Event-related potentials and oscillatory brain responses associated with semantic and Stroop-like interference effects in overt naming

Vitória Piai[a,b], Ardi Roelofs[a], and Roemer van der Meij[a]

[a] Radboud University Nijmegen, Donders Institute for Brain, Cognition and Behaviour, Centre for Cognition, Montessorilaan 3, 6525 HR, Nijmegen, The Netherlands

[b] International Max Planck Research School for Language Sciences, Wundtlaan 1, 6525 XD, Nijmegen, The Netherlands

V. Piai: v.piai@donders.ru.nl

A. Roelofs: a.roelofs@donders.ru.nl

R. van der Meij: r.vandermeij@donders.ru.nl

Correspondence concerning this article should be addressed to:
Vitória Piai
Radboud University Nijmegen
Donders Institute for Brain, Cognition and Behaviour
Centre for Cognition
Spinoza Building B.01.09
Montessorilaan 3
6525 HR Nijmegen
The Netherlands
Email: V.Piai@donders.ru.nl
Phone: +31.24.3612635
Fax: +31.24.3616066

Abstract

1   Picture-word interference is a widely employed paradigm to investigate lexical access
2   in word production: Speakers name pictures while trying to ignore superimposed distractor
3   words. The distractor can be congruent to the picture (pictured cat, word *cat*), categorically
4   related (pictured cat, word *dog*), or unrelated (pictured cat, word *pen*). Categorically related
5   distractors slow down picture naming relative to unrelated distractors, the so-called semantic
6   interference. Categorically related distractors slow down picture naming relative to congruent
7   distractors, analogous to findings in the colour-word Stroop task. The locus of semantic
8   interference and Stroop-like effects in naming performance has recently become a topic of
9   debate. Whereas some researchers argue for a pre-lexical locus of semantic interference and a
10 lexical locus of Stroop-like effects, others localise both effects at the lexical selection stage.
11 We investigated the time course of semantic and Stroop-like interference effects in overt
12 picture naming by means of event-related potentials (ERP) and time-frequency analyses.
13 Moreover, we employed cluster-based permutation for statistical analyses. Naming latencies
14 showed semantic and Stroop-like interference effects. The ERP waveforms for congruent
15 stimuli started diverging statistically from categorically related stimuli around 250 ms.
16 Deflections for the categorically related condition were more negative-going than for the
17 congruent condition (the Stroop-like effect). The time-frequency analysis revealed power
18 increase in the beta band (12-30 Hz) for categorically related relative to unrelated stimuli
19 roughly between 250 and 370 ms (the semantic effect). The common time window of these
20 effects suggests that both semantic interference and Stroop-like effects emerged during lexical
21 selection.
22   *Key words:* semantic interference, Stroop, ERP, time-frequency analysis, beta band
23
24

1. Introduction

The colour-word Stroop task (Stroop, 1935) has been widely used in the cognitive neurosciences to investigate various aspects of human cognition (e.g., Appelbaum et al., 2009; Balota et al., 2010; Bench et al., 1993; Bub et al., 2006; Lachter et al., 2008; Roelofs et al., 2006; Szucs & Soltész, 2010). In this task, people are required to name the ink colour of written words denoting colours (see MacLeod, 1991 for a review). The written word can be congruent with the name of the ink colour (e.g., *blue* printed in blue ink) or incongruent (e.g., *blue* printed in red ink). The central finding of the Stroop task is that response times (RTs) for the incongruent condition are longer than for the congruent condition, referred to as the Stroop effect. In the past few decades, researchers have also made use of a picture-word analogue of the colour-word Stroop task (e.g., Glaser & Glaser, 1989; La Heij, 1988). In the picture-word interference (PWI) task, speakers have to name pictured objects while trying to ignore written distractor words superimposed onto the pictures. The distractor can be the name of the picture (congruent condition: pictured cat, word *cat*), a categorically related word (related condition: pictured cat, word *dog*), or a categorically unrelated word (unrelated condition: pictured cat, word *pen*). A central finding obtained with PWI is that categorically related distractors slow down picture naming relative to unrelated distractors. Furthermore, categorically related distractors slow down picture naming relative to congruent distractors, a finding which is analogous to that in the colour-word Stroop task. In the remainder of this article, we refer to the PWI contrast between categorically related and congruent conditions as the Stroop-like effect, and the contrast between categorically related and unrelated conditions as semantic interference, following the convention in the literature (e.g., Roelofs, 2003).

Models of Stroop task performance assume processing stages of perceptual and conceptual encoding, response selection, response programming, and response execution (e.g., Dell'Acqua et al., 2007; Roelofs, 2003; Sanders, 1990). Models of picture naming assume perceptual and conceptual encoding, lexical selection, word-form encoding, and articulation as the processing stages (e.g., Levelt et al., 1999; Roelofs, 2003). Response selection in models of Stroop task performance corresponds to lexical selection in models of picture naming, response programming corresponds to word-form encoding, and response execution to articulation (Roelofs, 2003).

Different studies have made use of behavioural measures, neuroimaging, and computational modelling to examine the nature, time course, and neural underpinnings of performance in Stroop-like tasks (e.g., Liotti et al., 2000; Pardo et al., 1990; Roelofs et al., 2006; Van Maanen et al., 2009). Although these examinations have yielded many converging findings, researchers have not always found agreement on the interpretation of the results (e.g., Abdel Rahman & Aristei, 2010; Dell'Acqua et al. 2007; Mahon et al., 2007; Van Maanen et al., 2009). In the present study, we address one particular issue: the temporal locus of semantic interference and Stroop-like effects in naming tasks. The debate about the locus of the Stroop effect in colour-word Stroop task performance has a long history (see MacLeod, 1991, for a review). Whereas some accounts localise the effect in perceptual/conceptual encoding (e.g., Hock & Egeth, 1970), other accounts assume a locus close to articulation onset (e.g., Morton, 1969). More recently, computationally implemented accounts of the Stroop phenomenon (e.g., Cohen et al., 1990; Roelofs, 1992, 2003) have localised the Stroop effect in the stage of response selection. According to Roelofs (1992, 2003) and Starreveld and La Heij (1996), among others, the semantic interference effect in PWI also arises in response selection.

1   The idea that semantic interference and Stroop-like effects emerge during the same
2   processing stage (i.e., lexical selection) was recently called into question by a PWI study
3   conducted by Dell'Acqua and colleagues (2007), employing the psychological refractory
4   period (PRP) procedure. In their experiment, participants performed a manual tone
5   discrimination task (Task 1) and a PWI task (Task 2). On each trial, a tone stimulus and a
6   picture-word stimulus were presented, each requiring a quick and accurate response. The
7   stimulus onset asynchrony (SOA) between the two task stimuli ranged from 100 ms (short
8   SOA) to 1000 ms (long SOA). Participants were instructed to respond to the tone of Task 1
9   before responding to the picture-word stimulus of Task 2. RTs were measured to determine
10   the extent to which Task 1 delayed performance of Task 2.
11
12   Dell'Acqua et al. (2007) observed that the semantic interference effect was smaller at
13   the short than at the long SOAs. This result was different from earlier findings of Fagot and
14   Pashler (1992) using the colour-word Stroop task with a PRP design, where the effects of
15   Stroop condition (congruent vs. incongruent) and SOA were additive. That is, the magnitude
16   of the Stroop effect was the same at the short and long SOAs. Assuming that the locus of
17   dual-task interference is in response selection (Fagot & Pashler, 1992), this finding confirms
18   other evidence that the Stroop effect arises in selecting a colour-naming response (e.g.,
19   Roelofs, 2003). Dell'Acqua et al.'s finding of underadditive semantic and SOA effects was
20   interpreted by the authors as evidence that semantic interference for picture-word stimuli
21   arises earlier than response word selection and, thus, it is not a Stroop-like effect. According
22   to Dell'Acqua et al., the semantic interference effect arises during perceptual/conceptual
23   encoding.
24
25   Although Ayora and colleagues (Ayora et al., 2011) recently replicated the
26   underadditivity of semantic and SOA effects (in Italian), other researchers obtained additive
27   effects. Schnur and Martin (2012) conducted two experiments with different materials, and
28   slightly different experimental parameters, and failed to replicate Dell'Acqua et al. (2007).
29   One of the experiments used the picture names and word distractors from Dell'Acqua et al.
30   translated into English. In all experiments, equivalent PWI effects were obtained at short and
31   long SOAs following tone identification. Moreover, in our own lab (Piai, Roelofs, &
32   Schriefers, in preparation), we failed to replicate Dell'Acqua et al. in seven experiments with
33   different materials and parameters, including one experiment that used the picture names and
34   word distractors from Dell'Acqua et al. translated into Dutch. Just as Schnur and Martin, we
35   obtained equivalent PWI effects at short and long SOAs in all experiments. Following the
36   logic of Dell'Acqua et al., the additive semantic and SOA effects suggest that semantic
37   interference arises in lexical selection, unlike what Dell'Acqua et al. assume. Clearly, before
38   the empirical discrepancy between these studies is resolved (see Piai et al., 2011, for a
39   possible resolution in terms of executive control parameters), it seems premature to assume
40   that the issue of the locus of semantic interference in picture naming has been settled.
41
42   Moreover, the conclusion of Dell'Acqua et al. (2007) that the PWI effect is not a
43   Stroop effect was contested by Van Maanen and colleagues (Van Maanen et al., 2009), who
44   argued that the same interference mechanism underlies both effects, although the effects arise
45   at different processing stages. According to the authors, the discrepancy between the findings
46   of Dell'Acqua et al. and Fagot and Pashler (1992) should be attributed to differences in
47   processing speed between pictures (line drawings, which are relatively hard to process) and
48   colours (which are easily identifiable). Because pictures take longer to process than colours,
49   according to Van Maanen and colleagues, Stroop-like and semantic interference effects from
50   distractor words in picture naming occur during perceptual and conceptual encoding (i.e.,

before response selection), whereas the Stroop effect from distractor words in colour naming occurs during response selection. Van Maanen et al. (2009) presented the results of computer simulations corroborating their claim of a common mechanism but different loci for the effects in PWI and colour-word Stroop studies.

