Early goal-directed top-down influences in the production of speech

Kristof Strijkers[1]*, Yen Na Yum[2], Jonathan Grainger[3] and Phillip J. Holcomb[2]

[1] Department of Psicologia Bàsica, Universitat de Barcelona, Barcelona, Spain
[2] Department of Psychology, Tufts University, Medford, MA, USA
[3] CNRS, Aix-Marseille University, Marseille, France

It was recently reported that the conscious intention to produce speech affects the speed with which lexical information is retrieved upon presentation of an object (Strijkers et al., 2011). The goal of the present study was to elaborate further on the role of these top-down influences in the course of planning speech behavior. In an event-related potentials (ERP) experiment, participants were required to overtly name pictures and words in one block of trials, while categorizing the same stimuli in another block of trials. The ERPs elicited by the naming task started to diverge very early on (∼170 ms) from those elicited by the semantic categorization task. Interestingly, these early ERP differences related to task intentionality were identical for pictures and words. From these results we conclude that (a) in line with Strijkers et al. (2011), goal-directed processes play a crucial role very early on in speech production, and (b) these task-driven top-down influences function at least in a domain-general manner by modulating those networks which are always relevant for the production of language, irrespective of which cortical pathways are triggered by the input.

Keywords: speech production, goal-directed processing, lexical access, ERPs, categorization

INTRODUCTION

A substantial body of research in the last two decades has provided compelling evidence that top-down processes can have a powerful influence on certain early low-level perceptual processes (e.g., Corbetta et al., 1990; Posner and Dehaene, 1994; Desimone and Duncan, 1995; Luck et al., 1997; Hillyard et al., 1998; Kastner et al., 1999; Bar, 2003; Bar et al., 2006; Gilbert and Sigman, 2007). These findings have been crucial with respect to the way we conceptualize visual processing, in that the traditional automatic bottom-up view has changed dramatically when taking into account goal-directed behavior related to a visual act. That is, recognizing visual input does not solely proceed in a unidirectional manner from lower to higher levels of representation, but rather is achieved through the dynamical interplay of stimulus-driven processes with early top-down influences facilitating access to those representations that are relevant to the desired behavior. However, the role top-down processes play in regulating higher goal-directed behaviors, such as producing speech, has not received nearly as much attention. This would appear to be an important issue to explore though, since speaking is in principle an intention-driven activity. A speaker knows beforehand, at least most of the time, whether he/she wants to verbalize or not certain ideas, thoughts, and stimuli in his environment. Consequently, it seems reasonable to assume that these *a priori* goal-directed settings should exert an important influence on the way speech planning proceeds. Nevertheless, most speech production models assume that the influence of task intentions come about relatively late, namely after word selection (e.g., Dell, 1986; Caramazza, 1997; Levelt et al., 1999). Such a view is clearly in disagreement with the more dynamical accounts of brain processing that are emerging in the fields of vision and object

recognition. One of the reasons for this discrepancy might be that the temporal role of top-down influences of task intentionality has never been systematically investigated for language behavior. To this end, we aimed at increasing our understanding of when goal-directed influences associated with speech production affect the course of processing. In particular, we compared the brain's electrical response (event-related potentials; ERPs) in a task where one has the conscious intention to articulate the name of the stimulus (naming task) vs. a task where no naming intention is present semantic categorization task (SCT) both for non-linguistic stimuli (pictures) and linguistic stimuli (written words). By doing so, we aimed to see when the brain starts differentiating the same visual input as a function of task intention, and whether these top-down influences vary depending on the type of processing required for different input modalities.

There is a long-standing tradition in the field of language production that has explored whether items that a speaker does not intend to name nevertheless activate lexical representations. Within this tradition there have been a large number of studies investigating whether distractor items (items we do not wish to utter) affect the speed with which we name a target item, as well as a smaller number of studies assessing the presence of linguistic effects in non-verbal tasks (e.g., Kroll and Potter, 1984; Glaser and Glaser, 1989; Schriefers et al., 1990; Levelt et al., 1991; Roelofs, 1992, 2003, 2008; Jescheniak and Levelt, 1994; Starreveld and La Heij, 1996; Meyer et al., 1998, 2007; Cutting and Ferreira, 1999; Costa and Caramazza, 2002; Morsella and Miozzo, 2002; Navarrete and Costa, 2005; Bles and Jansma, 2008). However, unlike the current study, these studies did not aim at uncovering how goal-directed speech behavior influences the course of processing

within the system. Instead, they sought to assess whether activated concepts automatically transmit information to the lexical system and up to which level of representations this spreading activation extends. For instance, in linguistic Stroop-like tasks it is frequently reported that distractors having a semantic or phonological relationship with a target affect the speed with which that target is named. From these findings it has been concluded that both speech intended and speech non-intended concepts activate the lexicon in parallel. Such a conclusion is in accordance with models of lexical access that embrace the principle of spreading activation, according to which activated concepts percolate to the lexical system regardless of the intention of a speaker (e.g., Dell, 1986; Caramazza, 1997; Levelt et al., 1999; but see Bloem and La Heij, 2003). However, what has not been addressed is when and how intentional processes do come into play in order to eventually produce the desired behavior. In other words, previous comparisons between items we intend to verbalize and items we do not intend to verbalize have mainly served the purpose of improving our understanding of how concepts activate lexical representations (and beyond) during language production, but this line of research did not aim to uncover how this lexicalization process or even earlier processes leading up to lexical access may interact with task intentional processing. Therefore, in order to develop accounts of language production which do not solely accommodate language-related processing, but instead try to incorporate this within the broader spectrum of human information processing, it is important to study how language processing progresses in relation to goal-directed behavior.

Recently, Strijkers et al. (2011) investigated this question explicitly and reported evidence that the top-down intention to speak seems to affect the language system in a proactive manner; that is, prior to the spreading activation between concepts and words. Specifically, these authors compared the brain's electrical response for a variable known to affect lexical access, namely word frequency, during overt object naming and non-verbal object categorization. They found that during naming, ERPs elicited for objects with low frequency names started to diverge from those with high frequency names as early as 152 ms after stimulus onset (pP2), while during non-verbal categorization the same frequency comparison appeared 200 ms later at a qualitatively distinct component (N400). Two important conclusions were drawn from these findings: first, in line with spreading activation models of lexical access, activation in semantic representations percolates to the lexical system regardless of a speaker's intention; Second, and in contrast to the predictions of most language production models, initial access to the lexicon is instigated by the top-down intention to speak. Put differently, when there is conscious intention to name an object the brain will engage substantially faster in lexical access compared to when no such intention is present due to top-down signals pre-activating the lexical system prior to the conceptually driven activation of words. To our knowledge, this was the first study to directly demonstrate the vital and early role top-down processes play in facilitating the retrieval of words one intends to utter. This result places novel constraints on language production models in that initial access to the lexicon from *activated* concepts is not as automatic as originally thought, but can be better seen as a dynamical process driven by goal-directed intentions.

