Master's thesis

Cognitive Neuroscience – Psycholinguistics Specialisation

Radboud University Nijmegen

The time-course of word-form encoding in second language word production: an ERP study

Jana Hanulová[1]

Supervisor: Peter Indefrey[1,2]

Co-Supervisor: Doug J. Davidson[1,2]

[1]F. C. Donders Centre for Cognitive Neuroimaging, Nijmegen, the Netherlands

[2]Max-Planck-Institute for Psycholinguistics, Nijmegen, the Netherlands

Correspondence to: J. Hanulová, Kapittelweg 29, 6525 EN Nijmegen, the Netherlands

E-mail: jana.hanulova@fcdonders.ru.nl

Abstract

When naming pictures in their second language, even proficient L2 speakers generally initiate articulation later than in their L1 (Christoffels, De Groot, & Kroll, 2006). In a series of tacit naming electrophysiological studies using the N200 and LRP event-related potentials, we investigated whether this difference is due to the later accessibility of the L2 word-form. Proficient Dutch-English bilinguals silently named pictures, on which they first performed a dual-choice, go/no-go manual task based on a semantic (manmade or natural?), and a phonological (does the picture's name start with an "s' or not?) decision. In both languages, the N200 and LRP developed earlier when they were elicited by the semantics of the planned word than when they were based on its phonology. Furthermore, when the translation equivalents of the word initially overlapped (by having the same onset phoneme), the L2 task was easier than when their phonologies were different in each language, suggesting automatic activation of L1 phonology in L2 naming. We did not find consistent evidence that the duration of word-form encoding differs in the two languages. On the basis of these findings, we suggests that L2 word-form retrieval is only slower at postlexical stages of word production.

Keywords: bilingual word production, ERP, N200, LRP, nonselective lexical access

The time-course of word-form encoding in second language word production: an ERP study

Word production involves several processing stages, from conceptual preparation to articulation. The timing of the different sub-processes is a crucial element in developing plausible models of bilingual word production. In the monolingual domain, such information, accumulated from a long tradition of experimental psycholinguistic research, has resulted in computationally explicit models (e.g. Roelofs, 1992, 1997; Levelt, Roelofs & Meyers, 1999; Garrett, 1988; Dell & O'Seaghdha, 1992).

To date, however, little evidence on the time course of individual processes involved in second language (L2) lexical access exists. Moreover, the picture of bilingual word production is further complicated by studies involving between-language priming, picture word interference, spontaneous slips of the tongue and L2/L1 naming latencies, suggesting that the two lexicons of a bilingual are not separate, but interact to some degree (for example Costa, Miozzo & Caramazza, 1999; Colomé, 2001; Costa, La Heij, & Navarrete, 2006 for a review).

In this study, we will apply an ERP paradigm from the monolingual word-production domain (van Turennout, Hagoort, & Brown, 1997, 1998; Schmitt, Münte & Kutas, 2000) to the field of bilingualism to address the issue of the time course of lexical access in the second language, with emphasis on the time course of word-form retrieval. We will also directly address the question of whether the two lexicons of highly proficient L2 speakers influence each other at the level of phonology during single word production.

The time course of monolingual lexical access

There is general agreement in the psycholinguistic literature that before a word is articulated, at least three types of representations must be accessed: the meaning of the word (its conceptual representation), its syntax (represented at the lexical lemma level) and the appropriate word-form representations (Levelt, Roelofs & Meyers, 1999; Garrett, 1988, Dell & O'Seaghdha, 1992). Computational models of monolingual single word production capitalize on detailed information about the temporal dynamics of the retrieval of these three basic kinds of representations. To date, the available timing evidence comes from both behavioral and imaging studies, grouped into essentially three experimental paradigms: behavioral chronometric naming studies, word-picture interference paradigms, and neuroimaging research.

For instance, Schriefers et al. (1990) in a picture-word interference task with different stimulus-onset asynchronies between the distractor and picture presentation, were able to distinguish between an earlier phase of semantic activation and a later phase of phonological activation. More detailed time course estimates are summarized in a metanalysis on word-production studies by Indefrey and Levelt (2004), who proposed the following time windows on the respective retrieval processes during word planning: lemma selection 175-250 ms after the onset of the respective stimulus (e.g. a picture in a picture naming paradigm), followed by word-form access between 250 – 300 ms, syllabification at 330 – 455 ms, and finally phonetic encoding and articulation onset between 455 – 600 ms.

The above estimates include data from a new and promising research paradigm that recently entered the arena of timing studies on word production – event-related potential (ERP) studies. These employed delayed or tacit naming tasks, thus avoiding the problem of motor artifacts in the EEG signal caused by articulatory movements. In a series of production studies, the time course of conceptual and lexical activation processes in monolingual word production has been studied. In a pioneering study, van Turennout, Hagoort, & Brown (1997), made use of the lateralized readiness potential (LRP), and found that in lexical access, semantic activation of the planned word precedes its word-form activation. Later, the same paradigm was applied to probing the time course of syntactic and phonological encoding (van Turennout, Hagoort, & Brown, 1998), and another evoked potential, the N200, was also shown to be relevant for addressing the timing question (Schmitt et al. 2000, 2001) (for application of a similar design in comprehension research, see Rodriquez-Fornells, 2002).

The experimental paradigm devised by van Turennout, Hagoort, & Brown (1997, 1998) combines delayed picture naming with a manual two-choice go/no-go decision. Before naming the stimulus picture, participants were asked to classify it along two dimensions: phonological (whether the picture's name starts with a certain phoneme) and conceptual-semantic (whether the picture depicts an object or an animal), and carry out a button-press contingent on the combined outcome of both tasks. In the "go-semantics" version of the study, the results of the phoneme monitoring task determine whether the left or the right hand is used to press the button, and the animate/inanimate decision determines whether the response is executed or not. In the "go-phonology" version, these contingencies are reversed.

The dependent measure in these tasks was the latency of the LRP, a negatively-going waveform largest over the hemisphere contra-lateral to the responding hand, which is thought to index specific response preparation (e.g. Kutas & Donchin, 1980; Gratton, Coles, Sirevaag, Eriksen, & Donchin, 1988). Single cell recordings in the monkey have shown that the LRP is generated in the motor cortex (e.g. Miller, Riehle, & Requin, 1992); moreover, its onset is directly related to the onset of the motor response, although a motor response is only executed once the LRP reaches a certain threshold level (Gratton et al. 1988). Importantly for the present paradigm, it has been repeatedly demonstrated that the LRP is not postponed until after full stimulus evaluation, but starts developing even after *partial* stimulus evaluation, i.e. as soon as the earliest possible response-related information is available (Smid et al., 1992). That is, if a response is contingent on the result of two consecutive tasks, a negativity over the contralateral motor cortex will begin to develop immediately after the outcome of the first task signals the responding hand to "go". If the result of the second task negates this decision, the developing negativity gradually returns back to baseline.

In the word-production version of the binary go/no-go task, this LRP feature was used to distinguish which lexical code, during word planning, is activated earlier. If, during picture-naming, semantics is activated before phonology, then in the "go-phonology" version of the task an LRP should develop as soon as the semantic code of the word, signaling which response hand to prepare, is activated. Later, when on the basis of the activated phonological code the prepared manual response is inhibited (on "no-go" trials), the LRP returns back to baseline. In the "go-semantics" version of the task, however, the correct hand response can be determined only after retrieval of phonology; therefore, if semantics is activated before phonology, no "no-go" LRP will be observed. This was indeed the pattern of results observed in the van Turennout et al. studies, as well as in their later replications and modifications (Schmitt, Münte, & Kutas 2000; Schmitt, Schiltz, Zaake, Kutas, & Münte, 2001, but see also Abdel Rahman, van Turennout, & Levelt, 2003; Abdel Rahman & Sommer, 2003).

Schmitt and colleagues (2000, 2001) extended the logic of the paradigm to another ERP component, the N200. The no-go N200 is a negatively going, fronto-centrally distributed difference potential which develops about 100 - 300 ms after stimulus onset in trials in which a response is withheld (no-go trials) relative to go-trials (Kok, 1986; Gemba & Sasaki, 1989)[1]. The N200 has been successfully used in go/no-go tasks investigating various cognitive processes (e.g. visual perception by Thorpe et al., 1996), and its onset and peak latencies are used as measures of when in the course of a particular task enough information is gathered to send an inhibition signal to withhold the motor response. Thus, the reasoning here is based on response inhibition rather than preparation: in no-go trials, a manual response can be inhibited only at the time point when the information forming its basis (e.g. semantics) is already available to the system.

An alternative way of looking at the above studies is to focus on absolute estimates of the start and duration of the respective sub-processes of lexical access instead of their relative order, thus capitalizing on the excellent time resolution of the ERP methodology. The LRP and N200 latencies revealed that in monolingual word production, phonology is preceded by semantics by about 120 ms, and syntax is activated about 90 ms after semantics. Furthermore, van Turennout et al. (1998) found that a bisyllabic word is fully phonologically encoded in about 80 ms. One should keep in mind, however, that the relationship of the LRP and N200 estimates and the actual “start” and “duration” of the respective word-planning levels are only indirect. The timing of the LRP or N200 in the EEG signal reflects additional processes as well, such as the conscious decision to make/withhold a manual response based on the semantic/phonological task’s outcome. Moreover, electrical activity arising in the brain does not map onto the ERP signals in a simple, straightforward manner. In this respect, although both ERP measures allow us to detect the events of interest prior to the overt responses, their timing can only be interpreted as upper estimates of when the process in question took place.

While the prospect of “tapping into” individual sub-processes of word production is a very appealing one and the resulting timing data potentially highly informative for computational modeling, criticism has been raised regarding the unnaturalness of the LRP/N200 paradigm (e.g. Jescheniak, Schriefers, Garrett, & Friederici, 2002). Complex, metalinguistic decisions involved in the two binary tasks are in direct contrast to the fast, automatic, effortless, and not consciously accessible processes of semantic encoding and

[1] There are currently two main theories about its functional significance, linking the N200 effect to either motor response inhibition (Sasaki & Gemba, 1993; for a more general inhibition view see Pfefferbaum et al., 1985), or conflict-monitoring, where the N200 marks conflict detection in situations where two response tendencies are simultaneously active (Nieuwenhuis, Yeung, & Ridderinkhoff, 2003).

word-form retrieval in everyday word production. The crucial issue at stake is whether the semantic and phonological "stages" during word-planning are indeed characteristic of normal word production, or whether they are just an artifact elicited by the task at hand. Jescheniak et al. offer a partial answer to this in their ERP version of interference studies with the N400 component, where they found that during picture naming, the semantic and phonological codes are *automatically* activated, and namely in the order delineated by computational models of word production and word-production ERP tasks.

To conclude, there is a general agreement in the monolingual literature regarding the relative order of the individual sub-processes involved in single word production, with semantic information activation preceding the word-form encoding process. Two effective ERP paradigms, with virtually identical results, can be used to not only detect the relative ordering of the processes in question, but importantly, also the timing of these processes with respect to the stimulus onset.

Bilingual word production and nonselective lexical access

Although the basic architecture of bilingual models of word production mirrors the monolingual case (see e.g. Grosjean, 1998; Poulisse & Bongaerts, 1994), precise time-course information, such as that reviewed above for the monolingual domain, are not yet available in any detail. It has been shown that second language word production is less automatic and results in more errors (e.g. Temple, 2000). Another finding from L2 behavioral studies is that naming latencies tend to be significantly longer in L2 than in L1, even for highly proficient L2 speakers (e.g. Kroll, Bob, & Wodniecka, 2006). As semantic representations of first language words and their translation equivalents are hypothesized to be shared (Kroll & Stewart, 1994), naming latency differences between first and second language may be due to slower access to the word-form of L2 words (for a similar assumption see Guo & Peng, 2007).