A problem with the simulations of Van Maanen and colleagues (2009) is that they are based on the assumption of faster color than picture processing, which is questionable. High temporal-resolution examinations suggest estimates for the time course of colour processing that are not different from estimates for picture shape processing, namely 100-200 ms (e.g., Anllo-Vento et al., 1998; Müller & Keil, 2004; see Dell'Acqua et al., 2010, for a review of some of these studies). This evidence challenges the critical parameters in the simulations of Van Maanen and colleagues.

Another prominent account of the semantic interference effect is the response exclusion hypothesis (e.g., Miozzo & Caramazza, 2003). According to this hypothesis, the semantic interference effect arises after the lexical selection stage, closer to articulation onset. The effect is argued to emerge due to the exclusion of the distractor word from an articulatory buffer (cf. Morton, 1969). The temporal locus of the semantic interference effect stipulated by this account, close to articulation onset, is not easy to investigate with EEG because of artefacts emerging from speech production. Therefore, we do not address this hypothesis in the present study. However, there is accumulating evidence against the response-exclusion account of semantic interference, reported elsewhere (e.g., Abdel Rahman & Aristei, 2010; Mulatti & Coltheart, in press; Piai et al., 2011; Roelofs et al., in press; Starreveld et al., in press).

To summarise, whereas some models assume a common lexical locus of semantic interference and Stroop-like effects in PWI (Starreveld & La Heij, 1996; Roelofs, 1992, 2003), other accounts assume a perceptual/conceptual encoding locus for semantic interference and a lexical selection locus for the Stroop-like effect (Dell'Acqua et al., 2007), or a perceptual/conceptual encoding locus for both semantic interference and Stroop-like effects in the PWI task (Van Maanen et al., 2009).

1.1. The present study

The aim of the present study is to adjudicate between the different views on the temporal loci of semantic interference and Stroop-like effects using EEG measures such as event-related potentials (ERPs) and time-frequency representations (TFRs) of power. EEG is an ideal tool to address questions about the timing of processes as it allows for a fine-grained temporal resolution.

Estimates of the timing of processing stages underlying word production were provided by an influential meta-analysis (Indefrey & Levelt, 2004, see also Indefrey, 2011). According to these estimations, based on an average naming latency of 600 ms, the stage of perceptual and conceptual encoding is completed around 200 ms after picture onset, after which lexical selection starts. In the PWI task, mean naming latencies tend to be longer than 600 ms, namely within a range of 700 to 800 ms (e.g., La Heij, 1988; Roelofs, 1992). Taking 750 ms as the mean naming latency (corresponding to what was obtained in the present study), and using a proportional scaling of the estimates to this mean (see Indefrey, 2011), yields 250 ms as the end of the time window of perceptual and conceptual encoding and as the point in time at which the operation of word selection is initiated. This means that, according

to the proposal that semantic interference in PWI emerges during perceptual/conceptual encoding (e.g., Dell'Acqua et al., 2007), differences in brain responses between categorically related stimuli (e.g., pictured cat, word *dog*) and unrelated stimuli (e.g., pictured cat, word *pen*) should be seen in a time window that extends at most to 250 ms post picture onset. Moreover, if Stroop-like effects arise in lexical selection, differences between categorically related stimuli (e.g., pictured cat, word *dog*) and congruent stimuli (e.g., pictured cat, word *cat*) should be detected in a time window starting around 250 ms after picture onset. If, however, semantic interference and Stroop-like effects in PWI both arise in perceptual/conceptual encoding (Van Maanen et al., 2009), both effects should emerge before 250 ms post picture onset. Finally, if semantic interference and Stroop-like effects both arise during lexical selection (Starreveld & La Heij, 1996; Roelofs, 1992, 2003), these effects should be visible in a time window starting after 250 ms post picture onset.

Most previous investigations of performance in the colour-word Stroop task using ERPs found a negativity, associated with the incongruent condition relative to the congruent condition, occurring between 300-550 ms after stimulus onset with a centro-parietal scalp distribution (e.g., Liotti et al., 2000), suggesting a lexical selection locus of the Stroop effect. Investigations of performance on the PWI task using ERPs did not include the Stroop contrast of congruent versus incongruent stimuli (e.g., Aristei et al., 2011; Hirschfeld et al., 2008), except for a study by Xiao et al. (2010). These authors observed a negative-going potential between 280 and 400 ms for the categorically related condition relative to the congruent condition, with a fronto-central scalp distribution. There are, however, a few reasons why this study does not allow us to draw a conclusion about the temporal locus of semantic interference and Stroop-like effects in PWI. First of all, only three stimuli (i.e., three geometrical shapes) were used as pictures. In contrast, PWI studies typically include some 20-50 different pictures of various semantic domains, such as animals, tools, etc. (e.g., Aristei et al., 2011; Dell'Acqua et al., 2007; Roelofs, 2003). Moreover, participants responded to the geometrical shapes by pressing keys, so no overt naming was used. Furthermore, for the ERP analysis, there were no a priori defined time windows. The selection of time windows for statistical analyses was based on visual inspection of the data, a procedure prone to bias. Regarding semantic interference in PWI, recent attempts to find this specific effect with ERPs were not successful (Aristei et al., 2011; Hirschfeld et al., 2008), except for one study (Dell'Acqua et al., 2010), which obtained two semantic interference effects, one with an onset latency of 106 ms and the other starting at 320 ms post picture onset. However, this study did not include a Stroop-like contrast, precluding a direct comparison of semantic interference and Stroop-like effects.

In the present study, we addressed the issue regarding the timing of the Stroop-like effect and the semantic interference effect in the PWI task by means of electrophysiological measures while participants overtly articulated their responses. The measurement of EEG in overt speech production tasks had long been avoided because of the presumed movement artefacts caused by articulation. However, the use of overt articulation in EEG research is no longer considered problematic (see Eulitz et al., 2000, for a demonstration that ERPs can be analysed using overt naming up to stages of phonetic processing), and an increasing number of studies has made use of this combination successfully (e.g., Aristei et al., 2011; Costa et al., 2009; Hirschfeld et al., 2008; Laganaro & Perret, 2011; Strijkers et al., 2010; Verhoef et al., 2009, 2010). By comparing the brain responses among distractor conditions (i.e., categorically related, unrelated, congruent), it may be assessed when the underlying processes diverge from each other, indicating a time window associated with semantic and Stroop-like effects.

Besides the type of distractor used, we also manipulated the lexical frequency of the picture name. With this kind of manipulation, a word-frequency effect is usually observed: Pictures with high-frequency names are named faster than pictures with low-frequency names (e.g., Oldfield & Wingfield, 1965). Since this effect has been shown to be a lexical effect (Jescheniak & Levelt, 1994; Jescheniak et al., 2003), we used this manipulation as a possible extra marker of lexical processes in our experiment. Regarding the EEG, this lexical-frequency effect should also be observed in time windows related to lexical processes, starting no earlier than 250 ms (cf. Strijkers et al., 2010).

Since we do not have a specific hypothesis for the ERPs regarding the scalp distribution of our effects, an appropriate statistical method must be chosen that allows for testing numerous hypotheses (due to many comparisons of time point by channel), while dealing with the multiple-comparisons problem (cf. Lage-Castellanos et al., 2010). A method that combines the richness of the data with a strict control of the family-wise error rate is cluster-based permutation tests (Maris & Oostenveld, 2007). With this method, no pre-defined time windows are necessary nor is it required to average the signal within a time-window. This means that we have a method to determine a time window where brain responses differ between conditions in a unbiased way. Note that this method does not allow us to determine the exact starting point of the earliest divergences between conditions. However, it enables us to determine the time window of the main component of brain responses (cf. Letham & Raij, 2010).