Here we aimed to extend the findings of Strijkers et al. (2011) by examining when the goal-directed operations determined by task intention penetrate stimulus-evoked processing. Like Strijkers et al. (2011) we compared the time-course of ERP effects in a task requiring speech production (naming) vs. a task where speech is not necessary (semantic categorization). However, rather than exploring how the intention to speak may interact with a particular production operation, here we sought to directly ascertain the temporal dissociations related to task intention. Exploring the electrophysiological signature of task intention in isolation, should enable us to pinpoint when the most evident ERP deflections associated with naming and categorization intention occur. This is relevant for two reasons: first, it can provide an independent test of the conclusions reached by Strijkers et al. (2011). In that study, the authors argued that top-down intention to speak penetrates the lexico-semantic system prior to concept selection. If this conclusion is correct, ERP differences elicited by task intention in general (naming vs. categorization) should occur prior to, or in the temporal vicinity of the time-course uncovered by Strijkers and colleagues (i.e., around 150–200 ms after stimulus onset). Second, given that no particular linguistic operation such as lexical access is targeted, we will be able to provide a more general assessment of the processing differences between speech production and semantic categorization. In this way, potentially important ERP effects other than those that are lexically driven (for instance, already during visual and/or conceptual processing) may become apparent.

An important aspect of the current experiment is that we also explored whether top-down involvement differs as a function of presentation modality. By comparing effects across modalities we should be able to ascertain whether the top-down intention to speak affects the language network in a modality-general manner or whether this process elicits distinct modulations depending on the input. Compared to picture naming, where the production of speech entails active retrieval from memory, word naming directly conveys the linguistic information which has to be uttered (i.e., written words are automatically associated with the required output). As a consequence, early top-down processes that facilitate lexical retrieval may be especially relevant for more demanding processing situations such as picture naming compared to the more predominantly stimulus-driven processing associated with word naming. Given these differences in linguistic vs. non-linguistic input, and consequently, differences in the dynamics for retrieving the same words between the two tasks, it will be interesting to see whether goal-directed influences display the same or distinct time-courses.

In two blocks of trials undergraduate participants were presented with a combination of black and white pictures of common objects (picture condition) and the written words that were the names of objects (word condition). In one block participants were asked to rapidly name (naming task block), and in another block they were asked to categorize (SCT) the words or objects. ERPs time-locked to the onset of words and pictures in both blocks were recorded along with the overt response. This approach has only recently been employed successfully to study language production (e.g., Christoffels et al., 2007; Koester and Schiller, 2008; Chauncey et al., 2009; Costa et al., 2009; Dell'Acqua et al., 2010; Strijkers

et al., 2010). The design was a two (Modality: words vs. pictures) by two (Task: naming vs. semantic categorization) factorial. In light of the results of Strijkers et al. (2011) we predicted electrophysiological deflections within 200 ms of stimulus onset when depicted objects had to be named compared to being categorized. For printed word naming, if the top-down mechanisms related to task intention are qualitatively similar across modalities and processing dynamics, we expect to see the same early, and perhaps even slightly earlier, ERP dissociations between naming and categorization as for pictures. If, however, the influences generated by the higher goal-directed processes operate as a function of a particular input and/or whether the task-relevant representations have to be retrieved internally (pictures), then distinct and potentially later ERP modulations for written words compared to pictures might be obtained.

MATERIALS AND METHODS

PARTICIPANTS
Twelve native English speakers (six females, mean age 18.67 years) participated and were compensated for their time. All participants were right-handed, with normal or corrected-to-normal visual acuity and no history of neurological insult or language disability.

STIMULI
The picture stimuli consisted of 184 black and white line drawings of common objects, selected from a standardized inventory (Snodgrass and Vanderwart, 1980). The word stimuli were 184 English words that corresponded to the line drawings. Of these, 24 words and 24 images were so-called "probe" items. The ERPs to probe items were not analyzed. All stimuli were presented in white on a black background. Both pictures and words were presented together in two mixed blocks (see below), arranged in a pseudo-random order to prevent expectation and priming effects.

PROCEDURE
Participants were seated in a comfortable armchair facing a computer monitor in a sound-attenuated room for electrode placement. Each trial started with a fixation cross in the middle of the screen for 500 ms and a blank screen for 500 ms. The stimulus (a

picture or a word) then appeared for 400 ms, followed by a 1100 ms blank screen, and a blink signal for 1700 ms. This was followed by another blank screen for 500 ms and the fixation cross for the next trial (see **Figure 2** for examples of both trial types). Participants were asked to blink during the blink signal if necessary, and minimize eye movements for the rest of the time. There were two scheduled 1-min breaks during each block of the experiment.

In the first block participants engaged in a go/no–go SCT, in which they were instructed to press a button whenever they saw either a picture or a word referring to a human body part (so-called "probe" items). Probes made up 12% of trials with equal numbers of picture and word body parts. No response was required to the remaining 88% of critical stimuli which were averaged in the ERPs reported here.

In a second block of trials participants were told to rapidly name all picture and word stimuli. The exact same trial structure was used in this second block, although the word and picture stimuli were switched such that all items that were formerly pictures were now words and vice versa. Stimuli that appeared in picture format for one participant for a given task appeared in word format for another participant, and vice versa, and the order of the two blocks (semantic categorization and naming) was counterbalanced across participants.