Major debates in the bilingual production domain (but also in comprehension, e.g. Dijkstra, 2005) revolve around the *locus of selection* problem (for a review, see Costa, La Heij, & Navarrete, 2006). Central to this problem is the question whether the two languages of a bilingual are separated from the very beginning, perhaps by virtue of a conceptually encoded language cue, so that activation spreads exclusively to the lexical representations and word-forms of the target language (La Heij, 2005; for a similar view see Costa et al., 1999; Costa, 2005), or whether the two systems are shared, with a lexicon-external mechanism modulating the relative activation of the lexical candidates (e.g. inhibiting the non-target language lexical items (Green, 1998)). Furthermore, proponents of language nonselectivity hold differing opinions about where the conflict is resolved: either at the lemma level, and therefore non-target language phonology receives no activation (e.g. Hermans, Bongaerts, de Bot & Schreuder, 1998); or at the word-form level, with activation spreading all the way to the level of the nontarget language phonology (Kroll et al., 2000).

Evidence for nonselectivity in L2 lexical access comes essentially from four types of paradigms: cross-language slips of the tongue (Poulisse & Bongaerts, 1994), namely L1/L2 word blends and L1 lexical substitution errors; secondly, naming studies in which cognate words tend to be named faster than noncognates in L2, with some studies also reporting a smaller magnitude effect in the L1 (the so-called *cognate facilitation effect*, Costa, Caramazza, & Sebastian-Gallés, 2000); next, Stroop-like interference paradigms with stimuli and distractors coming from two different languages, reporting effects of the nontarget language distractor on L1 naming latencies (e.g. Kroll et al., 2000); and finally, phoneme monitoring tasks (Rodriquez-Fornells et al., 2005; Colomé, 2001).

The results of the phoneme monitoring studies offer convincing empirical evidence about simultaneous activation of both-language candidates up to the level of phonology. In Colomé's phoneme-monitoring study, highly fluent Spanish-Catalan bilinguals and a control group of Spanish monolinguals had to make a “yes” or “no” response based on the first sound of the presented picture. The pictures belonged to 3 groups: (a) those to which a “yes” was appropriate in the L1 of participants and “no” in the L2), (b) those to which a “yes” was appropriate in the L2 of participants and “no” in the L1, and (c) control pictures in which a “no” was required in both languages of the (bilingual) participants. Colomé found that the ”no” responses were initiated at longer latencies when there was conflict between the L1/L2 response (such as when L2 signaled a “yes” and L1 signaled a “no”) in contrast to control items. No such effect was observed in the monolingual group. In a similar ERP German-Spanish task, trials in which a phoneme-monitoring task would yield conflicting results in the two languages of a bilingual, and trials in which the results would coincide, exerted differential effects on the amplitude of the N200, interpreted by Rodriquez-Fornells et al. (2005) as indicative of a between-language competition located at the word-form level.

Studies of bilingual word production converge on the notion that the two language systems are shared, at least to some degree. However, a detailed look at the experimental paradigms on which most of the evidence for nonselectivity is based reveals that often there is a risk that the tasks might influence the degree of the nontarget language' activation. For example, bilingual picture-word interference paradigms deliberately introduce distractors in the nontarget language; and phoneme monitoring studies, by visually presenting the nontarget-language word's critical phoneme might also induced extra activation of the corresponding nontarget-language phonological candidate, and thus lead to the observed effects (Costa, La Heij, & Navarrete, 2006). In order to avoid this kind of criticism, studies that limit such deliberate activation are necessary.

The outcomes from different studies using various methodologies and subject groups seem to suggest that the locus of language selection, manifested at the behavioral level as different degrees of nontarget language interference, is not universally specified, but depends on many factors, amongst which language proficiency, age of L2 onset and task demands (Kroll et al., 2006), as well as the participants' position on the "language mode" continuum (Soares & Grosjean, 1984), play a crucial role (Finkbeiner, Gollan, & Caramazza, 2006). Under certain conditions, production will (appear to) be selective, because these factors converge to allow rapid lexical and language selection; however, in different contexts, lexical access will be nonselective (Kroll, Bobb, & Wodniecka, 2006).

ERP studies on the timing of lexical processes in L2 production

Recently, two electrophysiological studies investigated the time course of phonological (Rodriquez-Fornells et al., 2005), and semantic and phonological (Guo and Peng, 2007) encoding in the L2 vs. L1. Guo and Peng compared the results of a two-choice go/no-go task in Chinese monolingual speakers (Guo et al., 2005) and moderately proficient Chinese learners of English (Guo & Peng, 2007). Although, in both Chinese and English, the N200 effect based on the semantic decision preceded the N200 effect based on the decision about the words' phonology, there was no statistical difference between the L1 and L2 "semantic-to-phonology" intervals, suggesting that in the L2, word-form information is accessed at about the same time after the word's semantic code, as in the L1. One limitation of the study, however, was that the L1 and L2 were compared between subjects, rendering a direct timing comparisons problematic. It might be that the failure to observe a statistically significant difference between the L1 and L2 was simply a consequence of the not enough sensitive experimental design.

Rodriquez-Fornells et al. (2005) applied a single-choice go/no-go task to the issue of phonology retrieval in the two languages of Spanish-German bilinguals versus a monolingual German control group. Although their study was not primarily focused on timing issues, the results are nevertheless informative. The N200 effect on the phonological decision (consonant vs. vowel at the onset of the word) in the monolingual control group emerged earlier (in the 200 – 400 ms time window) than the two bilingual N200 effects (between 400 – 600 ms). Moreover, the N200 amplitude was significantly more negative for the *non-coincidence* condition, consisting of stimuli for which, in case of language nonselective access, the go/no-go decision would have been conflicting in L1 vs. L2 (i.e. one word started with a vowel in L1 and with a consonant in L2, such as *fresa* and *Erdbeere*) than the *coincidence* group, for which such a decision would always coincide (e.g. *mesa* and *Tisch*). At the same time, there was no significant no-go N200 amplitude difference in the monolingual group. Rodriguez-Fornells et al. interpreted these findings as favoring nonselective lexical access up to the level of phonology, as the higher N200 amplitude presumably indexed a conflict of both automatically activated L1 and L2 candidate word-forms.

One aspect of the design, however, limits the validity of this conclusion. Unlike the control group, which completed the task in one purely monolingual session, the bilinguals carried out the task in alternating blocks of L1 or L2 during a single session. Thus, the ”conflict effect“ between the bilinguals‘ two lexicons might have been an artifact of the task, as *both* language phonologies were intentionally kept active during the task. It remains to be seen whether the same effect would be observed if the two languages were restricted to separate sessions.

In summary, although there are a handful of studies directly or indirectly addressing the question of the timing of word-form retrieval in the L2, and the issue of competition between the two word-forms during production, the answer is still far from clear. As suggested in the previous section, a desirable design would be one employing a stricter between-subjects comparison and a test situation where the participant is kept at the monolingual end of the language mode continuum.

Objectives of the current study

The objectives of the present study are twofold. First, we attempt to investigate the time course of single word production in the participants‘ second language. Secondly, we aim to directly address the issue of *nonselective* bilingual lexical access at the word-form level. By making half of our stimuli coincidental in the two languages with respect to the phonology-classification task (see the *Methods* section), we hope that our temporally sensitive experimental technique will allow us to detect any L1 influence on L2 phonology, and/or vice versa.

In contrast to previous studies (Guo & Peng, 2007), we will employ a within-subject design, in which the individual variability associated with each participant's responses will enter into the two language conditions in the same way. Moreover, we will restrict the instances of L1 and L2 usage to separate sessions, thus avoiding the potential *bias* towards between-language competition (for a similar critique, see Finkbeiner et al., 2006; Kroll, Bobb, & Wodniecka, 2006).

Unlike in the Rodriguez-Fornells et al. (2005) study, our participants are highly proficient Dutch learners of English who are fluent in their L2, but still clearly more dominant in their native language. This choice is motivated by recent claims in the literature regarding the control mechanisms of balanced bilinguals being *qualitatively* different from highly proficient L2 speakers (Costa & Santesteban, 2003; Costa, Santesteban, & Ivanova, 2006). Therefore, we want to establish whether the effects Rodriquez-Fornells et al. observed will still hold for speakers with unequal dominance in their two languages.

Before proceeding to the specific hypotheses of the current study, a short summary of the paradigm's rationale ensues. The logic of the experimental paradigm is based on the following two assumptions: first, in a dual decision go/no-go task, participants prepare their response based on the information that is *first* available to them. Thus, on trials where response is executed (go-trials), we will observe an LRP onset at the time point when the information motivating the response is available to the system. As van Turennout, Hagoort, & Brown (1997) showed, an LRP can develop even on trials where no overt response is observed, given that the participants have the appropriate information to base their manual responses on (such as which hand to prepare). Secondly, information which gets available later, can effectively suppress any existing response preparation. The onset of the response preparation will be evidenced by a developing LRP; and, as soon as response preparation is abandoned, the LRP will return back to the baseline.

A similar reasoning applies to the N200 (Schmitt, Münte, & Kutas, 2000). A signal to inhibit a response can only start at/after the time point when information necessary to make the no-go decision becomes available. Thus, the onset of the N200 (i.e. the difference between the go and the no-go trials) in the version of the task where "go" responses depend on the semantic decision results, will be indicative of the relative time point where enough evidence about the word's semantics is available to the participant in order to withhold a button press. Similarly, the N200 onset in those trials where the go/no-go signal is contingent on the phoneme decision should only develop at/after the time point when the respective information is registered by the participant.

Although the logic of the current task is the same as in van Turennout et al., the logic of the paradigm differs in one important respect. While the main emphasis of the previous studies was on the order of the two stages in question during word production, ours is on the relative *timing* of phonological and semantic encoding during word planning. In other words, we are interested in whether the time it will take subjects to proceed from the activation of semantic information to the (start) of phonological encoding differs between the two languages.

With respect to the theoretical issues summarized above, we hypothesize that:

a) the order of lexical access processes involved in L1 and L2 word production will be identical. In both languages we expect an LRP developing on no-go trials in the "go-phonology" condition, suggesting that the earlier-available semantics motivated the choice of the hand. We also expect that an LRP on go-trials in the "go-semantics" condition will start later than the LRP on go-trials in the "go-phonology" condition *in both languages*. Similarly, in both languages, the onset of the "semantic" N200 effect will precede the "phonological" N200 effect.

b) the activation of the phonological code of L2 words, manifested by the relative latencies of the respective LRP and N200 evoked potentials, will come at a later time point than the same process in the L1 of participants. This will be manifested by a longer latency of the "phonological" N200 in relation to its respective the "semantic" N200 *in English relative to Dutch*, and by a similar LRP effect;

c) there will be a differential effect of the coincidence/non-coincidence stimuli on the N200 amplitude in the L2, and potentially L1 of the participants, expressed in amplitude differences between the respective „phonological" N200 waveforms.

Materials and Methods

Participants

Twenty right-handed native Dutch speakers (12 females), highly proficient in English, with a mean age of 21.4 years, participated in the experiment. Two participants were excluded from the main analysis due to technical problems with EEG recordings (1) or excessive number of eye-blinks (1). All participants can be characterised as late functional bilinguals whose parents were native Dutch speakers living in the Netherlands, and who were not exposed to the L2 before the age of four.

Participants' L2 proficiency was assessed by means of several measures: performance on 50 grammaticality judgment items in the Oxford Placement Test (OPT, Allan, 1992), performance on 60 vocabulary items in a non-speeded lexical decision test (LDT, Meara, 1996, adapted by Lemhöfer, Dijkstra, & Michel, 2004), self-ratings on six 5-point scales, and age of L2 onset. For the group, the OPT mean number of correct answers was 44.3 (SD = 2.89, highest score 50), LDT mean score was 50.1 (SD = 6.25, highest score 59), mean of self-rating on grammar, pronunciation, writing, reading, listening and speaking was 25.4 (SD = 2.5, highest score 30), and mean age of L2 onset was 8.9 years (SD=2.4, range 4 - 12 years). Based on these results, our participants can be described as highly proficient L2 speakers.

All participants gave their written informed consent prior to the experiments and were paid a small fee or assigned course credits for their participation. All had normal or corrected-to-normal vision and were neurologically healthy.