A further aim of the present study was to investigate how the semantic interference and the Stroop-like effects in the PWI task might be reflected in changes in oscillatory brain activity. Whereas ERPs capture mainly evoked activity, i.e., brain responses that are phase-locked to a stimulus, time-frequency analyses reveal changes in oscillatory induced activity, i.e., not necessarily phase-locked, in specific frequency bands over time. Oscillatory activity is modulated by a variety of cognitive factors and is thought to reflect the activity of large ensembles of synchronised neurons (e.g., Buzsáki & Draguhn, 2004). These two approaches to analysing electrophysiological data (i.e., ERP and oscillatory power) are complementary and in some cases, differences in brain signals not evident in terms of ERPs can be revealed in the time-frequency domain (cf. Mazaheri & Jensen, 2010).

1.2. Summary

The present study investigates the temporal locus of the semantic interference and Stroop-like effects in PWI using EEG with overt articulation (the majority of previous investigations of Stroop task performance used manual responding). Moreover, we conducted analyses of oscillatory power in the context of word production and the PWI task. Finally, we analysed the electrophysiological measures with a method that does not require specific time-windows and channels to be determined a priori, which means we avoid basing our analyses on biased or arbitrary time windows and channels. This method allows us to find a time window at which divergences in the EEG for different conditions become statistically significant, both in the ERPs and in the TFRs of power. If semantic interference occurs in perceptual/conceptual encoding and the Stroop-like effect in response selection, the semantic effect should emerge before 250 ms post picture onset and the Stroop-like effect after this moment in time. However, if the locus of both the semantic interference and Stroop-like effect in PWI is the perceptual/conceptual encoding stage, both effects should emerge in a time window ending before 250 ms, when perceptual/conceptual encoding is completed. Finally, if

the locus of semantic interference and Stroop-like effects is the response-selection stage, both effects should emerge after 250 ms post picture onset, when lexical selection is initiated.

2. Results

2.1. Behavioural data

The error percentages for the different distractor types were 1.2% for the congruent condition, 3.4% for the categorically related condition, and 2.6% for the unrelated condition; and 2.5% for high-frequency and 2.3% for low-frequency picture names. The logistic regression model indicated that, relative to the congruent condition, categorically related distractors caused the log-odds of an incorrect response to increase by a factor of 2.9, β $coefficient = -1.05$, $S.E. = .31$, $Wald\ Z = -3.3$, $p < .001$; and unrelated distractors increased the log-odds of an incorrect response by a factor of 2.7, $\beta\ coefficient = -1$, $S.E. = .32$, $Wald\ Z = -3.1$, $p = .002$. Categorically related distractors did not differ from unrelated distractors, $p = .800$. Lexical frequency was not a significant predictor in the regression model, $p = 1.0$.

Figure 1(A) shows the box-and-whisker diagram for the RTs, with the mean and the distribution for each condition (the ends of the whiskers represent the lowest and highest data point within 1.5 interquartile range). A main effect of distractor type was found, $F_1(2,38) = 150.86$, $p < .001$, $F_2(2,76) = 187.91$, $p < .001$. Contrasts revealed a semantic interference effect (26 ms), $F_1(1,19) = 34.04$, $p < .001$, $F_2(1,39) = 10.80$, $p = .002$, $95\%\ CI$ (17, 36), and a Stroop-like effect (125 ms), $F_1(1,19) = 200.66$, $p < .001$, $F_2(1,39) = 364.57$, $p < .001$, $95\%\ CI$ (107, 145). The mean RTs for the high-frequency and low-frequency conditions were 749 ms and 757 ms respectively. The lexical frequency effect was only found in the analysis by participants, $F_1(1,19) = 5.65$, $p = .028$, $F_2(1,38) < 1$, $95\%\ CI$ (-16, 0.2). No interaction was found between distractor type and lexical frequency, Fs < 1.

Since the lexical-frequency effect was not robust in the mean RTs, we also performed Vincentile analyses to examine the shape of the RT distribution. We divided the rank-ordered RTs for each participant into 20% quantiles and computed quantile means for each lexical-frequency condition. The quantiles were then averaged across participants. This technique allows the detection of opposite underlying effects possibly giving rise to null effects in the mean RTs (e.g., Heathcote et al., 1991).

Figure 1(B) shows the Vincentised cumulative distribution curves for the lexical-frequency effect. The high-frequency condition clearly yields shorter RTs than the low-frequency condition, except for the 20% slowest responses. Thus the RT distribution analysis shows that the lexical-frequency effect is only absent towards the tail of the distribution.

2.2. ERP data

Grand-average ERPs for the three distractor types, collapsed over high- and low-frequency conditions, are shown in Figure 2(A) for nine representative channels (coloured red in Figure 2(B); there is a one-to-one correspondence between the orientation of the channels in 2(A) and 2(B)). For the Stroop-like effect, starting around 250 ms, amplitudes for the congruent condition become larger than for the categorically related condition, and this difference tends to increase with time. For the semantic interference effect, there are no visible differences in the waveforms between the categorically related and unrelated conditions.

For the Stroop-like effect, a negative statistically significant cluster was detected, starting at 254 ms and lasting until the end of the segment (i.e., 500 ms), $p < .001$. The cluster was first detected in fronto-central electrodes (as shown in the left scalp topography in Figure 2(C)), extending later to centro-parietal electrodes (as shown in the right scalp topography in Figure 2(C)). For the semantic interference effect, no significant clusters were detected.

Grand-average ERPs for the two lexical-frequency conditions, collapsed over the categorically related and unrelated conditions, are shown in Figure 3(A) for nine representative channels (coloured red in Figure 3(B); there is a one-to-one correspondence between the orientation of the channels in 3(A) and 3(B)). Starting around 260 ms, amplitudes for the high-frequency condition become more positive-going relative to the low-frequency condition. A positive statistically significant cluster was detected between 288 ms and 390 ms, $p = .042$. The cluster was detected in fronto-central electrodes (as shown in the scalp topography in Figure 3(C)).

2.3. Time-frequency data

Power spectra revealed increased relative power in the beta band for the categorically related condition compared to the unrelated condition approximately between 210 and 380 ms (Figure 4(A)). Statistical testing identified a significant cluster with a frequency range of approximately 15-27 Hz and a time range of approximately 230-370 ms, present in 15 channels (coloured red in Figure 4(B)), $p = .019$. This effect is most prominent in central channels, slightly right-lateralised, as shown in the scalp topography in Figure 4(C).
No significant clusters were detected either for the Stroop-like effect or for the lexical-frequency effect. Finally, no significant clusters were detected in the high-frequency range (i.e., 30-100 Hz).

3. Discussion

The present EEG experiment examined the timing of semantic interference and Stroop-like effects in the PWI task. We investigated whether the timing of the ERP effects suggests a lexical selection locus of both effects (Roelofs, 2003), a perceptual/conceptual locus for the semantic interference effect and a lexical selection locus for the Stroop-like effect (Dell'Acqua et al., 2007), or a perceptual/conceptual encoding locus for both semantic interference and Stroop-like effects in PWI (Van Maanen et al., 2009).

Our RT data revealed a Stroop-like effect (i.e., pictures paired with categorically related distractors were named more slowly and less accurately relative to congruent pairs) and a semantic interference effect (i.e., categorically related pairs were named more slowly than unrelated pairs). The effect of lexical frequency in the RTs was less robust, and absent in the relatively slow responses. Regarding the ERPs, a statistically significant negativity was detected for categorically related stimuli relative to congruent stimuli (the Stroop-like effect), in a time window starting around 250 ms and lasting until the last analysed segment at 500 ms. The effect started with a fronto-central scalp distribution, spreading later slightly to centro-parietal channels. The scalp distribution of the effect, however, was very similar between 250-350 ms and 350-500 ms. A lexical-frequency effect was detected roughly between 290 ms and 390 ms mainly in fronto-central channels, with the high-frequency condition eliciting a more positive-going wave relative to the low-frequency condition. No statistically significant clusters were detected corresponding to the semantic interference

effect in the ERPs, in line with other studies (e.g., Aristei et al., 2011; Hirschfeld et al., 2008, but see Dell'Acqua et al., 2010). However, in the time-frequency domain, a statistically significant relative power increase in the beta-band was observed for categorically related stimuli compared to unrelated stimuli (the semantic interference effect) between 230 and 370 ms. Due to the intrinsic temporal smearing in the time-frequency estimations, the latency of this beta effect cannot be taken strictly as revealing the precise timing of the underlying neural processes, so the time window of this effect should be seen as an approximation of the time window of the semantic interference effect.