EEG RECORDINGS
Electroencephalograms were collected using 32-channel caps (Electro-cap International). The tin electrodes were arranged according to International 10–20 system (see **Figure 1**). In addition, an electrode below the left eye (LE) was used to monitor for blinks and vertical eye movements and an electrode beside the right eye (HE) was used to monitor for horizontal eye movements. Two electrodes were placed behind the ears on the mastoid bone; the left mastoid site (A1) was used as a reference for all electrodes, and the right mastoid site (A2) was recorded to evaluate differential mastoid activity. Impedance was kept at less than 5 kΩ for all electrode sites except the lower eye channel, which was below 10 kΩ. The EEG was amplified using an SA Bioamplifier (SA Instruments, San Diego, CA, USA) operating on a bandpass of 0.01 and 40 Hz. The digitizing computer continuously sampled
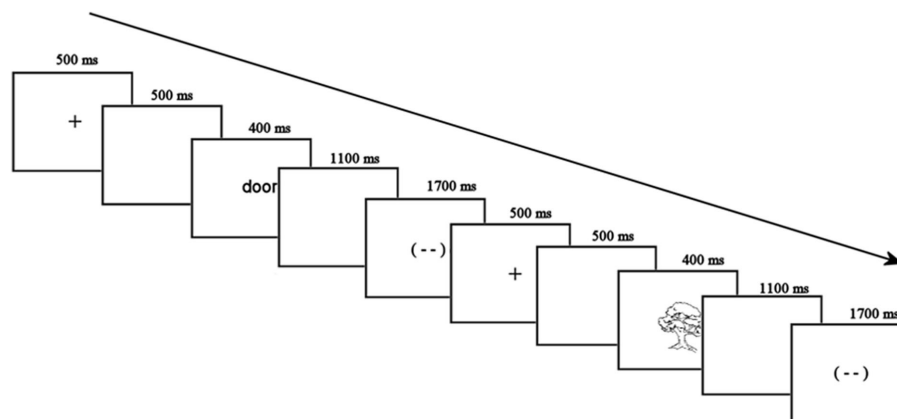

FIGURE 1 | Two sample trials, one with a word stimulus and one with a picture stimulus.

the EEG at a rate of 200 Hz while a stimulus computer simultaneously presented stimuli to a 19″ CRT monitor located 54″ in front of the participant (all stimuli subtended less than 7° of horizontal visual angle).

DATA ANALYSIS

Averaged ERPs were computed for all word and picture stimuli for each participant in each block of trials (task) at the 29 scalp sites shown in **Figure 2**. Epochs with eye movement artifacts between −100 and 600 ms post stimulus onset were excluded prior to averaging. The resulting ERPs were baselined between −100 and 0 ms and re-referenced to the average of the 29 scalp sites (average reference – Joyce and Rossion, 2005). The resulting ERP data were measured by calculating mean amplitudes within three latency windows: 150–250 ms and 250–350 ms for picture and word stimuli and also between 350 and 450 ms for picture stimuli (the late window was not used with the word stimuli because a substantial number of trials in the naming task had significant speech artifact starting just after 350 ms post-word onset – see the anterior sites in **Figure 5**).

The data were analyzed with repeated measures ANOVAs with four independent variables: TASK (SCT vs. Naming), MODALITY (words vs. pictures), ANTERIOR–POSTERIOR (prefrontal vs. frontal vs. temporal–parietal vs. occipital for average reference data and frontal, vs. central vs. parietal vs. occipital for mastoid reference data), and LATERALITY (left vs. midline vs. right – see **Figure 2** for the electrode sites included in the analysis). Significant interactions involving the TASK and MODALITY factors were followed up with planned ANOVAs breaking down the interaction.
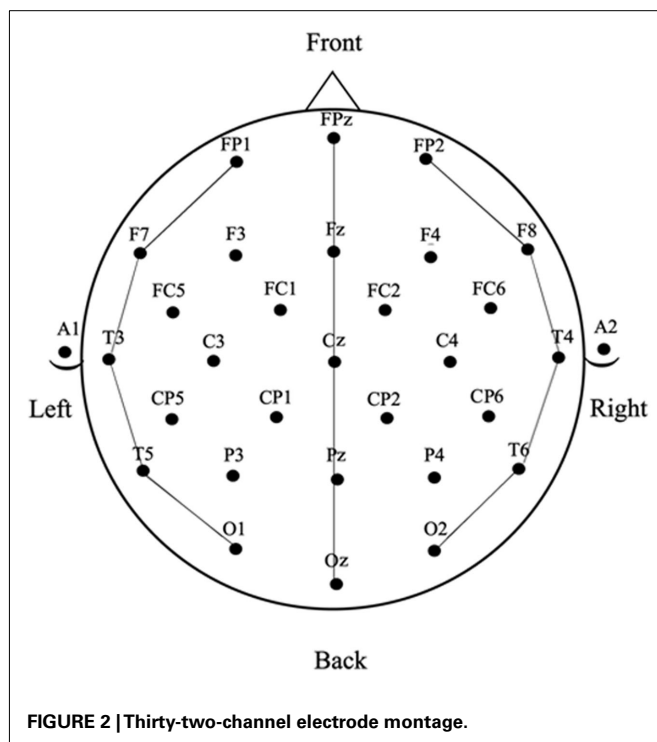
FIGURE 2 | Thirty-two-channel electrode montage.

RESULTS

BEHAVIORAL ANALYSES

The behavioral analyses were restricted to the naming data (given that for the categorization experiment the critical trials – no–go trials – did not require a behavioral response). Here it was observed that word naming resulted in faster reaction times (520 ms; SD: 69 ms) compared to picture naming (790 ms; SD: 78 ms). A paired t-test between word and picture naming confirmed that this difference was significant [$t(10) = 29.12$, $p < 0.001$].

ERP ANALYSES

150–250 ms. As can be seen in **Figure 3**, in this epoch there are substantial differences between ERPs to stimuli presented during a SCT compared to the same items presented during a naming task. This observation is supported by both a main effect of TASK [$F(1,11) = 5.41$, $p = 0.04$] as well as an interaction between TASK and the two scalp site variables [TASK × ANT–POST × LATERALITY: $F(8,88) = 4.39$, $p = 0.0002$]. **Figures 4** and **5** reveal that the effects of TASK differed as a function of MODALITY (words vs. pictures) and scalp site [TASK × MODALITY × ANT–POST × LATERALITY: $F(8,88) = 3.23$, $p = 0.0029$]. To better understand this interaction we conducted follow-up analyses examining the two modalities separately. In the analyses of ERPs recorded to picture stimuli we found significant differences between the TASKS as a function of scalp site [TASK × ANT–POST: $F(4,44) = 3.37$, $p = 0.0172$]. This interaction was due to the naming task producing more negative-going ERPs over posterior sites but more positive-going ERPs over the most anterior sites than the SCT (see the FP row of sites in **Figure 4**). It is important to note that there was no trend for this effect to differ as a function of laterality (i.e., TASK × ANT–POST × LATERALITY, and TASK × LATERALITY, ps > 0.1). For the word stimuli there was a similar pattern of posterior negativity and anterior positivity for naming compared to semantic categorization [TASK × ANT–POST: $F(4,44) = 3.66$, $p < 0.07$], but this trend tended to be greater over the left hemisphere and midline toward the back of the head [O1, Oz, and T5, TASK × ANT–POST × LATERALITY: $F(8,88) = 7.06$, $p < 0.0001$ – see voltage maps in the bottom left panel of **Figure 6**]. In summary, an N170-like negativity at posterior sites was larger for naming than semantic categorization, and this effect tended to be larger over the left than right hemisphere, but only for word stimuli.