Experimental task

As already sketched in the Introduction, the experimental task involved a dual-task go-no-go decision within the context of (delayed) tacit naming. Prior to (silently) naming a picture in either their L1 or L2, participants had to press the left button or the right button on a button box, based on the results of a *semantic decision*. However, whether they actually carried out the button-press or refrained from responding was determined by the result of a second task, the *phonological decision*. In addition to this "go-*phonology*" version, there was also a "*go-semantics*" version, in which the two contingencies were reversed: for half of the trials within each language, the go/no-go was based on the results of the phonological decision, while the left vs. right hand button press was based on the result of the semantic decision (see *Figure 1*).

In contrast to previous studies, the semantic decision did not involve an animacy decision (animate/inanimate), but rather a manmade/natural judgment. Our main motivation was to increase the absolute number of the available picture stimuli, and at the same time to make sure that the task prompts the participants to access semantic representations of the depicted objects (Glaser, 1992). The manmade/natural judgment has been previously used in tasks concerned with semantic representations (e.g. Zeelenberg and Pecher, 2003). Also the phonological decision was slightly modified in relation to previous studies. While previous studies used a simple binary phonological decision (consonant vs. vowel in Schmitt et al., 2000, and consonant X vs. consonant Y in Van Turennout, Hagoort, & Brown, 1997 and Guo & Peng, 2007), we opted for a slightly different comparison: names *starting* with a particular consonant (“C“) vs. names *not starting* with that consonant (“non-C“). In this way, we could be more flexible in choosing the target pictures, and thus profit from the advantage of using the same set of pictures in both languages, which should substantially reduce excess variance in the between-language comparison. Like previous authors, we assume that in order to make a decision regarding the identity of the first consonant of a particular word, its word-form representation has to be, at least partially, accessed.

Stimuli

For the main experiment, we used 38 black and white line drawings of manmade objects, and 38 drawings of natural objects, all selected from the International Picture Naming Project (IPNP) database (http://crl.ucsd.edu/~szekely/ipnp/sources.html). Half of the names of the depicted objects within these two categories started with certain target phonemes, and the other half did not (for details see below).

As our aim was to choose a set of pictures with fairly homogenous recognition and naming times, we pretested a large group of pictures for naming agreement in both languages, which is highly correlated with naming latencies (Severens, van Lommel, Ratinckx, & Hartsuiker, 2005). We tested 320 object pictures from the IPNP database (all object pictures after exclusion of pictures with Dutch/English cognate names) in an offline naming task with 12 native Dutch speakers attending English-taught university courses, who did not participate in the main experiment. In two sessions separated by 3 weeks, participants received a printed booklet with the pictures, and were asked to indicate their English (session 1) and Dutch (session 2) name(s). Only those pictures that received the same name by at least 75% of all participants in both languages were chosen for further pre-tests.

Next, four groups of pictures were created, with an equal number of items: manmade objects with names starting with different phonemes in Dutch and English (*non-coincidence* condition, e.g. *chair*/*stoel*), manmade objects with names starting with identical phonemes in both languages (*coincidence* condition, e.g. *scissors*/*schaar*); and two groups consisting of natural objects: natural non-coincidence (*girl*/*meisje*) and natural coincidence (*snail*/*slak*) groups.

Pictures from the coincidence and non-coincidence groups were assessed for visual complexity, which is known to be positively correlated with naming and decision latencies (Glaser, 1992). In the *picture complexity pretest*, 15 new right-handed subjects with diverse language backgrounds completed a computerized "old-new task" (see e.g. Van Turennout, Hagoort, & Brown, 1997). On a PC screen, they first viewed 34 randomly ordered filler pictures in a self-paced manner, with the instruction to remember them for later recognition. In the next phase, the 34 fillers were presented together with the 76 target pictures, each staying on the screen for 600 ms, with a 600 ms fixation cross before and 2800 ms blank screen after the picture. Upon seeing each picture, participants had to press the left button in case they have already seen it in the encoding phase (*old*), or the right button for the yet unseen pictures (*new*). As this task did not require the use of language, any significant differences in RTs for the right button presses (i.e. within the set of the target pictures), measured from the picture onset, were interpreted as reflecting the differing length of non-linguistic, perceptual processing of the pictures. An independent t-test of the mean reaction times for the target (*new*) pictures for the coincidence vs. the non-coincidence group showed that the mean perceptual complexity of the two groups was statistically identical ($t(14) = .935$, $p=.35$).

Out of the set of 76 target pictures, the three most common consonants were identified in each language, which formed the basis of the three experimental sub-blocks (B, K, S in Dutch, P, B, and S in English). The three sub-blocks were always presented in a sequence, making up an experimental block. As described in more detail below, each block was presented twice per instruction.

In order to balance the number of the GO/NO-GO and LEFT/RIGHT reactions in each sub- block, a number of fillers were added to the sub-blocks. Moreover, the first 8 items and the last item in each sub-block were always “warm-up“ filler items. In effect, each sub-block consisted of 38 trials with equal numbers of manmade/natural and “C“ vs. “non-C“ items. Overall, each block consisted of 114 pictures, out of which 76 were experimental and 38 were filler items (including warm-up trials). The number of experimental trials in each sub-block ranged from 24 and 27.

A repeated measure ANOVA with log-transformed lexical frequency data taken from the Dutch and English versions of the CELEX database (Baayen, Piepenbrock & van Rijn, 1993) with the factors *language* and *semantic status* revealed no significant differences between the natural and manmade pictures in any of the two languages. There was a significant main effect of language, with the English frequency counts being slightly higher than the Dutch frequency counts ($F(1,75) = 7.32$, $p = 0.01$). However, as the English frequencies in the CELEX database are representative of *native* English speakers, and L2 words of functional bilinguals are generally less often used than their L1 words, a difference in this direction is not problematic for our research question.

The resulting number of trials that could make up each ERP is as follows: all the ERPs in the “go-semantics“ and “go-phonology“ condition consisted of 152 trials per language per ERP (both “go“ and “no-go“ wave), and of 76 for each level of the coincidence condition per language.

Design

Each participant received eight different instructions in each language. In four of them, the response hand was contingent on the phonological decision and the go/no-go task was contingent on the semantic decision. Within each go and no-go decision, both assignments of response hands to decision outcomes were implemented (as left and right-cued instructions), i.e. left button press for the “C“ and right for the “non-C“ picture names, and *vice versa*. In the other four instructions, the left-right button assignment was contingent on the semantic decision and the go/no-go task was contingent on the phonological decision. Again, both left- and right-cued versions of each variant were implemented. The order of consonant sub-blocks within each instruction was randomized. An English example of the resulting eight instructions is listed in *Appendix 1*.

In order to counterbalance the order of instructions across participants, we created four lists of stimuli (A, B, C, D). Because we wanted to minimize the cognitive load due to frequent switching between and within the different instructions (in blocks and sub-blocks), the four “go-semantics“ (numbers 1-4 in *Appendix1*) and the four “go-phonology“ (numbers 5-6 in *Appendix1*) instructions were always presented back-to-back, whilst only the relative order of instructions *within* these two versions was varied. In this way, half of the participants started with the four ”go-semantics“ instructions first (lists A, C), and half with the four “go-phonology“ instructions first (B, D). Within these two alternatives, the order of the go and no-go decisions was varied. Thus, half of those starting with “go-semantics“ began with GO on manmade stimuli (A, order of instructions 1, 2, 3, 4), the other half with GO on natural stimuli (C, order 3, 4, 1, 2). Similarly, in the "go-phonology“-first version, half of the participants made a “GO“ response to pictures whose names started with a certain consonant (B, order of instructions 5, 6, 7, 8), the other half to pictures whose names did not start with a certain consonant (D, order of instructions 7, 8, 5, 6). In this way, each consecutive instruction changed in only one aspect with respect to the previous instruction (either the hands-assignment, or the go/no-go decision basis). The only more complex instruction change followed after a five-minute break in the middle of each eight-instruction run.

Participants received each complete eight-instruction set twice per session, with a twenty minute break in between. The order of instructions was kept constant in both languages.

Procedure

Preparation

The experiment was divided into two sessions, each carried out entirely in either the participant‘s L1 or L2. Half of the participants completed the first session in their L1, the other half in their L2. The two sessions were separated by a week. All interactions with the experimenter, as well as the task instructions, were in the target language of the session.

Each session started with ca 45 minute electrode application, during which participants filled out the English proficiency tests (in the L2 session) and familiarized themselves with the filler and target pictures. The experimenter showed them a printed booklet with all 148 pictures (including training pictures) and asked them to name the pictures in the session's target language. She also provided the correct name in case participants made errors.

Naming pre-test

Next, participants named all the pictures again in a naming pre-test. The purpose of the pre-test was twofold: to check their knowledge of the target words, and to record naming latencies for L1 and L2. Participants were seated behind a PC in a soundproof room and the 76 stimuli interspersed with fillers were presented in a randomized order in three blocks. Each picture was preceded by a fixation cross in the centre of the screen for a randomized period between 1000 – 2000 ms. The target picture stayed on the screen for 1500 ms, during which naming onset was recorded by means of a voice-key. After 1500 ms, the correct name of the target, spoken by an early balanced Dutch-English male bilingual speaker, was played through the loudspeakers as feedback. After an erroneous response, a filler item was inserted, in order to prevent post-error slowing on the consecutive trial(s).

Training

The main experiment was preceded by a training phase in which participants practiced four main types of instructions (see *Design* and *Appendix 1*, instructions bearing odd numbers). Training in the first session started at a much slower pace than the actual experimental pace (5 s response window) and responses were followed by visually presented feedback. Response windows got progressively shorter until experimental pace was reached. Training for the second session did not include longer response windows, as the participants were already familiar with the task.

Following the training phase, the experimenter made sure the participants based their phonological decisions on the sound of the first phonemes, and not on their orthography. Some training pictures were chosen in such a way that this distinction was made explicit (e.g. contrasting the phonemes /ʃ/ vs. /s/ in "sheep" and "sink").

Main experiment

Following the application of EMG electrodes, the main experiment began. In the main experiment, each sub-block started with a set of instructions and eight warm-up trials, which were excluded from further analysis. The sequence of the pictures was pseudo randomized for each participant and for each block, with no more than two items requiring the same response in a row.

Each trial began with a fixation cross presented in the centre of the screen for a randomized interval of 1500 – 3000 ms, after which the picture was presented in the middle of the screen for 1500 ms. Participants were asked to respond during this time interval. After one third of the trials, a naming cue appeared on the screen, urging them to overtly name the picture in the target language of the session. The naming window lasted for 3 s (see *Figure 2*). Each picture was named 6 times per session. The order of naming was pseudo-randomized, with no picture being named twice within the same instruction, and no sub-block having more than 50% of naming responses. The purpose of the naming trials was to prompt participants to always prepare the target names in the respective language to ensure that the tacit naming task effectively elicited (silent) naming. Naming latencies were not further analyzed. All pictures were viewed under an 8ºx8º visual angle. Each block lasted about 5 minutes; a typical session including electrode application and breaks lasted approximately 3.5 hours.

Participants were instructed to minimize their movement and to blink only during overt naming. Their hands were placed on the table in front of them, with both index fingers positioned on the respective button boxes. On the go-trials, they responded with a button press as quickly as possible; on the no-go trials, they were instructed to withhold their response.

ERP apparatus and recordings

Naming latencies in the naming pre-test and push-button latencies in the main experiment were measured from the picture onset, with early responses (RT< 300 ms) excluded from further analysis.

The electroencephalogram was recorded from 30 scalp sites by means of sintered Ag/AgCl electrodes mounted in an electrode cap, referenced to the right mastoid, and later re-referenced to the linked mastoid reference. EOG signal was recorded from the left upper and lower orbital ridge (blinks) and right and left external canthus (horizontal eye-movements). EMG was recorded from two sites on each arm, approximately above the responding muscles *M. Flexor digitorum superficialis* and *M. Flexor digitorum profundus*. The EEG electrodes were spaced according to the 10-20 system. Prior to the analysis, the signal was re-referenced to the mean of the linked-mastoid reference. EEG and EOG impedances were kept below 5 kΩ, EMG impedances below 30 kΩ.