The lexical-frequency effect in the current experiment was small and not robust in the mean RTs. Although we are not certain why, this weakness could be attributable to the use of the PWI paradigm (previous reports of the lexical-frequency effect did not use the PWI paradigm, e.g., Jescheniak & Levelt, 1994; Jescheniak et al., 2003; Oldfield, Wingfield, 1965). In the ERPs, although small, this effect was significant between roughly 290 ms and 390 ms. Our time course diverges slightly from the interval of 150 – 200 ms for the frequency effect obtained by Strijkers et al. (2010), but they used a standard picture naming task, with no distractor words. Standard picture naming yields shorter RTs than picture naming in PWI, as it was the case in our experiment (PWI, mean naming RT around 750 ms) compared to Strijkers et al. (standard naming, mean naming RT around 700 ms). Thus the time shift of our effect compared to the effect of Strijkers et al. may be due to the presence of the distractor word.

Our ERP results for the Stroop-like effect are similar to Xiao et al.'s (2010) study, in which a fronto-central negativity was observed for the categorically related relative to the congruent condition in two time windows, 280–400 ms (the time window paralleling our results) and 530–600 ms. Similarly, in the colour-word Stroop task, a fronto-central negativity between 350–500 ms was found for the incongruent relative to the congruent condition (Liotti et al., 2000).

Lexical selection in word production is estimated to start around between 200-250 ms after stimulus onset and to last until around 350 ms (e.g., Indefrey & Levelt, 2004; Sahin et al., 2009). The time window for which we found a significant cluster for the lexical-frequency effect largely agrees with these estimates. Moreover, significant clusters were found between 254 – 500 ms for the Stroop-like effect in the ERPs and 230 – 370 ms for the semantic effect in the TFRs. These time windows overlap with the estimated time window of the lexical selection stage. Finally, taking the lexical-frequency effect as a marker of lexical access in the present data, the time windows of the interference effects also generally overlap with the time window of the lexical-frequency effect. Thus, the results from the ERP and time-frequency analyses do not agree with accounts that localise the semantic and Stroop-like interference effects in a stage prior to lexical selection (i.e., Van Maanen et al., 2009) or the semantic effect prior to lexical selection (i.e., Dell'Acqua et al., 2007). Therefore, the present data are mostly consistent with accounts of semantic and Stroop-like effects that place both effects at the stage of lexical selection (Starreveld & La Heij, 1996; Roelofs, 1992, 2003).

One could argue that the overlap in time of the present effects is in fact, contrary to what we claim, minimal (only between 288 and 370 ms). We would like to argue, firstly, that especially the onset of the overlapping statistical effects is important, rather than their offset. This is because the onset of the stage of lexical selection is, in fact, easier to estimate than its duration. In the case of lexical selection, only conceptual processing occurred before it. Thus the onset of the lexical selection stage can be estimated on the basis of the estimated end of the previous process. Going further down the chain of processes, an accumulated margin of

error makes estimates less precise (see Indefrey, 2011; Indefrey & Levelt, 2004). Secondly, the timing of the effects used in our argument of overlapping timing is based on the exact timing of the statistically significant clusters detected in our data. Although we did not test explicitly for the onset of ERP latency differences (see Kiesel et al., 2008), our time windows for statistical testing were not selected a priori. The statistical method we used, cluster-based permutation, tends to be a conservative test when compared to testing pre-defined time windows (cf. Groppe et al., 2011). From this perspective, the fact that the clusters of the three effects were detected starting between 230 and 288 ms points to an overlap that is not minimal. Thirdly, the lack of an overlap of offsets across the three effects (500 ms for the Stroop-like effect, and 370 and 390 ms for the semantic interference and lexical frequency effects, respectively) may be only apparent. The overlap between the lexical frequency and semantic interference effects is clear. For the Stroop-like effect, the fact that the cluster extends until 500 ms could easily be explained by the overlap of ERP components following the component elicited around 250 ms (e.g., Woodman, 2010). So we do not consider the offset of 500 ms as indicative for the duration of the lexical selection stage in the present study. Given these considerations, the temporal overlap of the three effects becomes more evident. Finally, the onset of the effects is similar to the onset of the effects related to lexical selection reviewed by Indefrey (2011), which provided an updated estimate onset of 200 ms for the lexical selection stage.

An objection could be made to the proportional scaling we applied to our time estimates. For example, some researchers argue that the onset of lexical access is fixed (Costa et al., 2009). However, there is also a reason to assume that a certain amount of rescaling should be used (see also Indefrey, 2011). For example, when presenting pictures alone, perceptual and conceptual encoding will be recruited for processing the picture. But if a distractor word is presented on top of the picture, there is clearly more perceptual information being provided, which potentially affects the perceptual encoding stage. Evidence for this influence can be seen, for example, in the fact that picture naming RTs are always shorter for pictures alone relative to pictures with an unrelated word distractor, a pronounceable pseudoword, or even consonant strings or a series of Xs (e.g., Lupker, 1982). So some sort of proportional scaling of the early stages of perception for picture-word stimuli is not entirely implausible. But crucial for the argument being pursued in the present study, even if one does not rescale the time estimates, then 200 ms should be our reference point. All effects observed in the present study occur later than 200 ms, thus decreasing the plausibility of models which localise the Stroop-like and semantic interference effects in stages preceding lexical selection.

Finally, an apparent discrepancy is found between the present lack of an ERP effect for the semantic interference effect and the presence of such effect in Dell'Acqua et al. (2010). Dell'Acqua et al. obtained two effects related to semantic interference, one starting at 106 ms and the other at around 320 ms. Note, however, that other studies also did not obtain semantic interference effects from distractor words in the ERPs (Aristei et al., 2011; Hirschfeld et al., 2008). Moreover, Dell'Acqua et al.'s interpretation of the effect at 320 ms is that it reflects processing at the lexical selection stage, which is in line with the current proposal. Different from the present ERP study, Dell'Acqua et al. (2010) did not pre-expose the materials to participants prior to collecting ERP and behavioural data. This raises the possibility that the early 106-ms semantic effect observed by Dell'Acqua et al. arose during perceptual/conceptual encoding because of the first-time processing of the pictures and words. In contrast, the later semantic effect (at 320 ms in Dell'Acqua et al. and between 250-370 ms in our study) is obtained regardless of pre-exposure to the materials, which suggests that the semantic effect arising in lexical selection is independent of a familiarization with the pictures

and words.

3.1. Difference in the electrophysiology of semantic and Stroop-like interference

The difference between distractor-word effects in the ERPs (i.e., the Stroop-like effect) and in the TFRs (i.e., the semantic interference effect) raises an important question. If Stroop-like and semantic interference both arise during lexical selection, one could argue that both effects should surface as the same electrophysiological effect, for example, both effects as relative beta-power increase. The difference in electrophysiological effects may be interpreted as evidence for distinct functional loci of Stroop-like and semantic interference effects. For example, one may argue that the semantic effect in the TFRs reflects conceptual level processing, whereas the Stroop-like effect in the ERPs reflects lexical level processing, in line with the functional account of Dell'Acqua et al. (2007). However, a problem with this interpretation is that the assumption of different functional loci for the effects (i.e., conceptual and lexical stages) is not compatible with the timing of the effects in the TFRs and ERPs, which suggests that the electrophysiological effects occur in roughly the same time window. Moreover, in this time window, also the lexical frequency effect occurs, suggesting a lexical selection locus of all three effects (i.e., Stroop-like, semantic, lexical frequency). Therefore, we make a distinction between a functional locus interpretation and a temporal locus interpretation of the present effects, which we explain below.

We assume that the different electrophysiological effects (ERPs vs. TFRs) reflect a functional difference, i.e., the difference in processing and attentional control demands of the words in the different conditions (cf. Roelofs, 2003). The experimental contrast used for assessing the semantic interference effect involves two conditions for which the distractor word is incongruent with the picture name (i.e., categorically related and unrelated distractors words), whereas one of the conditions used in the Stroop-like contrast is a congruent condition (i.e., the name of the picture itself). The (categorically related and unrelated) incongruent conditions may recruit different or additional brain areas, or the same areas to difference degrees, as compared to the congruent condition, where even reading the distractor word would yield a correct response. Evidence suggests that incongruent distractors trigger attentional control processes that deal with the interference, whereas such a process is not (or to a lesser extent) needed for congruent distractors (e.g., Roelofs et al., 2006). The different processing and attentional demands of the different distractor types may affect the nature of the signal that is measured at the scalp, as observed in our experiment. However, given the time course evidence, it seems plausible to assume that these two effects have a common temporal locus, i.e., they emerge during the same planning stage (lexical selection) in the course of word production processes. To conclude, whereas the timing of the Stroop-like, semantic interference, and lexical frequency effects suggests a common lexical locus (the temporal interpretation), the different electrophysiological manifestation of the effects (i.e., ERPs vs. TFRs) suggests differences in processing and attentional demands among the distractor conditions (the functional interpretation).