250–350 ms. Differences between conditions continued into this second measurement window. There was not, however, a significant main effect of MODALITY as in the previous window, although this factor did interact with the TASK variable [TASK × MODALITY: $F(1,11) = 5.27 = 0.0423$]. There were also significant [TASK × MODALITY × LATERALITY and TASK × MODALITY × ANT–POST × LATERALITY interactions ($F(2,22) = 7.14$, $p = 0.0041$; $F(8,88) = 4.51$, $p = 0.0001$, respectively]. To better understand these interactions we conducted follow-up analyses for the two modalities separately. For the picture modality we found a significant interaction of the TASK variable with both of the scalp site variables [TASK × ANT–POST × LATERALITY: $F(8,88) = 3.33$, $p = 0.0022$]. Pictures in the naming task tended to produce more negative-going ERPs over left posterior and central sites, but more positive-going ERPs
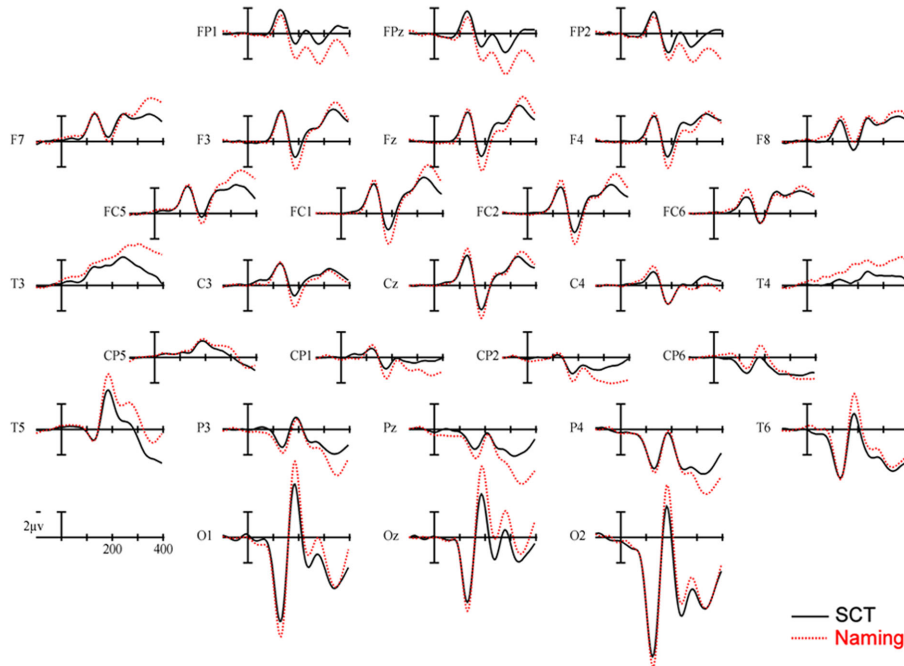
FIGURE 3 | Main effect of TASK. ERPs collapsed across item type. Solid lines are ERPs from all 28 scalp sites during a semantic categorization task (SCT) and dashed lines are ERPs recorded to the same stimuli in a naming task. Stimulus onset is the vertical calibration bar and each tic mark on the *x*-axis is 100 ms.
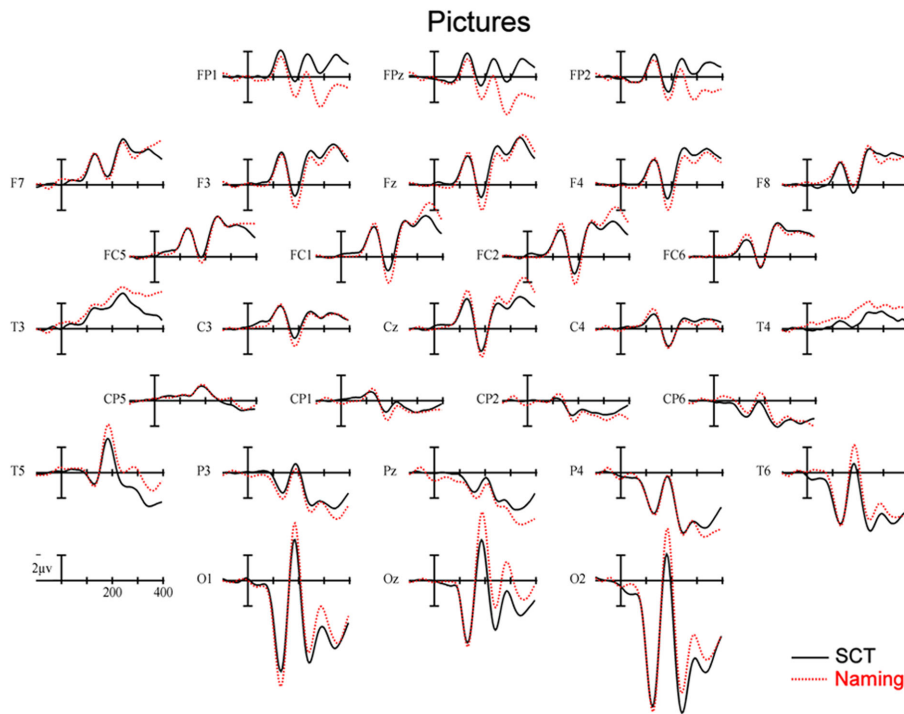
FIGURE 4 | Event-related potentials s recorded to pictures of objects, solid lines are ERPs during a semantic categorization task (SCT) and dashed lines are ERPs recorded in a picture naming task.
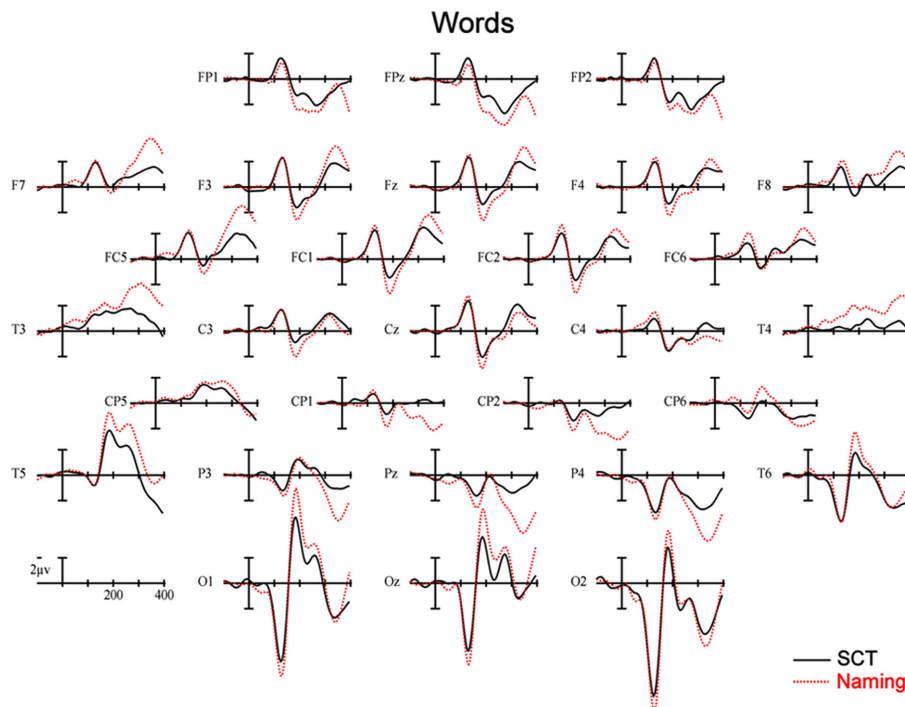
FIGURE 5 | Event-related potentials recorded to words, solid lines are ERPs during a semantic categorization task (SCT), and dashed lines are ERPs recorded in a word naming task.