All electrophysiological data recordings and processing were implemented with the Brain Vision Recorder/Analyzer software (Brain Products Inc.). EEG, EOG and EMG recordings were amplified with a BrainAmp DC amplifier with a 200 Hz high-cutoff, a 500 Hz sampling rate and a time constant of 10s. Offline, signals were filtered with a bandpass from 0.01 to 100 Hz and downsampled to 128 Hz.

Segments of 1500 ms after and 200 ms before the onset of picture stimuli were averaged separately for left- and right-cued go and no-go trials for each condition. Epochs in which deflections exceeded 150 µV were rejected. Based on the signal from the EOG channels, we corrected for eye blinks and horizontal eye-movements using the Gratton-Coles & Donchin algorithm (Gratton, Coles & Donchin, 1983).

The lower EMG channel was subtracted from the upper EMG channel for both arms, and the resulting signal was rectified. In order to detect movement on the no-go trials and/or movement of the incorrect finger in the go trials, the collated EMG channels were visually inspected, and in case wrong-channel activity occurred, the respective trial was rejected as error.

Lateralized Readiness Potentials were calculated at C3‘ and C4‘ electrode sites as the difference of activity at sites contralateral and ipsilateral to the responding hand according to the following formula:

LRP = right hand(C3‘-C4‘) – left hand(C3‘ – C4‘)

For each participant, four waves were computed: 1.) hand-semantics, go-phonology; 2.) hand-semantics, no-go-phonology; 3.) hand-phonology, go-semantics; 4.) hand-phonology, no-go-semantics.

The N200 was calculated by subtracting the go-trials from the no-go trials separately for the go-semantics and go-phonology conditions for each subject and each electrode site.

Results

Behavioral data

Picture- naming pre-test: Naming latencies

Data from 2 additional participants were excluded due to microphone failure in one of their sessions. Naming latencies below and above 3 standard deviations from the individual means were excluded. Trials during which errors, dysfluencies or voice-key failures occurred were excluded from the statistical analysis.

A repeated-measures ANOVA with two factors (*language*, *coincidence*) revealed a main effect of *language* ($F(1,15)=7.892$, $p < 0.05$), with naming latencies in English being significantly longer than naming latencies in Dutch (average Dutch naming latency 761 ms (SD = 127), English 862 ms (SD = 135)). No other main effects or interactions were significant, i.e. there was no facilitation effect for words starting with the same phoneme in both languages – neither overall, nor within one of the languages.

Another repeated-measures ANOVA was conducted with error frequencies. None of the above factors turned out to be significant. *Table 1* lists naming latencies and error counts for all 4 groups (*language*(2) x *coincidence*(2)).

	onset phoneme			
language	***coincidence*** *(pauw - peacock)*		***non-coincidence*** *(stoel - chair)*	
	RT (ms)	*errors (%)*	*RT (ms)*	*errors (%)*
Dutch	766 (*134*)	9.5	756 (*124*)	10.3
English	862 (*137*)	12.1	862 (*137*)	9.5

Table 1. *Mean naming latencies from 16 subjects for the Dutch and English picture stimuli, grouped according to whether they started with the same phoneme in both languages (coincidence group) or with different phonemes (non-coincidence group). Standard deviations are in parentheses. Data are in milliseconds, errors are given as percentages (N of trials per condition = 38).*

Main experiment: Push-button response times

To obtain the mean response times for the go-trials, the participants' (N=18) left- and right-hand button press reactions were averaged together, but separately for the "go-semantics" and "go-phonology" conditions in each language. Responses below 200 ms and above 2500 ms were not included in the subsequent analysis.

The average reaction times and their respective standard deviations are summarized in *Table 2*. A repeated-measures ANOVA with the factors *language, decision type* (go-semantics, go-phonology), and *experiment-half* (first vs. second half of the experiment) revealed no main effect of *language*, but a main effect of *decision type* ($F(1,17) = 17.56$, $p<0.001$) and *experiment-half* ($F(1,17) = 186$, $p< 0.000$), as well as a significant interaction of *experiment-half* with *decision-type* ($F(1,17) = 9.9$, $p<0.01$).

Post-hoc testing showed that the push-button latencies were in general shorter in the second half of the experiment (presumably due to practice effects), with faster responses in the *go-semantics* condition within both languages. As revealed by the significant *experiment-half x decision-type* interaction, the "go" responses in the *go-phonology* condition speeded up more than the "go" responses in the *go-semantics* condition (136 vs. 103 ms) in both languages.

The absolute numbers of incorrect responses from both "go" and "no-go" trials (wrong-hand responses, responses longer than 2500 ms, false alarms and misses) were subjected to a repeated-measures ANOVA, using the same factors as above.

The resulting pattern of effects was very similar to the push-button latencies' results. Again, there was no main effect of *language*, but main effects of *decision type* ($F (1,17) = 6.3$, $p < 0.02$) and *experiment-half* ($F (1,17) = 46.2$, $p< 0.000$), yet no significant interactions.

Post-hoc tests revealed that, on average, there were more errors in the *go-phonology* condition (11.2 vs. 9.4), and fewer errors in the second half of the experiment (12.2 vs. 8.4). In absolute numbers, there were more errors in the English part of the experiment (E = 11.2 vs. D = 9.4 per language/participant), but this difference was not statistically significant.

In order to estimate whether the participants used identical response criteria in all the conditions, we calculated the criterion value *c* (Macmillan & Creelman, 1990) for each subject and condition using the following formula[2]:

$$c = -((\Phi^{-1}(H) + \Phi^{-1}(F))/2).$$

The index expresses the distance between the criterion (i.e point where the noise and signal distributions overlap) and a point where neither yes ("go"), nor no ("no-go") response is favored (Stanislaw & Todorov, 1999). A repeated –measures ANOVA using the factors *language* and *decision type* revealed that the *c* values were statistically the same in all the experimental conditions. That is, participants applied roughly equal response criteria in all the experimental conditions.

	decision type			
language	***go-semantics (hand=phonology)***		***go-phonology (hand=semantics)***	
	RT (ms)	*errors (%)*	*RT (ms)*	*errors (%)*
Dutch	838 (*75*)	5.8	890 (*76*)	6.7
English	865 (*79*)	6.5	929 (*103*)	8.2

Table 2. *Push-button latencies for the go-semantics and go-phonology conditions in Dutch and English (in milliseconds). Standard deviations are in parentheses. Total errors (incl. wrong-hand responses, misses and false alarms) are reported as percentages (N of trials per condition = 304).*

Electrophysiological data

LRP

As described in the *Methods* section, electrophysiological recordings were first checked for errors (both performance and measurement errors, such as amplifier drifts), and corrupted trials were excluded. The remaining ca 81% of trials were included in the statistical analyses.

LRP grand averages (*Figure 3*) were calculated as the mean lateralized readiness potentials from the data of 13 participants. 5 participants were excluded from the analyses because of the nonstandard morphologies of their "go" LRPs, which were positive rather than negative. A similar practice was employed in previous LRP word-production studies

[2] Average of the z-score corresponding to the hit rate (H) and to the false alarm rate (F), multiplied by -1 (Stanislaw & Todorov, 1999).

(van Turennout al., 1997; Schmitt et al., 2000), and is recommended in ERP methodological literature (e.g. Handy, 2004).

For each participant and language, we calculated 4 waveforms: the go-responses and the no-go responses averaged across both hands (left, right) for the *go-semantics* and *go-phonology* conditions. Each LRP was calculated in 3 steps: 1.) waveforms representing the (C3' – C4') difference were calculated from all experimental trials in the respective condition; 2.) these were averaged separately for the left-hand cued and right-hand cued trials; and finally, 3.) the left-hand average was subtracted from the right-hand average. Because of the last subtraction, response-unrelated lateralization cancelled out, and the resulting waveform represents the average amount of lateralization due to response preparation (e.g. Luck, 2005).

In order to statistically determine the onset latency of each LRP, the waveforms were quantified as mean amplitude measures relative to the pre-stimulus baseline (-200 ms to 0 ms relative to the picture onset). Their onset and potential offset (for *no-go* waveforms) latencies were measured by means of one-tailed serial t-tests against zero mean in the 250 – 700 post-stimulus interval. This interval was chosen on the basis of previous word-production LRP studies (Schmitt et al., 2000), and visual inspection of the grand-average waveforms. In comparison to Schmitt et al. (2001), its lower boundary is extended by 50 ms.

The t-tests were carried out with step-size 7.8 ms, i.e. each step represented averaged voltage over a 7.8 ms period. For the "go" trials, LRP onset was determined as the first time-point which significantly deviated from the baseline, given that *all the consecutive time points* in the specified time window were significantly above the baseline. For the "no-go" LRPs, the latter criterion was relaxed to 5 consecutive steps, i.e. about 40 ms. Thus, the *offset* point of the "no-go" LRP was represented by the last time point of the significant series.

Figure 3 shows the grand averages (N=13) of each waveform for the 4 conditions in both languages.

Dutch LRP results. In both the *go-phonology* and *go-semantics* conditions, a typical "go" LRP is clearly visible for the go-trials. In no-go trials, a short development of an LRP is detectable in the *go-phonology* condition (returning back to the baseline after a short interval), but not in the *go-semantics* condition.

Statistical analysis confirmed the validity of these observations. In the *go-semantics* condition (*Figure 3A*), the go-LRP significantly deviated from the baseline at 452 ms (after this time-point, all $t(12) < -2.03$, $p < 0.05$). The go-LRP in the *go-phonology* condition (*Figure 3B*) deviated from the baseline at 413 ms (all the following $t(12) < -2.03$, $p < 0.05$). However, with a lower significance criterion $p < 0.1$, there was an additional negativity from about 63 ms before the 452 ms mark, at 390 ms post-stimulus, clearly detectable from the figure (*3B*). In this additional interval, all but two time bins were significantly ($p < 0.05$) more negative than the baseline, with the two 7.8 ms time bins being only marginally significantly different from the zero baseline ($p < 0.1$). Thus, there was a trend for the *go-phonology* LRP to start developing at 390 ms after stimulus onset.

In the case of the no-go trails, statistical analysis confirmed that there was no significant deviation from the baseline in the *go-semantics* condition, whereas there was a significant negativity in the *go-phonology* condition starting at 382 ms (the following 9 t-tests have $t(12) < -2.29$, $p < 0.05$) which lasted until 452 ms. A comparison of the go and no-go LRPs in the *go-phonology* condition by means of a serial paired-samples t-test revealed that the two waveforms started to significantly diverge from each other at 577 ms (after this time point, all $t(12) < -1.79$, all $p < 0.05$).

Thus, the go and no-go LRPs in the *go-phonology* condition started deviating from the baseline at about the same time, and were both significantly different from the baseline but not from each other for about 62 milliseconds (from 390 to 452 ms post-stimulus).

English LRP results. Just as for Dutch, the two English go-LRPs are clearly visible in both response-preparation conditions. Serial t-tests confirmed that the development of the *go-semantics* LRP started at 506 ms post-stimulus ($t(12) < -2.02$, $p < 0.05$). The *go-phonology* LRP began developing at 366 ms after the picture onset ($t(12) < -1.80$, $p < 0.05$).

The two no-go LRPs in the *go-semantics* and *go-phonology* conditions were not significantly different from the baseline in the specified time range.

N200

According to previous word-production studies using the go/no-go paradigm, the increased negativity in the no-go versus go conditions indicates the time points where the necessary information to inhibit a response is available (Schmitt et al., 2000, 2001; Guo & Peng, 2007). Therefore, we expected to see an N200 effect in both conditions and languages. However, based on our predictions, we assumed that the development of the two effects in the *go-semantics* and the *go-phonology* conditions would differ in time, with the N200 onset in the *go-semantics* condition preceding the N200 onset in the *go-phonology* condition. Moreover, we expected a difference in the time *interval* between the onsets of the two effects in the native and second language, with a longer interval in the L2 English.