3.2. Beta oscillations in PWI and word production

Beta-band activity has been reported especially in the sensorimotor domain in relation to motor preparation and execution, and recently also in relation to expectancy (e.g., Engel & Fries, 2010; Neuper et al., 2006). In the language domain, investigations of oscillatory activity so far have been confined to language comprehension (e.g., Bastiaansen et al., 2008; Röhm et al., 2001; Wang et al., in press) whereas there are no reports in the literature related to

language production paradigms (but see Saarinen et al., 2006; Salmelin & Sams, 2002 for oscillatory and motor-cortex activity related to mouth movements).

Since beta desynchronisation has been consistently found in relation to motor preparation (see e.g., Neuper et al., 2006; Saarinen et al., 2006), one could argue that our beta power increase for the semantic effect reflects differences in motor preparation between the categorically related and unrelated conditions. For example, in earlier studies, motor cortex activity was found while preparing and executing mouth movements, quantified by a pattern of suppression followed by rebound of the 20-Hz rhythm (Saarinen et al., 2006; Salmelin & Sams, 2002). Since participants are slightly faster in the unrelated than in the categorically related condition, motor preparation (and thus beta desynchronisation) in the former should start earlier than in the latter. This would explain why there is more power for the related relative to the unrelated condition.

There are reasons to believe, however, that differences in motor preparation between the categorically related and unrelated conditions cannot alone account for the present beta effect. Firstly, evidence suggests that around 250 ms, participants are still in the process of selecting the word to be produced (cf. Indefrey & Levelt, 2004; Sahin et al., 2009). It is unlikely that participants could start preparing the articulatory programme of a given word while not having selected the word. Models of word production agree on the assumption that motor preparation (referred to as phonetic encoding in Indefrey and Levelt, 2004) is the last step before articulation, which in our study happened on average around 750 ms after picture onset. In Sahin et al.'s (2009) study, which used intracranial EEG, phonological encoding did not start before 450 ms in a word generation task, which usually yields RTs around 600 ms (e.g., Roelofs, 2006). Our RTs were certainly longer, which should place the beginning of phonological encoding even further away from 450 ms. Besides, motor representations are only engaged in the last substages of phonological encoding (e.g., Indefrey & Levelt, 2004; Levelt et al., 1999). Moreover, as noted by Sahin and colleagues (2009), motor neuron commands are issued between 50 and 100 ms before speech onset. For the trials included in the EEG analyses in the present study, participants' individual mean RTs for the distractor-type conditions are above 732 ms. Working backwards from the RTs, our participants must have engaged in phonological encoding no earlier than 450 – 500 ms after picture onset. Furthermore, Saarinen and colleagues (Saarinen et al., 2006) observed that the onset of the 20-Hz suppression preceded the mouth electromyogram by no more than 150 ms on average. These time points cannot, of course, be taken as absolute when considering the time-frequency domain since time-frequency estimates are smeared both in time and in frequency.

Regarding the scalp topography of the beta power effect, using magnetoencephalography (MEG), Saarinen and colleagues found that the onset and offset of the 20-Hz activity in the left hemisphere preceded that in the right hemisphere. The scalp topography of our beta effect, on the contrary, is quite right lateralised, although a comparison between MEG and EEG scalp topographies is not straightforward. Finally, if the beta effect was simply reflecting motor preparation, a similar beta power modulation should have been found for the lexical-frequency effect and for the Stroop-like effect, or an even stronger modulation in the latter case, since differences in RTs are larger between the categorically related and congruent conditions than between the related and unrelated conditions.

Note that we do not exclude the possibility that there may be some kind of general motor preparation already at earlier stages of word production, for example, during lexical selection. Participants are engaged in a task for which they know a motor response is required

at every trial so general aspects of preparation may be at play quite early. However, we do not think that this general motor preparation should be condition specific already during the stage of lexical selection. Given the arguments outlined above, altogether, it seems that the beta power increase in the present study cannot be simply accounted for in terms of motor-preparation.

Alternatively, beta activity may relate to the engagement and disengagement of specific brain regions (e.g., Engel & Fries, 2010; Haegens et al., 2011; Jensen & Mazaheri, 2010; Van Wijk et al., 2009; Wang et al., in press). According to this view, neuronal synchronization in a specific band (e.g., gamma band) may reflect the engagement of certain brain areas in processing the current task, whereas other frequency bands (e.g., alpha band) are argued to play a role in inhibiting task-irrelevant areas (see e.g., Jensen & Mazaheri, 2010). The beta oscillations captured by our EEG recordings might be reflecting a similar inhibitory mechanism. In the categorically related condition, the disengagement of processes related to word reading must be stronger than in the unrelated condition. Note that this hypothesis is still speculative. More replications of this effect are needed before conclusions can be drawn regarding what aspects of the lexical-selection process and motor preparation are being reflected in the oscillatory activity. For example, source localisation of the beta modulation could provide very helpful information to help interpret this effect. This question is being pursued in an MEG experiment in our lab (Piai, Roelofs, Bonnefond, & Jensen, in preparation).

3.3. Summary and conclusions

In summary, we investigated participants' overt naming performance in the PWI task while recording their EEG. Naming RTs showed the expected semantic interference and Stroop-like effects. The ERP waveforms for congruent stimuli started diverging statistically from categorically related stimuli around 250 ms with more negative-going deflections than the congruent condition: the Stroop-like effect. The time-frequency analysis revealed oscillatory power increase approximately between 15-27 Hz for categorically related stimuli relative to unrelated stimuli roughly between 230 and 370 ms: the semantic interference effect. Finally, effects of lexical frequency emerged between 288 and 390 ms. The common time window of these effects suggests that both semantic interference and Stroop-like effects emerged during lexical selection.

4. Experimental procedure

4.1. Participants

Twenty native speakers of Dutch (6 male) from the participant pool of the Radboud University Nijmegen participated in the experiment. They received a reward of 15 Euros for their participation. All participants were right-handed. They had normal or corrected-to-normal vision and no history of neurological deficits. Participants gave oral informed consent to their participation after they were completely informed about the nature of the study.

4.2. Materials and design

Forty pictures of common objects were selected from the picture gallery of the Max Planck Institute for Psycholinguistics, Nijmegen, together with their Dutch basic-level names. The pictures were white line drawings on a black background, scaled to fit into a frame of 4

cm x 4 cm. Twenty pictures had low-frequency names (mean: 6.49 per million, range: 0.16 – 14.76) and twenty had high-frequency names (mean: 176.74 per million, range: 42.71 – 827.45). Each target picture was paired with three different distractor words: 1) the Dutch basic-level name of the picture (congruent condition); 2) a semantically related word (categorically related condition); and 3) a semantically unrelated distractor word (unrelated condition). The unrelated condition was formed by re-pairing the pictures with the categorically related distractors, creating semantically unrelated pairs. The lexical frequency of the distractor words was kept within the range of 14.57 – 48.11 per million (mean: 28.59) and each distractor was paired once with a picture from the low-frequency condition and once with a picture from the high-frequency condition. There were 120 picture-distractor pairs and the distractor words were not members of the response set, except for the congruent distractors, which were the names of the pictures. The distractor words were presented in font Arial size 36, colour white. The picture-word pairs were presented three times in a blocked manner, i.e., a given picture could only appear for the second time after all pictures had already been presented once, and so on. The two independent variables (distractor type and lexical frequency of the picture name) were manipulated within-participants. The lexical frequency of the picture name was manipulated between-items and distractor type was manipulated within-items. One unique stimulus list was used per participant using Mix for randomisation (Van Casteren & Davis, 2006). The following constraints were applied: a) there were at least 15 pictures intervening between one presentation of a certain picture and its next presentation; and b) a given distractor type condition and a given lexical-frequency condition could not be repeated more than three consecutive times. The 40 pictures were also used for a naming practice session, in which they appeared twice. One unique randomisation was used per participant for the naming practice.

The experimental pictures were evaluated in a pre-test with respect to differences in ease of recognition, using a picture recognition task (e.g., Özdemir et al., 2007). Ten native speakers of Dutch (3 male), none of which took part in the EEG experiment, performed the picture recognition task. Forty additional pictures, which were used as filler items, were selected from the same picture gallery as for the experimental materials, also subdivided into high-frequency (20 pictures) and low-frequency names (20 pictures), with the same lexical characteristics as the experimental items. One unique randomisation was used per participant with the same lexical-frequency type appearing at most in three consecutive trials. A written word was presented in the centre of the screen for 500 ms. A black screen followed for 500 ms, followed by a picture presented in the centre of the screen for 1 s. Participants indicated with a button press whether the written word and the object referred to the same entity, by pressing the yes button, or not, by pressing the no button. Whether the yes button would be the right or the left button was counterbalanced across participants. In the case of the experimental items, the word and the picture always referred to the same entity; the filler pictures were always preceded by a different word. We analysed only the RTs to the experimental items with correct button presses. Errors were not analysed, due to their very low occurrence. Mean RTs were 455 ms for the high-frequency condition and 467 for the low-frequency condition. These conditions did not differ in ease of recognition, Fs < 1. So we can conclude that, if we find differences in the naming RTs or in the EEG between high-frequency and low-frequency picture names, then these effects are related to lexical effects, since the pictures are recognised equally fast in the two lexical-frequency conditions.