over midline and left hemisphere sites than did ERPs to the same items during the SCT (see **Figure 4**). While the basic pattern of TASK effects was similar for words, the left lateralized nature of the posterior negativity and anterior positivity was stronger than for pictures [TASK × LATERALITY: $F(2,22) = 5.79$, $p = 0.012$; TASK × ANTERIOR–POSTERIOR × LATERALITY: $F(8,88) = 8.12$, $p < 0.00001$].

350–450 ms. As previously mentioned, because of the earlier onset of articulation in word naming, we only analyzed the ERP data for picture stimuli in this epoch. As can be seen in **Figure 4**, ERPs to pictures in this epoch tended to be more negative-going in the naming task than the SCT at most scalp sites with the notable exception of the most anterior electrodes (FP1, FPz, and FP2) and Pz. At these sites the naming tasks generated clearly more positive-going ERPs than did the SCT [TASK × ANT–POST × LATERALITY interaction: $F(8,88) = 4.29$, $p = 0.0002$]. As can be seen in **Figure 6**, the pattern for picture stimuli in this epoch is similar to the pattern seen in the previous epoch for words (compare **Figure 6** bottom right with top right).
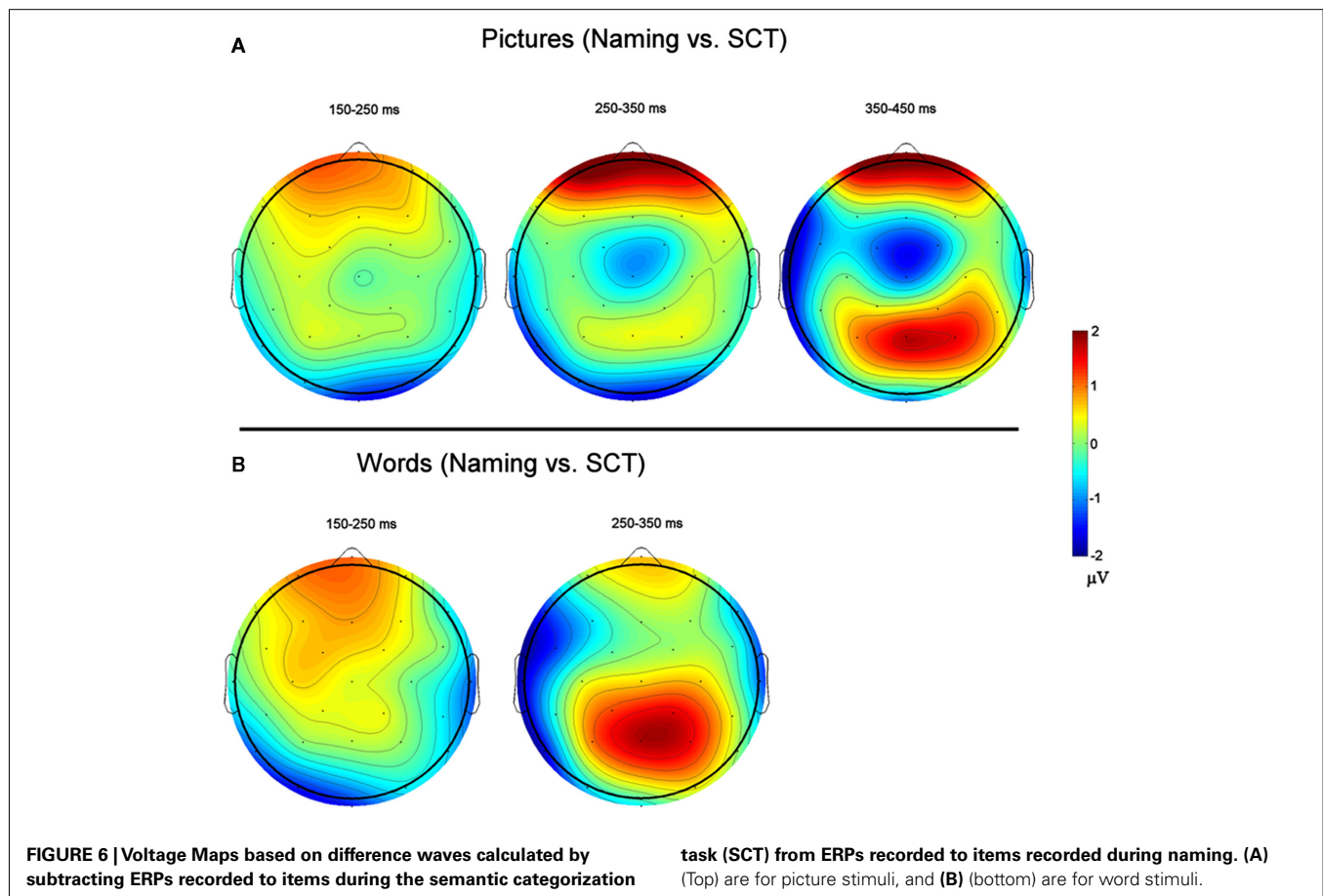
DISCUSSION