In both conditions and languages, a clear N200 is visible, largest over the frontal sites (N = 18) (*Figures 4* and *5 A, B*). In order to detect whether the latencies of the effects differ in the predicted manner, we statistically analyzed the onset and peak latencies of the "semantic" (*go-semantics*) and "phonological" (*go-phonology*) N200s. Although a standard practice in the ERP literature is comparing peak latencies (e.g. Handy, 2004), our analyses focused on the onsets of the N200 effects as well, since from the perspective of the temporal course of information encoding, the onsets are also informative of when enough information is available to withhold a response.

All statistical analyses were carried out using the frontal midline *Fz* electrode measurements, as according to the literature (Falkenstein et al., 1999) the no-go N200 effect reaches its maximum at this scalp location[3].

N200 onset latencies

An N200 onset is the first time point over a specified interval at which the go and no-go waveforms significantly diverge from each other. Statistically, this was implemented by means of serial t-tests, in a procedure similar to the one described above. We conducted a series of one-tailed paired-samples t-tests (no-go vs. go waveforms) in steps of 7.8 ms in the time interval between 200 and 700 ms (Schmitt et al., 2000). An onset was determined as the first time point in the no-go waveform that was significantly ($p < 0.05$) more negative than the same time-point in the "go" waveform, given that the next 9 time-bins were also significantly more negative (~ 80 ms).

This procedure resulted in the following values: for Dutch, the *go-semantics* N200 started at 307 ms post-stimulus and lasted for 242 ms; the Dutch *go-phonology* N200 effect started at 393 ms and lasted for 195 ms. For English, the *go-semantics* N200 effect started

[3] Schmitt et al. (2000) used two additional electrodes, the frontal Fp1 and Fp2. However, in our recordings these two electrodes were extremely noisy, as they are located on the subjects' foreheads and are thus most influenced by eye movements and sweating.

at 245 ms post-stimulus and lasted for 273 ms; the English *go-phonology* N200 effect started at 331 ms after the onset of the picture and continued for 280 ms (for a summary see *Table 3*).

That is, the interval (= difference in onset times) between the “semantic” and “phonological” N200 effects was identical for both languages: 86 ms.

N200 peak latencies

The peak latencies of the N200 effects were determined as the time-points where the overall area comprised by the effect (the *no-go* minus *go* waveforms) reached its half (for a similar, 25% peak area measure, see Hillyard & Hansen, 1980). We first calculated the overall negative area under the difference curve of the *no-go minus go* waveforms in the specified time-interval (200-700 ms). In the second step, we determined the time point at which 50% of this area was reached. According to Hillyard and Hansen (1980), the *peak-area latency* is a more robust estimate of the latency of an ERP than the classical peak-latency measure (see also Handy, 2004).

The results are as follows: the mean peak latency relative to the picture onset in the Dutch *go-semantics* condition was 474.5 ms (SD = 100.9) and in the *go-phonology* condition 517.4 ms (SD = 54.4); the mean peak latency in the English *go-semantics* condition was 461.9 ms (SD = 102.3), and in the go-phonology condition 484.0 ms (SD = 89.2).

A repeated-measures ANOVA using these values with the factors *language* and *decision_type* (go-semantics, go-phonology) revealed a marginally significant (F (1,17) = 3.14, p = 0.094) main effect of *decision_type,* with the *go-phonology* peak latencies being later than the *go-semantics* peak latencies. Although the difference between the latencies of the *go-semantics* and *go-phonology* N200 effects seems smaller in English than in Dutch (23 vs. 43 ms), the interaction *language* x *decision_type* yielded no significant effect (F (1.14) = .419, p = .526).

Adding the between-subject factor *proficiency* (based on a median-split of the averaged English proficiency tests’ scores) yielded no significant effects.

type of measure	***language***	***semantics***	***diff.***	***phonology***
LRP onsets	Dutch (L1)	390	**62**	452
	English (L2)	366	**140**	506
N200 onsets	Dutch (L1)	307	**86**	393
	English (L2)	245	**86**	331
N200 peaks	Dutch (L1)	475	**42**	517
	English (L2)	462	**22**	484

Table 3*. Summary of the data indicating the availability of the semantic and phonological codes as revealed by different types of measures (the respective N200 and LRP onsets). In cases where the exact timing is not clear, more values are listed (a lower/upper value range). The middle row represents the difference between the semantic and word-form code availability. Numbers are averaged to the nearest millisecond.*

N200 coincidence analysis

The original coincidence/non-coincidence groups of stimuli were designed according to whether the two translation equivalents coincided or did not coincide with respect to the critical first phoneme, i.e. whether they shared the first phoneme in both languages (such as *peacock* - *pauw*) or not (*chair* - *stoel*). We hypothesized (see also Rodriquez-Fornells et al., 2005) that if second language lexical access is nonselective, i.e. upon producing L2 words also L1 words are automatically activated (and perhaps also vice-versa, see *Introduction*), there should be facilitation vs. inhibition effects based on whether the activated words belong to the coincidence or the non-coincidence group.

This logic would imply that facilitation should arise in a situation where both languages are giving the same “signal” (e.g. “go”) in contrast to a situation where each language is giving a conflicting signal (e.g. Dutch “go” and English “no-go”). The facilitation (or conflict) might be reflected in N200 onset latencies (similar to the *cognate facilitation effect* in L1/L2 picture naming latencies, see *Introduction*), and/or amplitudes), as the signal to withhold a response can have different intensity, depending on the proportion of the conflict/support it gets.

However, in the course of the experiment we discovered that this reasoning did not take one important factor into account – that is, the particular task demands, i.e. the specific phoneme, which is "dictated" by the phoneme decision task. In other words, the "coincidence/non-coincidence" division only makes sense with respect to the decision task. If, for example, the instructions required participants to "go" only if the picture's name started with the sound "s", a *coincidence* stimulus would be of two types: either a pair where both translation equivalents started with an *s* (e.g. *schaar*/*scissors*), or both *did not* start with the target phoneme (e.g. *ober*/*waiter*). On the contrary, in the non-coincidence group, the two starting phonemes would always give the participant conflicting answers, e.g. *stoel/chair*. Here, the Dutch word is telling the participant to go, whereas the English word is telling her to withhold the response.

Unfortunately, our original *coincidence* group consisted only of the first type of stimuli (the *schaar/scissors* type), and the *non-coincidence* group combined the latter two categories in roughly equal proportions (the *ober/waiter* and *stoel/chair* types in relation to the task – phoneme decision on the "s" phoneme). In effect, the original *non-coincidence* group confused conflict and nonconflict stimuli.

Therefore, in case our predictions were true and the non-target language lexical candidates would always enter into the task, there might be a facilitation effect for the *ober/waiter* type of stimuli, and an inhibition (or at least no facilitation) effect for the *stoel/chair* types of stimuli in the "s" phoneme decision task. In fact, the inhibition and facilitation effects would cancel each other out, and we would observe no effects at all. In order to solve this situation, we created two new groups out of the stimuli at hand: the new *coincidence* category comprised all the original coincidence stimuli (same onsets in both languages, N=38), and the new *non-coincidence* group consisted only of those 15 pictures which were truly conflicting with respect to the phoneme decision task (see *Appendix 2*). The new non-coincidence groups in English and Dutch did not contain the same pictures, as the target phonemes in the L1 and L2 phoneme decisions were different (see the *Methods* section); there was only roughly 30% stimuli overlap.

A negative consequence of this re-categorization was that the average log frequency of the Dutch *non-coincidence* stimuli was significantly lower than the average log frequency of the Dutch *coincidence* group (numbers represent log-transformed frequencies for the Dutch coincidence and the *new* non-coincidence groups: 2.82 (SD=0.58) vs. 3.18 (SD=0.50), $p<0.05$). The two new English groups were statistically equal.

The results reported here are thus based on the following number of trials per waveform: the coincidence group 76 (2x38), and the non-coincidence group 30 (2x15) trials, minus about 18% of rejected trials.

As noted above, our hypothesis was not specific with respect to whether the potential influence of the non-target language during target-language production would manifest itself as a timing effect (earlier N200 onset in the “go-phonology” coincidence condition with respect to the non-coincidence condition) or as a relative amplitude difference between the coincidence and non-coincidence conditions, as in the study by Rodriquez-Fornells et al. (2005). Therefore, in addition to the serial t-test onset analyses, we also compared mean amplitudes of the go/no-go effects and looked for effects of coincidence.

N200 onsets for the coincidence/non-coincidence stimuli

Table 4 summarizes the onset times of the *go-phonology* N200 effect in Dutch and English for the coincidence/non-coincidence groups. The Dutch *non-coincidence* go-semantics onset times are not reported, as despite visual appearance, there were not enough statistically significant time-points in the specified time range to denote a clear onset. Similarly, the effect for the English N200 was (marginally) significant only after 450 ms.

A direct comparison of the *go-phonology* N200 effects seems to indicate that while there is a 55 ms head-start for the coincidence *go-phonology* N200 effect relative to the non-coincidence stimuli in Dutch, the two English conditions are virtually identical. However, these differences are difficult to interpret in light of the incomplete data from the *go-semantics* condition.

language	***onset phoneme***	***go-semantics***	***go-phonology***
Dutch	*coincidence*	374	428
	non-coincidence	NS	483
English	*coincidence*	366	397
	non-coincidence	NS	397

Table 4. *Mean onset times of the N200 effects for the coincidence and non-coincidence stimuli in the two response-preparation conditions. The non-coincidence go-semantics onsets were not statistically detectable. Numbers are in milliseconds.*

Mean amplitude measurements[4]

Figure 6 shows the amplitude differences between the coincidence and non-coincidence groups for English. The go and no-go waveforms were quantified as mean amplitudes over 100 ms intervals in the 200 to 700 ms period. A series of repeated measures ANOVA's was conducted for each interval with the factors *go-nogo* and *coincidence*. *Table 5* summarizes the findings.

For the *go-semantics* conditions, there was a significant difference between the no-go and go waveforms (no-go more negative), which corresponds to a semantic no-go N200 effect in this time range for both coincidence and and non-coincidence stimuli.

In the *go-phonology* condition, in addition to a significant no-go effect between 300 and 600 ms after the picture onset, there was also a simultaneous coincidence effect, with the coincidence stimuli go and no-go waveforms being more negative than the non-coincidence go and no-go waveforms. The coincidence effect lasted until about 600 ms after the stimulus onset.

decision_type	factors	200 - 300	300 - 400	400 - 500	500 - 600	600 - 700
go-semantics	go/nogo		6.13*			
	coincidence					
go-phonology	go/nogo		14.47***	5.18**		
	coincidence		12.50**	14.59***	14.02**	

Table 5*. Results of mean amplitude comparisons for the English coincidence and non-coincidence groups. Individual cells contain F-values, significance levels are denoted by stars (* p<0.05, ** p< 0.01, *** p < 0.001). Empty cells mean that none of the factors was significant. There were no interaction effects.*

[4] Because of the reported frequency difference between the Dutch coincidence and non-coincidence groups, we only report coincidence results for the English language. The frequency is a potentially confounding variable, and thus any observed differences are possibly influenced by it.

Discussion

The principal aim of the study was to gain insights into the time course of word-form encoding in the L2 using a well-established tacit picture naming ERP paradigm. The timing of second language word-form encoding was, in a within-subject design, directly compared to the time course of first language word-form encoding, while both values were assessed against a “baseline” of semantic encoding in the respective languages. We employed the onset latencies of the lateralized readiness potential, and the onset and peak latencies of the N200 effect as indicators of when, in the course of tacitly naming the stimulus picture, semantic and word-form information was first available to the participants. In addition, to learn whether during L2 production, L1 word-form representations are also automatically activated, we compared the relative differences in the N200 amplitude for two types of pictures: those, whose names initially overlapped phonologically in the target and non-target languages with respect to the phoneme monitoring task, versus pictures which had conflicting onsets in each language with respect to the monitoring task.

Our study yielded two main findings: firstly, we found that during L2 picture naming, there is a clear temporal difference between the semantic and word-from information encoding, which is in line with previous studies, and supports both serial and cascading models of word production. Secondly, we found that during both first and second language word planning, semantic information encoding precedes word-form encoding. However, the ERP evidence is inconclusive regarding the timing of word-form encoding in the L2 relative to the L1. Whilst according to the LRP results, L2 word-form encoding takes longer in the L2 than in the L1, the N200 findings speak for an equally long semantics-to-phonology interval.