4.3. Procedure and apparatus

Participants were seated comfortably in an electrically and acoustically shielded booth

in front of a computer monitor, approximately 50 cm away from it. The presentation of stimuli and the recording of responses were controlled by Presentation Software (Neurobehavioral Systems). Vocal responses were measured with a voice key. Before the experiment, participants were given a booklet to get familiarised with the pictures used in the experiment and their names. They were instructed to name the pictures and to ignore the printed words. Next, a naming practice was administered, during which the experimental pictures appeared unobstructed on the screen for 1 second, followed by a black screen for 500 ms. Participants named the pictures and were corrected after the naming practice if they had used the wrong name. A PWI practice session of eight trials followed, in which 3 non-experimental pictures were presented with semantically unrelated distractors following the same trial structure as the experimental session. The experiment proper followed the practice sessions.

An experimental trial began with the picture and the distractor word being presented on the centre of the screen for 250 ms. Then a black screen was presented for 1,250 ms plus a jitter. Jitter durations varied randomly between 350, 500 and 750 ms. The use of this trial structure was motivated by findings that anticipated stimuli evoke a slow wave in the EEG (Walter et al., 1964). The absence of a fixation point at the beginning of a trial and the varying jitter durations prevent participants from anticipating the beginning of each trial, thereby minimizing the presence of expectancy slow-waves in the data. The registration of the vocal responses started as soon as the stimuli were displayed on the screen. There were in total six short breaks, during which participants were allowed to drink water and rest, and they indicated when they were ready to proceed. The whole session, including participant preparation, lasted approximately 1 hour and 15 minutes.

4.4. EEG recording and pre-processing

EEG was recorded from 60 scalp electrodes mounted equi-distantly in an elastic cap, positioned according to the international 10-20 system, using the Acticap system, amplified with BrainAmps DC amplifiers (500 Hz sampling, 0.016 – 100 Hz band-pass). Each electrode was referenced on-line to the left mastoid and re-referenced off-line to averaged mastoids. The horizontal electrooculogram was reconstructed from the recorded EEG from the electrodes placed on the left and right temples. The vertical electrooculogram was reconstructed from EEG recorded from the electrodes positioned below and above the left eye. Electrode impedance was kept below 5 kΩ. Four channels (T7, T8, F7, F8) were excluded from subsequent analyses due to large amount of noise in the data of four participants. All EEG analyses were performed using the FieldTrip open source Matlab toolbox (Oostenveld et al., 2011).

4.5. RT analysis

At each trial, the experimenter evaluated the participants' vocal responses. Trials in which the voice key was triggered by a sound which was not the participant's response and naming RTs shorter than 200 ms were discarded. Responses which contained a disfluency, a wrong pronunciation of the word or a wrong response word were coded as errors. These trials were included in the error analysis and were subsequently excluded from the analyses of the naming RTs.

We submitted RTs to by-participant (F_1) and by-item (F_2) analyses of variance with the independent variables distractor type (congruent, categorically related, and unrelated) and lexical frequency (high and low). Moreover, 95% confidence intervals (*CI*) are reported for

the semantic interference, lexical frequency and Stroop-like effects. Errors were submitted to logistic regression analyses (Jaeger, 2008).

4.6. ERP analysis

All trials excluded from the RT analysis were also excluded from the ERP analysis. Additionally, trials with RT lower than 600 ms were also excluded to avoid contamination of the EEG data with artefacts from articulation onset. Single waveforms were filtered with a bandpass filter of 0.1 to 20 Hz. Next, the continuous EEG was segmented into stimulus time-locked epochs, starting at 200 ms before stimulus onset and lasting until 500 ms after stimulus onset. The segments were then baseline-corrected using the average EEG activity from the 200 ms prior to stimulus onset. Trials which contained eye movements, electrode drifting and muscular artefacts within the epoch were rejected. In total, 12.3% of the data was discarded, either already in the RT analysis or after artefact rejection. At least 60 trials remained in each level of each independent variable for each participant. Averaged ERPs were computed for each participant across trials for each level of the distractor type condition (collapsed over lexical frequency) and for each level of the lexical-frequency condition (only the categorically related and unrelated conditions were included). The reason for collapsing over conditions was the following. The congruent condition is from a different nature than the categorically related and unrelated conditions in the sense that in the former, the name of the picture is also the distractor word, whereas in the latter, the distractor words are never the name of the picture, i.e., both conditions are "incongruent" in this respect. If we computed the lexical-frequency brain responses pooling over all conditions, we could affect the nature of the signal. Furthermore, the RTs for categorically related and unrelated conditions tend not to differ more than 50 ms on average, whereas they tend to differ much more from the congruent condition (e.g., Roelofs, 2007). By averaging over conditions with such different RTs, effects could be washed-out. Therefore, we only collapsed over conditions with more similar RTs for the lexical-frequency conditions. Finally, both levels of the lexical-frequency manipulation occurred in all three distractor-type conditions, so in principle, collapsing over levels of one or the other condition is not problematic.

4.7. Time-frequency analysis of oscillatory power

All trials excluded from the RT analysis and trials with RT lower than 600 ms were excluded from the time-frequency analysis. Additionally, trials which contained eye movements, electrode drifting and muscular artefacts within the epoch were rejected (13% of the data). Two participants were excluded from this analysis due to the high occurrence of noise in the higher frequency bands. Note that this difference in rejection rate is due to the low-pass filter applied for the ERPs (20 Hz), which was not used for the time-frequency procedure (and the two excluded participants had too much noise in the frequencies above 20 Hz). In order to remove power line noise, the data was filtered by removing the two coefficients centred at 50 and 100 Hz from its Fourier transform, followed by the inverse Fourier transform. The continuous EEG was then segmented in epochs starting at 200 ms prior to picture onset until 500 ms. To analyse changes in oscillatory power, time-frequency representations (TFRs) of power were computed using a sliding time-window approach. To optimize the trade-off between frequency and time resolution, two different approaches were used. For the low-frequency range (5-30 Hz), power was calculated per trial using 200 ms time-windows. The data in each time-window was multiplied with a Hanning-taper followed by a Fourier transform to get a power estimate. Due to the length of the time-windows, power can only be estimated between 100 ms pre-stimulus and 400 ms post-stimulus. For the high-

frequency range (30-100 Hz), we used a multi-taper approach (Percival & Walden, 1993) with 200 ms sliding time-windows. The data of each time-window was multiplied with an orthogonal set of tapers taken from the Discrete Prolate Spheroidal Sequences. A frequency smoothing of $\Delta f = 16$ Hz was used, resulting in 3 tapers being applied to the data. Power values were then obtained by taking the Fourier transforms of the tapered data-segment and averaging over the tapers per trial (see Nieuwenhuis et al., 2008, for similar settings).

The TFRs of power were then averaged over trials per participant for each distractor type condition and differences in power between conditions were calculated as a relative difference (i.e., the difference in power between two conditions divided by the sum of the power of those two conditions). This relative power difference was subsequently analysed with cluster-based permutation tests.

4.8. Statistical analysis of ERPs and TFRs of power

Significance of the differences between conditions was tested using a non-parametric cluster-based permutation procedure (Maris & Oostenveld, 2007; see also Groppe et al., 2011). This statistical approach allows one to take full advantage of the multiple time and channel information in the data, while dealing with the multiple-comparisons problem. We briefly describe the procedure here, but we refer to Maris and Oostenveld (2007) for a detailed description of the approach.

First, for every channel-time point of the ERPs or channel-time-frequency point of the TFRs of power, a dependent-samples t-value is calculated. Note that these t-values are not used for statistical inference nor are they used to calculate the significance probability of the cluster. For the analyses of the semantic interference and Stroop-like effects, all available time points were taken, i.e., from 200 ms pre-stimulus to 500 ms post-stimulus. For the lexical-frequency analysis, the selection of time points included in the analyses (from 180 ms to 500 ms post-stimulus) was based on the effect found by Strijkers and colleagues (Strijkers et al., 2010). All pairs whose t-values are larger than a pre-determined threshold, in our case ± 1.75, are selected and clustered on the basis of temporal and spatial adjacency (and frequency adjacency for the TFRs). For the spatial adjacency, channels were set to have, on average, two neighbours. For the temporal adjacency, the criterion was one time point and for the frequency adjacency, 1 frequency unit. For each cluster, a cluster-level statistic is calculated by taking the sum of the t-values within that cluster. The significance of the clusters is then calculated with a *Monte Carlo* method. For that, a permutation distribution is created in the following way: A random partition is created by randomly pairing participant averages of one condition to the other condition, followed by calculating dependent-samples t-values. These are then thresholded and subsequently clustered (same as above). The maximum of the cluster-level summed t-values is then taken and selected to enter the permutation distribution. This procedure is repeated 1,000 times. All cluster-level statistics from the observed data are then compared to the resulting permutation distribution. The proportion of random partitions that yielded a larger test statistic than that of the observed cluster is then taken as the *Monte Carlo* estimate of the p-value. Using a critical alpha-level of .05, we conclude that two experimental conditions are significantly different if this *Monte Carlo* p-value is smaller than the alpha-level. The family-wise error rate is kept at .05 because all clusters are compared to the permutation distribution constructed using the maximum cluster-level statistic (cf. Maris & Oostenveld, 2007).