We found that ERPs elicited by the naming task started to diverge very early on (∼170 ms) from those elicited by the SCT, and these effects were qualitatively similar for pictures and words, aside from some hemispheric differences potentially reflecting the distinct visual input. Concretely, naming elicited enhanced negative brain responses peaking around 170 ms at occipital electrodes (N170) and enhanced positive amplitudes peaking around 200 ms at anterior sites (P2) compared to semantic categorization. In later

time-windows some expected differences between pictures and words became apparent: (a) picture naming implied additional frontal ERP enhancements, arguably reflecting a more effortful word retrieval for the former; (b) word naming resulted in an earlier onset of motor preparation than picture naming did, nicely mimicking the behavioral naming latencies. The results demonstrate that within 200 ms, top-down processes associated with task intentionality start influencing the course of brain processing. The data extends the increasing body of evidence from other modalities demonstrating that cognitive processing arises as a consequence of the early interplay between externally and internally driven processes (e.g., Desimone and Duncan, 1995; Ullman, 1995; Kastner and Ungerleider, 2000; Engel et al., 2001; Gilbert and Sigman, 2007) to the realm of language processing, and more specifically is in agreement with recent work arguing that the brain's fast engagement in lexical access is driven by the intention to speak (Strijkers et al., 2011).

In addition, the current results suggest that these early goal-directed modulations play a central part in the flow of activation regardless of the modality in which participants had to perform the task. This finding is consistent with the hypothesis that the mechanism behind these top-down influences is domain-general. That is, the manner in which the task intentional mechanism guides activity rapidly toward relevant representations is independent of the input processes[1]. A straightforward manner in which to interpret

---

[1]Note that, under the right conditions, this does not exclude the potential involvement of input specific top-down processes in speech production as well.

FIGURE 6 | Voltage Maps based on difference waves calculated by subtracting ERPs recorded to items during the semantic categorization task (SCT) from ERPs recorded to items recorded during naming. **(A)** (Top) are for picture stimuli, and **(B)** (bottom) are for word stimuli.

this finding is to assume that the intention to speak sets the brain in a general naming state, enhancing those representations which are always relevant for speech production regardless which cortical pathways are triggered by the input. A similar scenario unfolds for the SCT in that the system is tuned toward rapid activation of those representations that are crucial for classifying items as belonging to a particular category. In support of this notion is the fact that ERP task differences are similar for the two input modalities, even before the deflections associated with the different inputs become apparent. A final conclusion that can be derived from the similar early enhancements caused by task intention across input modalities is that top-down processes are not simply a helpful asset for the brain to calculate a motor program for speech production. Rather it suggests that early top-down processing forms an integral part of any speech production act. To summarize, the main contribution of the present study is the demonstration that the intentional act of producing speech requires the early involvement of domain-general top-down processes enhancing information relevant for producing speech irrespective of the input and the processing dynamics associated with that input. In what follows we will tentatively suggest potential loci where these goal-directed influences may take place in the course of speech planning.

A first potential locus for the early influence of top-down processes identified in the present study is at the lexical level. As mentioned in the Introduction, Strijkers and colleagues already

demonstrated that in picture naming, lexical access is facilitated in a proactive manner by the intention to speak. Given the overlapping time-course ($\sim$150–200 ms) between the ERP effects reported in that study and the current study, a similar conclusion as proposed by Strijkers et al. (2011) could be entertained here. In the case of naming, a task intentional mechanism will *a priori* increase the baseline activity of the lexical system so that when the actual stimulus is presented, there is facilitated and privileged access to word representations in order to rapidly and efficiently produce speech. In the case of semantic categorization, there is no need for top-down signals associated with task intention to enhance speech-related representations. Instead, goal-directed activity will "push" the sensory-driven processing of the incoming stimulus toward those semantic features that are relevant for deciding to which category the input belongs to, allowing us to categorize objects independently from the lexical information associated with that object (e.g., Dell'Acqua and Grainger, 1999).

Within this view, the ERP components we identified (or at least one of them), an anterior P2 and a posterior N170, can be conceived as reflecting the task intentional mechanism itself. That is, if the ERP modulations would reflect the top-down signals acting upon a particular process, lexical access in the present framework, we should have seen a similar pP2 (descriptively labeled *production* P2–pP2) modulation as reported by Strijkers et al. (2011), at least in the case of pictures. This posteriorly distributed positive-going

component, which has been shown to be sensitive to a range of lexical variables (Costa et al., 2009; Strijkers et al., 2010), was selectively modulated for naming intention (Strijkers et al., 2011). Strijkers et al. proposed that its modulation likely is engendered by the interaction between top-down processing and lexical activation levels. Here, no such pP2 modulation between naming and categorization was found, but instead a frontally maximal P2 and an occipital N170 were elicited. Thus, if we wish to maintain that these cortical deflections represent top-down influences which affect lexicalization, we must assume that the particular electrophysiological expression(s) observed here are associated with the top-down processes in isolation. If so, naming appears to require increased top-down involvement, which seems reasonable given that naming is a process requiring a much more specific behavioral response compared to semantic categorization.

Alternatively, the task driven ERP differences uncovered here may not be reflecting the same top-down signals as in the study of Strijkers et al. (2011). In that study, a particular linguistic stage was targeted by manipulating lexical frequency, whereas in the present study goal-directed influences might be acting upon other levels of processing. This conclusion finds support in a comparison of the effects seen in two early components affected by task in the present study, the anterior P2 and posterior N170, with the results from studies in other fields shown to modulate these components. In visual search paradigms the amplitude of the P2 is larger for attended stimuli and target-relevant information, which has been proposed to reflect attention-driven enhancements of the perceptually-relevant features of the input (e.g., Hillyard and Munte, 1984; Luck and Hillyard, 1994; Mangun and Hillyard, 1995). In language comprehension, similar P2 amplitude increases are found when words or pictures are highly expected in a given sentence, or follow related prime words (e.g., Federmeier and Kutas, 2002, 2005; Federmeier et al., 2005, 2007; Federmeier, 2007). As in vision, these effects are thought to index top-down driven anticipatory activation of the perceptual features related to an expected upcoming word. Given that the current P2 modulation, and in contrast to the pP2 encountered by Strijkers et al. (2011), has a similar frontal distribution as the P2 modulations reported in the above studies, it is possible that we are dealing with a similar effect as encountered in visual perception and language comprehension. In this case, our findings would suggest that the intention to speak can already affect processes related to the input. Such findings would be highly intriguing since they would indicate that the mere intention to engage in speech behavior not only influences the manner in which we retrieve lexical representations (Strijkers et al., 2011), the crucial units which must be retrieved for engaging in a speech act, but already alters the way our brain reacts to the input. Put differently, the intention to speak would change the manner in which we "perceive" an object and a word very early on compared to the visuo-conceptual processing of the same stimuli when the perceptual and/or semantic goals are different from those required for a speech act.