In addition to these main findings, we found some evidence that during L2 word-form encoding, the L1 phonological candidates are also active.

Behavioral data

Consistent with previous studies on second language picture naming, we found that picture naming latencies were longer in the L2 by about 100 ms. The speed of naming pictures in English was not affected by whether the picture names shared their onset phonemes with the Dutch words or not; clearly, a first-phoneme overlap does not lead to a similar facilitation effect as in the case of L1/L2 cognate words (Costa, Caramazza, & Sebastian-Gallés, 2000).

The push-button latencies in the main experiment revealed that when the go-decision is based on the word's semantics, it is approximately 50 – 60 ms faster than when it is based on its phonology. There was neither a main effect of language, nor a significant interaction of language and the type information the go-decision was based on, which suggest that the time interval between the retrieval of semantic and word-form information about the word was not longer in the L2.

LRP

Our Dutch LRP findings replicated the results of van Turennout et al. (1997) and Schmitt et al. (2000) on the time-course of monolingual word production. As predicted, there was a clear LRP detectable on go-trials in both conditions, i.e. when the response hand was determined by the semantic decision and the go/no-go response by the phonological decision, and *vice versa*. Furthermore, there was a short LRP clearly visible on the no-go trails only when the choice of the response hand was contingent on the semantic information, indicating that response preparation was initiated on the basis of the semantic information available prior to word-form information.

The timing of the LRPs in the L1 version of the study fits with the hypothesized time course of L1 word production: the go and no-go LRPs in the *go-phonology* (hand-semantics) condition emerged at roughly equal time-points (382 and 390 ms post-stimulus), indicating that at this time point, the semantic code of the planned word started driving the (left or right hand) response preparation. The two potentials continued developing together for 62 milliseconds, i.e. until 452 ms after the picture onset. The go-LRP in the *go-semantics* (hand-phonology) condition was statistically distinguishable from its respective no-go waveform at 452 ms, which is, in fact, exactly when the two go and no-go LRPs diverged from each other, indicating that around this time-point after the picture's onset, enough phonological information was available to influence the developing manual responses.

These results seem to support the previously delineated time course of L1 semantic and word-form encoding during picture naming, in which encoding of semantic information precedes word-form encoding by several tenths of milliseconds (62 in our case). It is reassuring that even though all the so-far published LRP picture-naming studies on the same matter used a slightly different semantic decision (animacy) or even a different phonological decision (vowel/consonant classification in Schmitt et al.), their findings are of comparable magnitude to ours: 40 ms in van Turennout et al. (1997) and 80 ms in Schmitt et al. (2000) study.

For the L2, the situation is less clear. Just as in the L1 case, the two “go” conditions yielded clear evoked potentials separated in time: the *go-phonology* LRP preceded the *go-semantics* LRP by 140 ms (366 vs. 506 ms post-stimulus), and the *go-phonology* diverged from its no-go at about 522 ms. As expected, there was no detectable LRP on no-go trials in the *go-semantics* condition, indicating that by the time word-form information could direct the choice of the responding hand, the earlier-available semantic code had already halted its response preparation. Contrary to our expectations, though, there was no reliable no-go LRP in the *go-phonology* condition, either.

This could be accounted for in two ways: one possibility is that the time-window of the word-form encoding in the L2 preceded or fully overlapped with the time-window of the L2 semantics encoding, and consequently, there was no partial-response transfer accountable for the no-go LRP development (as in Abdel Rahman & Sommer, 2003, when a difficult (=long) semantic decision almost fully overlapped with the interval during which word-form became available). Such a model is, however, in direct opposition to our English N200 findings as well as the only previously published L2 timing study (Guo & Peng, 2007), which both favour two consecutive, clearly separated information-encoding processes. Another way of explaining the lack of no-go LRP in the English *go-phonology* condition is through response errors: perhaps the no-go LRP was masked by incorrect-hand preparation, i.e. (covert) wrong-hand responses occurring on no-go trials. Wrong-hand preparation surfaces into a positively rather than negatively going LRP (Luck, 2004). If in the L2, a fairly large number of wrong-hand preparations occurred on the no-go trials, it could have “cancelled out” the developing no-go negativity. Clearly, as wrong-hand preparation in the no-go trials is not overtly detectable, we can only estimate its magnitude from the proportion of wrong-hand responses on the go-trials in the respective condition. The number of wrong-hand go-trial responses is, however, only slightly increased in the L2 (on average, 11.2 errors per participant in the L1 vs. 12.5 in the L2), which does not support our explanation either. Thus, at present, we don’t have a fully satisfying account for why there is no predicted no-go negativity observed in the L2, especially in the light of the *go-semantics* LRPs being visible 140 ms before the go-phonology LRP.

This unexpected pattern of results makes any timing interpretation less straightforward than in the L1. One option is to estimate the word-form encoding length from the onset of the two go-LRPs alone: this would lead to the already mentioned 140 ms semantics-to-phonology interval. This strategy, however, is not optimal, as the go LRPs by themselves were not always consistent with the prediction of the experimental paradigm: for example Schmitt et al. (2000) observed the two go-LRPs at basically identical timepoints.

In sum, the available LRP data from 13 subjects seem to speak in favor of a longer duration of the L2 semantics - to - word-form availability interval, with the difference between the two languages being 78 milliseconds.

N200

For both L1 and L2, the values obtained from 18 subjects clearly indicate that semantic encoding precedes word-form encoding during tacit picture naming. Although the N200 logic is based on inhibition rather than response preparation, the relative timeline of lexical access that our findings support is identical with the LRP pattern, which is based on response-preparation data.

Our L1 results replicate Schmitt et al. (2000) findings. The onset and peak of the N200 in the *go-semantics* condition, which develops only after the necessary *semantic* information about the planned word has been retrieved and can feed into response inhibition, precedes the *go-phonology* N200. While the onsets of the two effects are separated by 86 ms (307 vs. 393 post-stimulus), their peaks are a shorter, 42 ms interval, apart (475 – 517 ms).

Contrary to what we found with the LRP measure, with N200 the same 86 ms *onset* latencies' difference in L1 holds for L2 as well. The interval between the N200 *peak* latencies, although considerably closer together than in the case of the onset latencies, is still statistically indistinguishable from the Dutch *semantics to word-form* interval. Interestingly, these L2 findings are in line with the thus far only comparable N200 study by Guo and Peng (2007), who also found that the L1 (Chinese) and L2 (English) semantics to phonology intervals during tacit picture naming were statistically identical.

Although the two ERP measures are fairly consistent in their L1 semantics-to-phonology estimate (62 (LRP) vs. 86 (N200 onsets)), the L2 pattern of findings is inconclusive. An obvious concern with the above accounts is that the two estimates (N200 and LRP) are based on different numbers of subjects; thus, the difference between the LRP and N200 findings might be due to the five extra subjects in the N200 analysis. However, a reanalysis of the N200 onset data for the 13 ("LRP") subjects does not change the N200 pattern of results[5]: there is still practically no difference (15 ms) between the timing of the semantic and phonological encoding, this time with the Dutch interval being slightly longer (101 in Dutch vs. 86 ms in English). Needless to say, the two groups (13 vs. 18 subjects) are essentially identical in English proficiency, which is considered to play an important role in L2 performance (Kroll, Bob, & Wodniecka, 2006). According to the combined means of the two proficiency tests, the two groups are on average virtually equal, their scores being 47.2 (SD = 4.2, N=18) vs. 47.6 (SD = 4.0, N=13) correct answers.

As the conflicting results from the two ERP measures cannot be attributed to the subjects, the difference must lie within the (sensitivity of the) measures themselves. Each ERP capitalizes on a slightly different aspect of the situation: while LRPs are, except for the no-go LRP, a reflection of the motorically realized go-responses, N200 builds upon covert, no-go responses. A possible, although clearly tentative, explanation for the earlier N200/later LRP onsets is that while the N200 is sensitive to very early inhibition-related activity, the LRP only starts to develop once the system registers enough "evidence" for a positive response in order to start a *motor* response preparation. In the interval following the first (early) reaction, additional processes might come into play, which slow down the onset of the response preparation. In the less proficient, less automatic L2 processing, more of this "intervening" cognitive activity might occur, leading to a discrepancy between the respective N200 and LRP onsets. The onset of the N200 is in both languages indeed earlier than the onset of the motor response preparation (see *Table 5*). However, we are not aware of any studies detailing the time course of neural inhibition and response preparation, except for an account by Furster (1997), who claims that during motor tasks, information is first evaluated in the prefrontal cortex (inhibition), and only in the next step is it transmitted to the motor cortex (LRP) (see Schmitt et al., 2000, for a similar point). Indeed, all the available LRP/N200 studies on the time course of word production report the N200 onset

---

[5] For the sake of completeness, we reanalyzed the push-button responses for the 13 subjects who were included in the LRP average. As expected, there was a significant effect of response preparation condition (with the *go-phonology* condition push-button latencies being longer in both languages). While there was no significant interaction between the "language" and "response-preparation" in the N=18 case, this time there was a marginally significant trend ($p < 0.07$) towards to "semantic" – "phonology" interval being slightly longer in English.

as the earliest measure (Schmitt et al., 2000, 2001; Abdel Rahman, van Turennout & Levelt, 2003; Abdel Rahman & Sommer, 2003).

Even though we cannot account for the difference between the two ERP patterns of data, the underlying logic of the timing paradigm speaks in favor of the earlier measure, which is the N200. Both N200 and LRP only indirectly index the availability of the two sub-processes of lexical access: a manual response can be either executed (LRP) or inhibited (N200) once there is enough evidence about the positive/negative result of the respective decision. This implies that even the earliest ERP index is somewhat later than the semantic or word-form encoding processes have taken place. Presumably, the more extra time enters between the actual process and its ERP index, the more different additional processes are likely to be involved. Thus, the earliest overt measure (in our case, the N200 onset timing), although still surfacing only after the actual process has taken place, should have the relatively most direct relation to the actual "start" of the process in question. This, however, does not mean that the LRP measure is useless; it would still be helpful to understand the functional interplay of the N200 – LRP in more detail.

It has been an implicit assumption in our interpretations thus far that the semantic and phonological encoding follow a serial order, i.e. the word-form encoding does not start until semantic access is (at least partially) finished. Instead of individual values we have been looking at intervals; that is the timing of the word-form encoding in relation to the semantic encoding of the planned word: expressed either in terms of the go-semantics N200 to the go-phonology N200 temporal distance, or as the common start of the go and no-go response preparation when hand-choice was contingent on semantics, until their significant divergence. The timing of the semantic encoding process is thus implicitly used in place of a "baseline" towards which the word-form encoding timing is compared. While most of the published ERP timing studies on monolingual and bilingual word-production adopt the same approach, Abdel Rahman and colleagues (Abdel Rahman & Sommer, 2003; Abdel Rahman, van Turennout, & Levelt, 2003) showed that this might not always be good practice, as the two encoding processes might, in fact, run in parallel. In their studies, the authors exploited the difference between an "easy" (short) and "difficult" (presumably longer) semantic decision (or between a blocked and mixed semantic decision), and consequently its influence on the timing of the word-form access. They concluded that the timing of the word-form encoding process is not always contingent on the semantic encoding, as the absolute timing of the word-form encoding did not change together with differing "length" of the semantic encoding stage. It is still possible, however, to reconcile our "interval" estimate with Abdel Rahman's results. As Jescheniak et al. (2002) point out, there might be two ways a word's semantic code is retrieved: under normal conditions, certain "essential" features are retrieved automatically, while other, more specific semantic features (like e.g. the "herbivore/carnivore" features for animals in Abdel Rahman & Sommer 2003 study) are only retrieved "on demand" through a parallel, indirect way. Thus, the longer duration of the difficult semantic decision has no bearing on the "regular" picture naming process, where such a detailed semantic analysis is not necessary. For the present purposes, we conform to the logic delineated by most ERP word-production timing studies, and assume that phonological encoding builds upon the – at least partly encoded – semantic code of the word. Then again, even if we view our findings through the lenses of the parallel account, the interval between the semantic and phonological code availability is still an informative measure, at least in relative terms. Even though it does not directly reflect the timing of the word-form encoding, it can indirectly inform us of the *relative* differences in the time course of the word-form encoding in the L1 vs. L2, as later availability of the process in L2 will still make the L2 semantic-interval longer than the respective L1 interval. Moreover, independent ERP studies where *no conscious task* was required confirmed the automatic availability of the semantic code of the word before its phonological code (Jescheniak et al., 2002).