References

Abdel Rahman, R., & Aristei, S. (2010). Now you see it ... and now again: Semantic interference reflects lexical competition in speech production with and without articulation. *Psychonomic Bulletin and Review*, *17*, 657-661.

Anllo-Vento, L., Luck, S. J., & Hillyard, S. A. (1998). Spatio-temporal dynamics of attention to color: Evidence from human electrophysiology. *Human Brain Mapping*, *6*, 216-238.

Appelbaum, L. G., Meyerhoff, K. L., & Woldorff, M. G. (2009). Priming and backward influences in the human brain: processing interactions during the stroop interference effect. *Cerebral Cortex*, *19*, 2508-2521.

Aristei, S., Melinger, A., & Rahman, R. A. (2011). Electrophysiological chronometry of semantic context effects in language production. *Journal of Cognitive Neuroscience, 23*, 1567-1586.

Balota, D. A., Tse, C.-S., Hutchison, K. A., Spieler, D. H., Duchek, J. M., & Morris, J. C. (2010). Predicting conversion to dementia of the Alzheimer's type in a healthy control sample: the power of errors in Stroop color naming. *Psychology and Aging*, *25*, 208-218.

Bastiaansen, M. C. M., Oostenveld, R., Jensen, O., & Hagoort, P. (2008). I see what you mean: Theta power increases are involved in the retrieval of lexical semantic information. *Brain and Language, 106*, 15-28.

Bench, C. J., Frith, C. D., Grasby, P. M., Friston, K. J., Paulesu, E., Frackowiak, R. S., et al. (1993). Investigations of the functional anatomy of attention using the Stroop test. *Neuropsychologia*, *31*, 907-922.

Bub, D. N., Masson, M. E. J., & Lalonde, C. E. (2006). Cognitive control in children: stroop interference and suppression of word reading. *Psychological Science*, *17*, 351-357.

Buzsáki, G., & Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science*, *304*, 1926-1929.

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: a parallel distributed processing account of the Stroop effect. *Psychological Review*, *97*, 332-361.

Costa, A., Strijkers, K., Martin, R. C., & Thierry, G. (2009). The time course of word retrieval revealed by event-related brain potentials during overt speech. *Proceedings of the National Academy of Sciences*, *106*, 21442-21446.

Dell'Acqua, R., Job, R., Peressotti, F., & Pascali, A. (2007). The picture-word interference effect is not a Stroop effect. *Psychonomic Bulletin and Review*, *14*, 717-722.

Dell'Acqua, R., Sessa, P., Peressotti, F., Mulatti, C., Navarrete, E., & Grainger, J. (2010). ERP evidence for ultra-fast semantic processing in the picture-word interference paradigm. *Frontiers in Psychology*, *1*, 1-10.

Engel, A. K., & Fries, P. (2010). Beta-band oscillations-signalling the status quo? *Current Opinion in Neurobiology*, *20*, 156-165.

Eulitz, C., Hauk, O., & Cohen, R. (2000). Electroencephalographic activity over temporal brain areas during phonological encoding in picture naming. *Clinical Neurophysiology, 111*, 2088-2097.

Fagot, C., & Pashler, H. (1992). Making two responses to a single object: Implications for the central attentional bottleneck. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 1058–1079.

Glaser, W. R. , & Glaser, M. O. (1989). Context effects in stroop-like word and picture processing. *Journal of Experimental Psychology: General, 118*, 13-42.

Groppe, D. M., Urbach, T. P., & Kutas, M. (2011). Mass univariate analysis of event-related brain potentials/fields I: A critical tutorial review. *Psychophysiology*, *48*, 1711-1725.

Haegens, S., Handel, B. F., & Jensen, O. (2011). Top-down controlled alpha band activity in somatosensory areas determines behavioral performance in a discrimination task.

Journal of Neuroscience, *31*, 5197-5204.

Heathcote, A., Popiel, S. J., & Mewhort, D. J. (1991). Analysis of response time distributions: An example using the Stroop task. *Psychological Bulletin*, *109*, 340-347.

Hirschfeld, G., Jansma, B., Bölte, J., & Zwitserlood, P. (2008). Interference and facilitation in overt speech production investigated with event-related potentials. *Neuroreport*, *19*, 1227-1230.

Hock, H. S., & Egeth, H. (1970). Verbal interference with encoding in a perceptual classification task. *Journal of Experimental Psychology*, *83*, 299-303.

Indefrey, P. (2011). The spatial and temporal signatures of word production components: A critical update. *Frontiers in Psychology*, *2*(October), 1-16.

Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*, 101-144.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*, 434-446.

Jensen, O., & Mazaheri, A. (2010). Shaping functional architecture by oscillatory alpha activity: Gating by inhibition. *Frontiers in Human Neuroscience*, *4*, 1-8.

Jescheniak, J. D., & Levelt, W. J. M. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 824-843.

Jescheniak, J. D., Meyer, A. S., & Levelt, W. J. M. (2003). Specific-word frequency is not all that counts in speech production: Comments on Caramazza, Costa, et al.(2001) and new experimental data. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 432-438.

Kiesel, A., Miller, J., Jolicoeur, P., & Brisson, B. (2008). Measurement of ERP latency differences: a comparison of single-participant and jackknife-based scoring methods. *Psychophysiology*, *45*, 250-274.

Lachter, J., Ruthruff, E., Lien, M.-C., & McCann, R. S. (2008). Is attention needed for word identification? Evidence from the Stroop paradigm. *Psychonomic Bulletin and Review*, *15*, 950-955.

Laganaro, M., & Perret, C. (2011). Comparing electrophysiological correlates of word production in immediate and delayed naming through the analysis of word age of acquisition effects. *Brain Topography*, *24*, 19-29.

Lage-Castellanos, A., Martínez-Montes, E., Hernández-Cabrera, J. A., & Galán, L. (2010). False discovery rate and permutation test : An evaluation in ERP data analysis. *Statistics in Medicine*, *29*, 63-74.

La Heij, W. (1988). Components of Stroop-like interference in picture naming. Memory and Cognition, 16, 400-410.

Letham, B., & Raij, T. (2010). Statistically robust measurement of evoked response onset latencies. *Journal of Neuroscience Methods*, *194*, 374-379.

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

Liotti, M., Woldorff, M. G., Perez, R., & Mayberg, H. S. (2000). An ERP study of the temporal course of the Stroop color-word interference effect. *Neuropsychologia*, *38*, 701-711.

Lupker, S. J. (1982). The role of phonetic and orthographic similarity in picture-word interference. *Canadian Journal of Psychology*, *36*, 349-367.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, *109*, 163-203.

Mahon, B. Z., Costa, A., Peterson, R., Vargas, K. A., & Caramazza, A. (2007). Lexical selection is not by competition: A reinterpretation of semantic interference and

facilitation effects in the picture-word interference paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 503-535.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*, 177-190.

Mazaheri, A., & Jensen, O. (2010). Rhythmic pulsing: linking ongoing brain activity with evoked responses. *Frontiers in Human Neuroscience*, *4*, 177.

Miozzo, M., & Caramazza, A. (2003). When more is less: A counterintuitive effect of distractor frequency in the picture-word interference paradigm. *Journal of Experimental Psychology: General*, *132*, 228-252.

Morton, J. (1969). Interaction of information in word recognition. *Psychological Review*, *76*, 165–178.

Mulatti, C., & Coltheart, M. (in press). Picture-word interference and the response exclusion hypothesis. *Cortex*.

Müller, M. M., & Keil, A. (2004). Neuronal synchronization and selective color processing in the human brain. *Journal of Cognitive Neuroscience*, *16*, 503-522.

Neuper, C., Wörtz, M., & Pfurtscheller, G. (2006). ERD/ERS patterns reflecting sensorimotor activation and deactivation. *Progress in Brain Research*, *159*, 211–222.

Nieuwenhuis, I.L., Takashima, A., Oostenveld, R., Fernandez, G., & Jensen, O. (2008) Visual areas become less engaged in associative recall following memory stabilization. *Neuroimage, 40*: 1319-1327.

Oldfield, R., & Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology*, *17*, 273-281.

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.

Özdemir, R., Roelofs, A., & Levelt, W. J. M. (2007). Perceptual uniqueness point effects in monitoring internal speech. *Cognition*, *105*, 457-465.