Interestingly, the fact that another early component, the N170, was affected by task intention, fits nicely with this conclusion. Typically, this occipito-temporal negative going component is enhanced for objects with which we have great expertise, such as faces (e.g., Bentin et al., 1996; Tanaka and Curran, 2001; Gauthier et al., 2003; Rossion et al., 2003; but see Liu et al., 2002; Thierry et al., 2007). It is worth mentioning, especially in the current context, that a few studies also reported similar N170-like modulations for expertise with written words (e.g., McCandliss et al., 2003; Yum et al., 2011). If the N170 modulation reported here can be considered similar to the one thought to be sensitive to visual expertise, this would mean that naming leads to more specific visual identification processes compared to categorization, which is not that surprising given the differences in task demands. To sum up, the fact that the input-related N170 was affected by differences in task intention rather than differences between stimuli provides support for our tentative proposal that the intention to speak may already affect the manner in which we visually and/or semantically process stimuli. This is a highly interesting possibility which clearly merits further investigation in future work.

CONCLUSION

In the present study, by comparing brain responses elicited by the same visual input in two different tasks, we were able to establish that goal-directed top-down influences penetrate and affect early processing of words and objects. These results concur with the evidence reported in a recent study (Strijkers et al., 2011) emphasizing the crucial role of early intention-driven processes for the production of speech. In addition, the results of the current study showed that these top-down signals are, at least partly, generated by a domain-general system, likely to be functional each time we want to speak. Whether these influences operate at the level of word representations, as demonstrated in previous work, or even during visuo-conceptual processing could not be determined explicitly on the basis of the present results. Nevertheless, the current results contribute some important insights regarding a topic which has received little attention in the literature so far, and paves the way for future research that will help to gradually narrow down the functional nature and sources of goal-directed processing in language production.

REFERENCES

Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *J. Cogn. Neurosci.* 15, 600–609.

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmidt, A. M., Dale, A.

M., Hämäläinen, M. S., Marinkovic, K., Schacter, D. L., Rosen, B. R., and Halgren, E. (2006). Top-down facilitation of visual recognition. *Proc. Natl. Acad. Sci. U.S.A.* 103, 449–454.

Bentin, S., McCarthy, G., Perez, E., Puce, A., and Allison, T. (1996).

Electrophysiological studies of face perception in humans. *J. Cogn. Neurosci.* 8, 551–565.

Bles, M., and Jansma, B. M. (2008). Phonological processing of ignored distractor pictures, an fMRI investigation. *BMC Neurosci.* 9, 20. doi:10.1186/1471-2202-9-20

Bloem, I., and La Heij, W. (2003). Semantic facilitation and semantic interference in word translation: implications for models of lexical access in language production. *J. Mem. Lang.* 48, 468–488.

Caramazza, A. (1997). How many levels of processing are there in

lexical access? *Cogn. Neuropsychol.* 14, 177–208.

Chauncey, K., Holcomb, P. J., and Grainger, J. (2009). Primed picture naming within and across languages: an ERP investigation. *Cogn. Affect. Behav. Neurosci.* 9, 286–303.

Christoffels, I. K., Firk, C., and Schiller, N. O. (2007). Bilingual language control: an event-related brain potential study. *Brain Res.* 1147, 192–208.

Corbetta, M., Miezin, F. M., Dobmeyer, S., Shulman, G. L., and Petersen, S. E. (1990). Attentional modulation of neural processing of shape, color, and velocity in humans. *Science* 248, 1556–1559.

Costa, A., and Caramazza, A. (2002). The production of noun phrases in English and Spanish: implications for the scope of phonological encoding in speech production. *J. Mem. Lang.* 46, 178–198.

Costa, A., Strijkers, K., Martin, C., and Thierry, G. (2009). The time course of word retrieval revealed by event-related brain potentials during overt speech. *Proc. Natl. Acad. Sci. U.S.A.* 106, 21442–21446.

Cutting, J. C., and Ferreira, V. S. (1999). Semantic and phonological information flow in the production lexicon. *J. Exp. Psychol. Learn. Mem. Cogn.* 25, 318–344.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychol. Rev.* 93, 283–321.

Dell'Acqua, R., and Grainger, J. (1999). Unconscious semantic priming from pictures. *Cognition* 73, B1–B15.

Dell'Acqua, R., Sessa, P., Peressotti, F., Mulatti, C., Navarrete, E., and Grainger, J. (2010). ERP evidence for ultra-fast semantic processing in the picture-word interference paradigm. *Front. Psychol.* 1:177. doi:10.3389/fpsyg.2010.00177

Desimone, R., and Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* 18, 193–222.

Engel, A. K., Fries, P., and Singer, W. (2001). Dynamic predictions: oscillations and synchrony in top-down processing. *Nat. Rev. Neurosci.* 2, 704–716.

Federmeier, K. D. (2007). Thinking ahead: the role and roots of prediction in language comprehension. *Psychophysiology* 44, 491–505.

Federmeier, K. D., and Kutas, M. (2002). Picture the difference: electrophysiological investigations of picture processing in the twocerebral hemispheres. *Neuropsychologia* 40, 730–747.

Federmeier, K. D., and Kutas, M. (2005). Aging in context: age-related changes in context use during language comprehension. *Psychophysiology* 42, 133–141.

Federmeier, K. D., Mai, H., and Kutas, M. (2005). Both sides get the point: hemispheric sensitivities to sentential constraint. *Mem. Cognit.* 33, 871–886.

Federmeier, K. D., Wlotko, E., De Ochoa-Dewald, E., and Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Res.* 1146, 75–84.