This brings us to the next point, namely the apparently slightly earlier availability of English semantics with respect to Dutch semantics (see *Table 5*), which is clearly at odds with the predictions of bilingual word-production models. Most published studies on bilingual semantics, spanning across different experimental paradigms, are consistent with the view that translation equivalents, especially concrete objects, share a common conceptual representation (for a review see Francis (2005)). Even if we assume that in order to make the manmade/natural semantic decision, the picture's conceptual/semantic representation must first be accessed indirectly, through its L1 lexical representation, L2 semantics retrieval would still be *slower* (see Zeelenberg & Pecher, 2003, for a discussion about models of bilingual semantic memory). The most plausible option is to assume that the seemingly later access to semantic information in the L1 is a by-product of our decision task. If the manmade/natural decision task is for some reason easier in one of the two languages, this difference might be reflected in the latency of the N200 *go-semantic* onset (or of the respective LRP onset). In English, the opposition “natural/manmade” clearly specifies the two categories in question (things “created” by nature, things made by people). However, such a simple “natural/manmade” distinction had to be translated into Dutch as “door de mensen gemaakt/niet door de mensen gemaakt” (= made/not made by people). It is plausible that this rather descriptive characterization of the two semantic categories makes the decision aspect of the semantic task slightly longer in Dutch than in English. There is no doubt that the participants knew which stimuli fall into each category; however, “translating” this dichotomy into the button press might have taken slightly more time in Dutch than in English. In follow-up experiments, pretests need to ensure that the two, seemingly equal semantics tasks, are indeed equally complex.

Another way of accounting for the apparently earlier “semantics” availability in the L2 is in terms of decision criteria. The observed pattern of results might be due to a more lenient response criterion in English than in Dutch, i.e. participants carried out the manual responses in the L2 even if they were less subjectively certain about their correctness. However, as described in the *Results* section, response criteria were statistically equal in both experimental conditions and languages, which makes this explanation implausible.

Taken together, the results of our earliest available ERP measures show that in the L2, word-form information about the planned word is available at about the same latency after semantics as in the L1. At the same time, pretest data indicate that simply naming our picture stimuli takes on average longer in English than in Dutch. One would, naturally, like to make a direct link between the longer L2 picture naming latencies and more time invested into accessing the right word-form. The crucial question, to which our data do not provide a conclusive answer, is whether the locus of this difference indeed lies in the word-form encoding stage.

Before considering the different options about what our results mean in relation to the L1/L2 naming difference, let us shortly look at to what extent our experimental design dealt with all of the stages of the word-form encoding process. Roelofs (2003), in a series of form preparation experiments, showed that word-form encoding in the L2 copies the general, rightward incremental fashion of the L1 word-form encoding (his *Experiment 1*). This implies that making a phoneme decision on the first phoneme of the stimulus words tapped into the stage where (at least) the word’s first constituent morpheme was retrieved from the lexicon, and phonologically specified with the appropriate phonemic segments. However, it is likely that the phoneme decision on the first phoneme did not yet involve post-lexical processes, i.e. inserting the word’s constituent phonemes into syllable frames and building up the phonological word (after which, phonetic encoding follows).

Knowing this gives us indirect arguments about where we can expect L2 word-form encoding to proceed slower than in the L1. Roelofs (2003) characterizes L2 word-form encoding as a process where representations at the phonemic segments’ level are largely shared between the two languages (especially for phonetically similar languages like Dutch and English); but where the word’s constituent morphemes are not, at least for noncognate L2 words. This implies that any possible L2 “slowing” in picture naming is likely to be caused by less automatic access to the word’s morphemes from memory during the initial stages of word-form encoding.

Our earliest (N200) evidence speaks against this option. As word-form encoding is incremental in a left-to-right fashion, phoneme decision on the first phoneme requires the accessing of at least the first morpheme. Clearly, according to the N200 findings, we observe no reliable slowing at this stage, although LRP data do partly support this.

Another option is that the slowing arises at later stages of word-form encoding, during post-lexical syllabification. Roelofs & Verhoef (2006) support such a view, as they claim that even though phonemic segments are shared in the mind of a bilingual, later aspects of phonological encoding, including syllabification rules, are language-specific. Consistent with this later, post-lexical account, are several lines of evidence: Firstly, Indefrey (2006), in a metaanalysis of bilingual imaging studies, found that the only area which was more strongly activated in L2 than in L1 during picture naming was the left posterior inferior frontal gyrus. This area is, according to Indefrey and Levelt (2004), involved in postlexical syllabification, and not in the early stages of word-form encoding, which are invoked in our task. Moreover, MEG naming studies (e.g. Salmelin, Hari, Lounasmaa, & Sams, 1994) found activation in this area in the 400 – 600 ms post-stimulus, which is slightly later than our N200 “phonology” onset times. These studies jointly imply that the locus of the L1/L2 picture naming slowing is indeed not to be expected before the postlexical syllabification stage.

Thus, a more definite test of this hypothesis would require that the phoneme decision task taps not only into the earlier, but also into the later stages of word-form encoding, such as when the phoneme decision is based on the last phoneme of a polysyllabic word. It is reasonable to expect that for a last-phoneme phoneme decision, the word's morpho-phonological spellout is complete and phonemic segments are inserted into their respective syllable frames, making up the phonological word. This is directly supported by the results of van Turennout et al. (1997), who found that the semantics to phonology interval was about 80 ms longer when the phoneme decision was based on the last phoneme of a disyllabic word than when it was based on its first phoneme (see also Wheeldon & Levelt (1995), for similar findings regarding phoneme decisions on different positions in the target word). Conclusive evidence for the postlexical account of L2/L1 word-form encoding latency difference would be possible if the (last phoneme – first phoneme) difference in the N200 onsets was reliably larger in the L2 than in the L1. Moreover, this difference should get bigger with each new syllable added to the planned word, while the first "starting point" (the N200 onset for the phoneme decision on the beginning of the word) should stay equal for both languages. If, on the other hand, L2 slowing arises at an even later stage, such as retrieving syllable scores prior to articulation, there will be no difference between the L1 and L2 last phoneme decision N200 (LRP) timing.

Some authors (Poulisse & Bongaerts, 1994) conceive of the L2/L1 picture naming latencies' difference as arising purely from differences in the functional frequency of the L1/L2 words. The psycholinguistic literature is not united on where the locus of the frequency effect is. For example, Almeida, Knobel, Finkbeiner & Caramazza (forthcoming) persuasively demonstrated that the frequency effect' locus lies in the lexical stages of word planning, e.g. later than the visual recognition and earlier than articulation; Jescheniak and Levelt (1994) locate it, more specifically, at the lexeme level, but their evidence is consistent also with a phonological word-level location, i.e. in the post-lexical stages of word-form retrieval. Finally, Cholin, Levelt & Schiller's (2006) study argues for a syllable frequency effect, which might arise only once syllable scores are retrieved from the syllabary. Given these different loci for frequency effects, the interpretation of the L1/L2 difference in picture naming latencies in terms of lower L2 functional frequency cannot be ruled out: if, for example, the locus of the frequency effect is indeed at the lexeme level, our N200 data speak against an L1/L2 difference. By contrast, frequency effects arising only later during word-form encoding or even during syllable score retrieval would be compatible with our data.

Nonselective lexical access in L2

Finally, we wish to discuss one additional finding regarding non-selective lexical access in L2 up to the level of phonology. In the English *go-phonology* condition, those L2 stimuli which shared onset phonemes with their L1 translations, led to more negative go and no-go waveforms than the noncoincidental stimuli. Importantly, this effect was specific to the *go-phonology* condition, and it overlapped with the N200 effect.

A presence of a specific, phonology-based, differential ERP amplitude effect seems to hint at a differential influence of the non-target language phonology during L2 phonological processing. We hypothesize that if, in the no-go coincidence condition, both the target (L2) and the non-target (L1) word-forms of the stimuli at hand are active and thus unequivocally support an inhibition response, this can lead to a more pronounced negativity (relative to those L2 stimuli which are in conflict with the, presumably simultaneously active, L1 word-forms). In addition, such an effect should be completely absent from the no-go waveforms in which the inhibitory response is based on the pictures' semantic properties, as at the semantic level, word-forms are not yet active, and there is therefore no conflict/coincidence between the two languages.

This is exactly the pattern of results we observed. However, one finding does not fit into the overall picture: the go-waveforms display the same coincidence effect as the no-go waveform, i.e. they are more negative than the non-coincidence go-waveforms. If the non-target phonology involvement affects the go-responses in the same way as the no-go responses, the go-waveforms amplitude should either not be affected by the stimuli status, or the non-coincidence waveform should even be more negative, as the non-target-language phonological candidates act against the response execution[6]. In our statistical analysis, this would surface as a significant interaction between the *go-nogo* and *coincidence* factors; however, we only observed two main effects.

On the other hand, an interpretation of the coincidence effect as resulting from some general, condition-unrelated factor (e.g. picture effect) is also not appropriate, as it is indeed specific to the *go-phonology* condition, and almost exclusively to the time window of the go/no-go effect (300 – 600 ms post-stimulus), suggesting the involvement of the non-target language phonology.

[6] In effect, if the non-target language stimuli are simultaneously active, they should lead to a small no-go effect, making the non-coincidence go-wave more negative than the coincidence go-wave.

To sum up, we did find a clear coincidence effect arising from simultaneous activation of the target- and nontarget-language phonologies, but we also found evidence that coincidence does not simply modulate the no-go responses, but the go-responses as well.

A “nonselective activation” result is consistent with the results of Rodriquez-Fornells et al. (2005), who found simultaneous activation of the non-target language phonological candidates during both L1 and L2 production. Importantly, we showed that the nonselectivity between the two lexicons seems to hold even when the experimental design restricts the situation to one-language only (by constraining each language to a separate session), and for non-balanced bilinguals.

On the basis of the present results, we cannot conclude whether the non-target language phonological candidates are only activated (Costa 2005; Costa, Miozzo, & Caramazza, 1999), or also selected. A “no difference” result in the naming pretest between the coincidence/non-coincidence groups in both Dutch and English speaks against the “selection” account. However, a more definite test of this claim is clearly necessary.

Conclusions

In a second language, the basic architecture of lexical access corresponds to the L1 case, with semantic encoding preceding word-form encoding. Furthermore, we found evidence that the time interval (40-80ms) between the availability of semantic and phonological information does not differ between L1 and L2. Given that L2 picture naming is nonetheless slower than L1 picture naming, we conclude that the difference must arise at a post-lexical processing stage.

Finally, we have found evidence that during L2 picture naming, L1 phonological candidates are active, suggesting nonselective lexical access in bilingual speakers.

References

Abdel Rahman, R., van Turennout, M., & Levelt, W.J.M. (2003). Phonological encoding is not contingent on semantic feature retrieval: an electrophysiological study on object naming. *Journal of Experimental Psychology: Learning, Memory and Cognition,* 29(5), 850 - 860.

Abdel Rahman, R., & Sommer, W. (2003). Does phonological encoding in speech production always follow the retrieval of semantic knowledge? Electrophysiological evidence for parallel processing. *Cognitive Brain Research*, 16, 372-382.

Almeida, K., Finkbeiner, M., & Caramazza, A. (forthcoming). The locus of the frequency effect: when recognition is not enough. *Psychonomic Bulletin and Review.*

Baayen, H., Piepenbrock, R.,& van Rijn, H. (1993). The CELEX Lexical Database [CD ROM]. Philadephia, PA: University of Pennsylvania, Linguistic Data Consortium.

