Dufor, O., Wendling, F., & Hassan, M. (2018). Dynamic reshaping of functional brain networks during visual object recognition. *Journal of Neural Engineering, 15*(5), 056022. <https://doi.org/10.1088/1741-2552/aad7b1>
- Roos, N. M., & Piai, V. (2020). Across-session consistency of context-driven language processing: A magnetoencephalography study. *European Journal of Neuroscience, 52*(5), 3457–3469. <https://doi.org/10.1111/ejn.14785>
- Roos, N. M., Takashima, A., & Piai, V. (2023). Functional neuroanatomy of lexical access in contextually and visually guided spoken word production. *Cortex, 159*, 254–267. <https://doi.org/10.1016/j.cortex.2022.10.014>
- Rugg, M. D., Soardi, M., & Doyle, M. C. (1995). Modulation of event-related potentials by the repetition of drawings of novel objects. *Cognitive Brain Research, 3*(1), 17–24. [https://doi.org/10.1016/0926-6410\(95\)00014-3](https://doi.org/10.1016/0926-6410(95)00014-3)
- Stojanoski, B., & Cusack, R. (2014). Time to wave good-bye to phase scrambling: Creating

- controlled scrambled images using diffeomorphic transformations. *Journal of Vision*, 14(12), 6–6. <https://doi.org/10.1167/14.12.6>
- Stoppelman, N., Harpaz, T., & Ben-Shachar, M. (2013). Do not throw out the baby with the bath water: Choosing an effective baseline for a functional localizer of speech processing. *Brain and Behavior*, 3(3), 211–222. <https://doi.org/10.1002/brb3.129>
- Strijkers, K., Costa, A., & Thierry, G. (2010). Tracking Lexical Access in Speech Production: Electrophysiological Correlates of Word Frequency and Cognate Effects. *Cerebral Cortex*, 20(4), 912–928. <https://doi.org/10.1093/cercor/bhp153>
- van Casteren, M., & Davis, M. H. (2006). Mix, a program for pseudorandomization. *Behavior Research Methods*, 38(4), 584–589. <https://doi.org/10.3758/BF03193889>
- Zioga, I., Weissbart, H., Lewis, A. G., Haegens, S., & Martin, A. E. (2023). Naturalistic Spoken Language Comprehension Is Supported by Alpha and Beta Oscillations. *Journal of Neuroscience*, 43(20), 3718–3732. <https://doi.org/10.1523/JNEUROSCI.1500-22.2023>