Gauthier, I., Curran, T., Curby, K. M., and Collins, D. (2003). Perceptual interference supports a non-modular account of face processing. *Nat. Neurosci.* 6, 428–432.

Gilbert, C. D., and Sigman, M. (2007). Brain states: top-down influences in sensory processing. *Neuron* 54, 677–696.

Glaser, W. R., and Glaser, M. O. (1989). Context effects on Stroop-like word and picture processing. *J. Exp. Psychol. Gen.* 118, 13–42.

Hillyard, S. A., and Munte, T. F. (1984). Selective attention to color and location: an analysis with event-related brain potentials. *Percept. Psychophys.* 36, 185–198.

Hillyard, S. A., Vogel, E. K., and Luck, S. J. (1998). Sensory gain control (amplification) as a mechanism of selective attention: electrophysiological and neuroimaging evidence. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 353, 1257–1267.

Jescheniak, J. D., and Levelt, W. J. M. (1994). Word frequency effects in speech production: retrieval of syntactic information and of phonological form. *J. Exp. Psychol. Learn. Mem. Cogn.* 20, 824–843.

Joyce, C., and Rossion, B. (2005). The face-sensitive N170 and VPP components manifest the same brain processes: the effect of reference electrode site. *Clin. Neurophysiol.* 116, 2613–2631.

Kastner, S., Pinsk, M., De Weerd, P., Desimone, R., and Ungerleider, L. (1999). Increased activity in human visual cortex during directed attention in the absence of visual stimulation. *Neuron* 22, 751–761.

Kastner, S., and Ungerleider, L. G. (2000). Mechanisms of visual attention in the human cortex. *Annu. Rev. Neurosci.* 23, 315–341.

Koester, D., and Schiller, N. (2008). Morphological priming in overt language production: electrophysiological evidence from Dutch. *Neuroimage* 42, 1622–1630.

Kroll, J. F., and Potter, M. C. (1984). Recognizing words, pictures, and concepts: a comparison of lexical, object, and reality decisions. *J. Verbal Learn Verbal Behav.* 23, 39–66.

Levelt, W. J. M., Roelofs, A., and Meyer, A. S. (1999). A theory of lexical access in speech production. *Behav. Brain Sci.* 22, 1–75.

Levelt, W. J. M., Schriefers, H., Vorberg, D., Meyer, A. S., Pechmann, T., and Havinga, J. (1991). The time course of lexical access in speech production: a study of picture naming. *Psychol. Rev.* 98, 122–142.

Liu, J., Harris, A., and Kanwisher, N. (2002). Stages of processing in face perception: an MEG study. *Nat. Neurosci.* 5, 910–916.

Luck, S. J., Chelazzi, L., Hillyard, S. A., and Desimone, R. (1997). Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *J. Neurophysiol.* 77, 24–42.

Luck, S. J., and Hillyard, S. A. (1994). Electrophysiological correlates of feature analysis during visual search. *Psychophysiology* 31, 291–308.

Mangun, G. R., and Hillyard, S. A. (1995). "Mechanisms and models of selective attention," in *Electrophysiology of Mind*, eds M. D. Rugg and M. G. H. Coles (Oxford: Oxford University Press), 40–85.

McCandliss, B. D., Cohen, L., and Dehaene, S. (2003). The visual word form area: expertise for reading in the fusiform gyrus. *Trends Cogn. Sci.* 7, 293–299.

Meyer, A. S., Belke, E., Telling, A. L., and Humphreys, G. W. (2007). Early activation of object names in visual search. *Psychon. Bull. Rev.* 14, 710–716.

Meyer, A. S., Sleiderink, A. M., and Levelt, W. J. M. (1998). Viewing and naming objects: eye movements during noun phrase production. *Cognition* 66, 25–33.

Morsella, E., and Miozzo, M. (2002). Evidence for a cascade model of lexical access in speech production. *J. Exp. Psychol. Learn. Mem. Cogn.* 28, 555–563.

Navarrete, E., and Costa, A. (2005). Phonological activation of ignored pictures: further evidence for a cascade model of lexical access. *J. Mem. Lang.* 53, 359–377.

Posner, M. I., and Dehaene, S. (1994). Attentional networks. *Trends Neurosci.* 17, 75–79.

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition* 42, 107–142.

Roelofs, A. (2003). Goal-referenced selection of verbal action: modelling attentional control in the Stroop task. *Psychol. Rev.* 110, 88–125.

Roelofs, A. (2008). Attention to spoken word planning: chronometric and neuroimaging evidence. *Lang. Linguist. Compass* 2, 389–405.

Rossion, B., Joyce, C. A., Cottrell, G. W., and Tarr, M. J. (2003). Early lateralization and orientation tuning for face, word, and object processing in the visual cortex. *Neuroimage* 20, 1609–1624.

Schriefers, H., Meyer, A. S., and Levelt, W. J. M. (1990). Exploring the time course of lexical access in language production: picture–word interference studies. *J. Mem. Lang.* 29, 86–102.

Snodgrass, J. G., and Vanderwart, M. (1980). A standardized set of 260 pictures: norm for name agreement, image agreement, familiarity, and visual complexity. *J. Exp. Psychol. Hum. Learn. Mem.* 6, 174–215.

Starreveld, P. A., and La Heij, W. (1996). Time-course analysis of semantic and orthographic context effects in picture naming. *J. Exp. Psychol. Learn. Mem. Cogn.* 22, 896–918.

Strijkers, K., Costa, A., and Thierry, G. (2010). Tracking lexical access in speech production: electrophysiological correlates of word frequency and cognate effects. *Cereb. Cortex* 20, 912–928.

Strijkers, K., Holcomb, P., and Costa, A. (2011). Conscious intention to speak proactively facilitates lexical access in object naming. *J. Mem. Lang.* 65, 345–362.

Tanaka, J. W., and Curran, T. (2001). A neural basis for expert object recognition. *Psychol. Sci.* 12, 43–47.

Thierry, G., Martin, C. D., Downing, P., and Pegna, A. J. (2007). Controlling for interstimulus perceptual variance abolishes N170 face selectivity. *Nat. Neurosci.* 10, 505–511.

Ullman, S. (1995). Sequence seeking and counter streams: a computational model for bidirectional information flow in the visual cortex. *Cereb. Cortex* 1, 1–11.

Yum, Y. N., Holcomb, P. J., and Grainger, J. (2011). Words and pictures: an electrophysiological investigation of domain specific processing in native Chinese and English speakers. *Neuropsychologia* 49, 1910–19220.