Pardo, J. V., Pardo, P. J., Janer, K. W., & Raichle, M. E. (1990). The anterior cingulate cortex mediates processing selection in the Stroop attentional conflict paradigm. *Proceedings of the National Academy of Sciences of the United States of America*, *87*, 256-259.

Percival, D. B., Walden, A. T. (1993). *Spectral analysis for physical applications: multitaper and conventional univariate techniques*. Cambridge: Cambridge UP. 583 p.

Piai, V., Roelofs, A., & Schriefers, H. (2011). Semantic interference in immediate and delayed naming and reading: Attention and task decisions. *Journal of Memory and Language, 64*, 404-423.

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: Modeling attentional control in the Stroop task. *Psychological Review*, *110*, 88-125.

Roelofs, A. (2006). The influence of spelling on phonological encoding in word reading, object naming, and word generation. *Psychonomic Bulletin and Review*, *13*, 33-37.

Roelofs, A. (2007). Attention and gaze control in picture naming, word reading, and word categorizing. *Journal of Memory and Language*, *57*, 232-251.

Roelofs, A., Piai, V., & Schriefers, H. (in press). Context effects and selective attention in picture naming and word reading: Competition versus response exclusion. *Language and Cognitive Processes*.

Roelofs, A., Turennout, M. van, & Coles, M. G. H. (2006). Anterior cingulate cortex activity can be independent of response conflict in Stroop-like tasks. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 13884-13889.

Röhm, D., Klimesch, W., Haider, H., & Doppelmayr, M. (2001). The role of theta and alpha

oscillations for language comprehension in the human electroencephalogram. *Neuroscience Letters*, *310*, 137-140.

Saarinen, T., Laaksonen, H., Parviainen, T., & Salmelin, R. (2006). Motor cortex dynamics in visuomotor production of speech and non-speech mouth movements. *Cerebral Cortex*, *16*, 212-222.

Sahin, N. T., Pinker, S., Cash, S. S., Schomer, D., & Halgren, E. (2009). Sequential processing of lexical, grammatical, and phonological information within Broca's area. *Science*, *326*, 445-449.

Salmelin, R., & Sams, M. (2002). Motor cortex involvement during verbal versus non-verbal lip and tongue movements. *Human Brain Mapping*, *16*, 81–91.

Sanders, A. F. (1990). Issues and trends in the debate on discrete vs. continuous processing of information. *Acta Psychologica*, *74*, 123-167.

Schnur, T.T., & Martin, R. (2012). Semantic picture-word interference is a post-perceptual effect. *Psychonomic Bulletin & Review*. doi: 10.3758/s13423-011-0190-x.

Starreveld, P. A., & La Heij, W. (1996). Time-course analysis of semantic and orthographic context effects in picture naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 896–918.

Starreveld, P. A., La Heij, W., & Verdonschot, R. G. (in press). Time course analysis of the effects of distractor frequency and categorical relatedness in picture naming: An evaluation of the response exclusion account. *Language and Cognitive Processes.*

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

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*, 643-662.

Szucs, D., & Soltész, F. (2010). Stimulus and response conflict in the color-word Stroop task: a combined electro-myography and event-related potential study. *Brain Research*, *1325*, 63-76.

Van Casteren, M., & Davis, M. H. (2006). Mix, a program for pseudorandomization. *Behavior Research Methods*, *38*, 584-589.

Van Maanen, L., van Rijn, H., & Borst, J. P. (2009). Stroop and picture-word interference are two sides of the same coin. *Psychonomic Bulletin and Review*, *16*, 987-999.

Van Wijk, B. C. M., Daffertshofer, A., Roach, N., & Praamstra, P. (2009). A role of beta oscillatory synchrony in biasing response competition? *Cerebral Cortex*, *19*, 1294-1302.

Verhoef, K., Roelofs, A., & Chwilla, D. J. (2009). Role of inhibition in language switching: evidence from event-related brain potentials in overt picture naming. *Cognition*, *110*, 84-99.

Verhoef, K. M. W., Roelofs, A., & Chwilla, D. J. (2010). Electrophysiological evidence for endogenous control of attention in switching between languages in overt picture naming. *Journal of Cognitive Neuroscience*, *22*, 1832-1843.

Walter, W., Cooper, R., Aldridge, V., McCallum, W., & Winter, A. (1964). Contingent negative variation: an electric sign of sensori-motor association and expectancy in the human brain. *Nature*, *203*, 380–384.

Wang, L., Jensen, O., Van den Brink, D., Weder, N., Hagoort, P., & Bastiaansen, M. C. M. (in press). Beta oscillations relate to the N400m during language comprehension. *Human Brain Mapping.*

Woodman, G. F. (2010). A brief introduction to the use of event-related potentials in studies of perception and attention. *Attention, Perception, & Psychophysics*, *72*, 2031–2046.

Xiao, X., Zhang, Q., Jia, L., Zhang, Y., & Luo, J. (2010). Temporal course of cognitive control in a picture-word interference task. *Neuroreport*, *21*, 104-107.

Appendix
Materials used in the experiment (English translations between parentheses)

Picture name	Categorically related distractor	Unrelated distractor
High frequency		
arm (arm)	heup (hip)	koe
bed (bed)	kast (wardrobe)	kaars
been (leg)	duim (thumb)	appel
broek (trousers)	rok (skirt)	pan
fiets (bicycle)	kar (cart)	hoed
fles (bottle)	pot (pot)	horloge
geweer (rifle)	zwaard (sword)	bal
huis (house)	kasteel (castle)	trommel
jas (jacket)	hemd (shirt)	duif
jurk (dress)	trui (sweater)	lepel
kaas (cheese)	ham (ham)	pistool
kat (cat)	varken (pig)	mes
kerk (church)	fabriek (factory)	pop
koffer (suitcase)	tas (bag)	piano
maan (moon)	zon (sun)	kom
oog (eye)	pols (wrist)	wortel
oor (ear)	teen (toe)	paleis
paard (horse)	beer (bear)	tempel
schoen (shoe)	laars (boot)	konijn
tafel (table)	fauteuil (armchair)	aap
Low frequency		
banaan (banana)	appel (apple)	heup
beker (cup)	lepel (spoon)	duim
dolk (dagger)	pistol (gun)	hemd
egel (hedgehog)	konijn (rabbit)	zwaard
fakkel (torch)	kaars (candle)	pols
giraffe (giraffe)	koe (cow)	kast
gitaar (guitar)	trammel (drum)	kasteel
hert (deer)	aap (monkey)	tas
igloo (igloo)	temple (temple)	teen
kan (jug)	mes (knife)	beer
ketel (kettle)	pan (pan)	fauteuil
molen (mill)	paleis (palace)	trui
muts (tuque)	hoed (hat)	fabriek
tol (spintop)	pop (doll)	varken
tomaat (tomato)	wortel (carrot)	rok
trompet (trumpet)	piano (piano)	ham
vlieger (kite)	bal (ball)	pot
vork (fork)	kom (bowl)	zon
wekker (alarm clock)	horloge (watch)	kar
zwaan (swan)	duif (pigeon)	laars

Figure Legends

Figure 1. Behavioural Data. (A) Box-and-whisker diagram of the naming RTs as a function of distractor type. (B) Vincentised cumulative distribution curves for the high-frequency (pink line) and low-frequency (black line) conditions. RT = response time; ms = milliseconds.

Figure 2. ERPs to Stimulus-type Conditions. (A) ERPs to the three distractor types: congruent (green line), categorically related (red line) and unrelated (blue line). (B) EEG-cap configuration. Each rectangle corresponds to a channel. The red channels are the channels for which the ERPs are shown. There is a one-to-one correspondence between the orientation of the channels in red in 2(B) and the ERPs in 2(A). (C) Scalp distribution of the Stroop-like effect, averaged over the time window 250-350 ms (left) and 350-500 ms (right).

Figure 3. ERPs to Lexical-Frequency Conditions. (A) ERPs to the high-frequency (pink line) and low-frequency (black line) conditions. (B) EEG-cap configuration. Each rectangle corresponds to a channel. The red channels are the channels for which the ERPs are shown. There is a one-to-one correspondence between the orientation of the channels in red in 3(B) and the ERPs in 3(A). (C) Scalp distribution of the lexical-frequency effect (high-frequency minus low-frequency condition), averaged over the time window 288-390 ms.

Figure 4. Time-frequency Data. (A) Power spectrum of the semantic interference effect, averaged over the channels in red in panel B. (B) EEG-cap configuration. Each rectangle corresponds to a channel. The red channels are the channels for which the averaged power spectrum is plotted. (C) Scalp distribution of the semantic interference effect, averaged over the time window 230-370 ms.

A. — Congruent — Categorically related — Unrelated

B.

C. [µV] 250 - 350 ms 350 - 500 ms

288 - 390 ms