Christoffers, I.K., de Groot, A.M.B., & Kroll, J.F. (2006).Memory and language skills in simultaneous interpreting: the role of expertise and language proficiency. *Journal of Memory and Language*, 54, 324-345.

Cholin, J., Levelt, W.J.M, & Schriefers, H. (2006). Effects of syllable frequency in speech production. *Cognition*, 99 (2), 205-35.

Coles, M. G., & Gratton, G. (1988). Detecting early communication: using measures of movement-related potentials to illuminate human information processing. *Biological Psychology,* 26 (1-3), 69-89.

Colomé, A. (2001). Lexical Activation in Bilinguals' Speech Production: Language-Specific or Language-Independent? *Journal of Memory and Language,* 45 (4), 721 - 736.

Costa, A., Miozzo, M., & Caramazza, A. (1999). Lexical selection in bilinguals: Do words in bilingual's lexicon compete for selection? *Journal of Memory and Language,* 41, 365-397.

Costa, A., Caramazza, A. & Sebastian-Gallés, N (2000). The cognate facilitation effect: implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26 (5), 1283-1296.

Costa, A. (2005). Lexical Access in bilingual production. In J. Kroll & A.M.B. de Groot (Eds.) *Handbook of Bilingualism: Psycholinguistic Approaches.* (pp.308-325).

New York, Oxford University Press.

Costa, A., & Santesteban, M. (2005). On the facilitatory effects of cognate words in bilingual speech production. *Brain & Language,* 94 (1), 94-103.

Costa, A., Santesteban, M., & Ivanova, E. (2006). How do highly proficient bilinguals control their lexicalization process? Inhibitory and language-specific selection mechanisms are both functional. *J Exp Psychol Learn Mem Cogn*, 32(5), 1057-74.

Costa, A., La Heij, W., & Navarrete, E. (2006). The dynamics of bilingual lexical access. *Bilingualism: Language and Cognition*, 9, 137-151.

Dell, G. S. & O'Seaghdha, P.O. (1992). Stages of lexical access in language production. *Cognition*, 42(1-3), 287-314.

Dijkstra, A. F. J. (2005). Bilingual visual word recognition and lexical access. In J. Kroll & A.M.B. de Groot (Eds.) *Handbook of Bilingualism: Psycholinguistic Approaches.* (pp. 178-201). New York, Oxford University Press.

Finkbeiner, M., Gollan, T., & Caramazza, A. (2006). Lexical access in bilingual speakers: What's the (hard) problem? *Bilingualism: Language and Cognition*, 9, 153-166.

Finkbeiner, M., & Caramazza, A. (2006). Now you see it, now you don't: on turning semantic interference into facilitation in a Stroop-like task. *Cortex*, 42 (6), 790-6.

Francis, W. S. (2005). Bilingual semantic and conceptual representations In J. Kroll & A.M.B. de Groot (Eds.) *Handbook of Bilingualism: Psycholinguistic Approaches.* (pp. 251 - 267). New York, Oxford University Press.

Furster, J. M. (1997). The prefrontal cortex: Anatomy, physiology and neurophysiology of the frontal lobe. Philadelphia, Lippincot-Raven.

Garrett, M. F. (1988). Processes in Language Production. In: F. J. Newmeyer (Ed.) *The Cambridge survey of linguistics.* (69 - 96). Cambridge, Cambridge University Press.

Gemba, H., & Sasaki, K. (1989). Potential related to no-go reaction of go/no-go hand movement task with color discrimination in human. *Neuroscience Letters*, 101(3), 263-268.

Glaser, W. R. (1992). Picture naming. *Cognition*, 42, 61-105.

Gratton, G., & Coles, M.G., (1983). A new method for off-line removal of ocular artifacts. *Electroencephalogr Clin Neurophysiol*, 55(4), 468-84.

Gratton, G., Coles, M., & Donchin, E. (1988). Pre- and poststimulus activation of response channels: a psychophysiological analysis. *Journal of Experimental Psychology: Humand Perception and Performance,* 14(3), 331-44.

Green, D. W. (1998). Mental Control of the Bilingual Lexico-Sematic System. *Bilingualism: Language and Cognition*, 1, 61-81.

Grosjean, F. (1998). The on-line processing of speech: Lexical access in bilinguals. In: Bhatt, P.D.(Ed.) *The Linguistic Brain.* Toronto, Canadian Scholars' Press.

Guo, T., & Peng, D. (2007). Speaking words in the second language: From semantics to phonology in 170 ms. *Neuroscience Research*, 57, 387-392.

Guo, T. M., Peng, D.L., Lu, C.M., & Liu, H.Y. (2005). The temporal course of semantic and phonological activation in Chinese word production: an ERP study. *Acta Psychologica Sinica*, 37, 569-574.

Handy, T. C. (2004). Event Related Potentials: A Methods Handbook. MIT Press.

Hansen, J. C., & Hillyard, S.A. (1980). Endogenous brain potentials associated with selective auditory attention. *Electroencephalogr Clin Neurophysiol,* 49(3-4), 277-90.

Hermans, D., Bongaerts, T., de Bot, K., & Schreuder, R. (1998). Producing words in a foreign language: can speakers prevent interference from their first language? *Bilingualism: Language and Cognition*, 1(3), 213-230.

Indefrey, P., & Levelt, W.J.M. (2004). The spatial and temporal signatures of word production components. *Cognition,* 92(1-2), 101-44.

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., Schriefers, H., Garrett, M. F., & Friederici, A. D (2002). Exploring the activation of semantic and phonological codes during speech planning with event-related brain potentials. *Journal of Cognitive Neuroscience,* 14, 951-964.

Kok, A. (1986). Effects of degradation of visual stimuli on components of the event-related potential (ERP) in Go/Nogo reaction tasks. *Biological Psychology*, 23(21-38).

Kroll, J. F., & Stewart, E. (1994). Category interferences in translation and picture naming: Evidence for asymmetric connection between bilingual memory representation. *Journal of Memory and Language,* 33: 149-174.

Kroll, J. F., Dijkstra, A., Janssen, N. & Schriefers, H. (2000). Selecting the language in which to speak. Experiments on lexical access in bilingual production. 41st Annual Meeting of the Psychonomic Society, New Orleans, LA.

Kroll, J. F., Bobb, S. C., & Wodniecka, Z. (2006). Language selectivity is the exception, not the rule: Arguments against a fixed locus of language selection in bilingual speech. *Bilingualism: Language and Cognition*, 9(2), 119-135.

Kutas, M., & Donchin, E. (1980). Preparation to respond as manifested by movement-related brain-potentials. *Brain Research*, 202: 95-115.

La Heij, W. (2005). Selection Processes in Monolingual and Bilingual Lexical Access. In J. Kroll & A.M.B. de Groot (Eds.) *Handbook of Bilingualism: Psycholinguistic Approaches.* (pp. 289-307). New York, Oxford University Press..

Lemhöfer, K. M. L., Dijkstra, A.F.J., & Michel, M. (2004). Three languages, one ECHO: cognate effects in trilingual word recognition. *Language and Cognitive Processes*, 19(5), 585-612.

Levelt, W. J.M., Meyer, A., & Roelofs, A. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22(1), 1-38.

Levelt, W. J., & Wheeldon, L. (1994). Do speakers have access to a mental syllabary? *Cognition*, 50(1-3), 239-69.

Luck, S. J. (2005). *An Introduction to the Event-Related Potential Technique*. MIT Press.

Macmillan, N. A., & Creelman, C.D. (1990). Response bias: Characteristics of detection theory, threshold theory, and "nonparametric" measures. *Psychological Bulletin*, 107, 101-113.

Meara, P. M. (1996). *English Vocabulary Tests:10k*. Swansea UK, Center for Applied Language Studies.

Miller, J., Riehle, A., & Requin, J.(1992). Effects of preliminary perceptual output on neuronal activity of the primary motor cortex. *Journal of Experimental Psychology: Humand Perception and Performance*, 18(4): 1121-38.

Nieuwenhuis, S., Yeung, N., van den Wildenberg, & Ridderinkhof, W. (2003). Electrophysiological correlates of anterior cingulate function in a go/no-go task: effects of response conflict and trial type frequency. *Cogn Affect Behav Neurosci*, 3(1), 17-26.

Poulisse, N., & Bongaerts, T. (1994). First language use in second language production. *Applied Linguistics*, 15, 36-57.

Rodriguez-Fornells A, v. d. Lugt, A., Rotte M., Britti B, Heinze HJ, & Münte TF (2005). Second language interferes with word production in fluent bilinguals: brain potential and functional imaging evidence. *Journal of Cognitive Neuroscience*, 17(3), 422-433.

Rodriquez-Fornells, A., Rotte, M., Heinze, H.J., Noesselt, T., & Muente, T.F. (2002). Brain potential and functional MRI evidence for how to handle two languages with one brain. *Nature,* 415, 1026-1029.

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, 42(1-3), 107-42.

Roelofs, A. (1997). The WEAVER model of word-form encoding in speech production. *Cognition*, 64(3), 249-84.

Roelofs, A., & Verhoef, K. (2006). Modeling the control of phonological encoding in bilingual speakers. *Bilingualism: Language and Cognition,* 9, 167-176.

Roelofs, A. P. A. (2003). Shared phonological encoding processes and representations of languages in bilingual speakers. *Language and cognitive processes,* 18(2), 175-204.

Salmelin R, Hari, R., Lounasmaa O., & Sams M. (1994). Dynamics of brain activation during picture naming. *Nature*, 368, 463-465.

Schmitt, B. M., Munte, T.F., & Kutas, M. (2000). Electrophysiological estimates of the time course of semantic and phonological encoding during implicit picture naming. *Psychophysiology,* 37(4), 473-84.

Schriefers, H., Meyer, A. S., & Levelt, W. J. M. (1990). Exploring the time course of lexical access in language production: picture-word interference studies. *Journal of memory and language*, 29(1), 86-102.

Severens, E., Van Lommel, S., Ratinckx, E. , & Hartsuiker, R. J. (2005). Timed picture naming norms for 590 pictures in Dutch. *Acta Psychologica*, 119 (2), 157 – 189.

Smid, H. G., Mulder, G., Mulder, L.J., & Brands, G.J. (1992). A psychophysiological study of the use of partial information in stimulus-response translation. *Journal of Experimental Psychology: Humand Perception and Performance*, 18(4), 1101-19.

Soares, C. & Grosjean, F. (1984). Bilinguals in a monolingual and a bilingual speech mode: The effect on lexical access. *Memory and Cognition,* 12(4), 380-386.

Stanislaw, H. & N. Todorov (1999). Calculation of signal detection theory measures. *Behav Res Methods Instrum Comput*, 31(1), 137-49.

Thorpe, S., Fitze, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, 381, 502-522.

Turennout, M., Hagoort, P., & Brown, C. M. (1998). Brain activity during speaking: from syntax to phonology in 40 milliseconds. *Science,* 280 (5363): 572-4.

van Turennout, M., Hagoort, P., & Brown, C.M. (1997). Electrophysiological evidence on the time course of semantic and phonological processes in speech production. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 23(4), 787-806.

Zeelenberg, R., & Pecher, D. (2003). Evidence for long-term cross-language repetition priming in conceptual implicit memory tasks. *Journal of Memory and Language,* 49, 80-94.

Acknowledgments

I want to thank Peter Indefrey and Doug Davidson for their weekly supervision; I'm also grateful to Sander Baerends and the entire TG group for their help with the EEG recordings. Thanks to Kirsten Weber for her help and support; and the same goes for Henk Cremers, Giovanni Piantoni, Michael Ploechl, Julia Udden, Iris Grothe, Christiane Fritz, Ferry Hommerson, Frauke Hellwig and all my friends from the Cognitive Neuroscience Master Program. Thanks to Micha Klein and Jonas Moons for our discussions, To Rene Scheeringa for helpful BVA advice, and to Tildie for her cheerful presence. And, most of all, thanks to my Ivan.

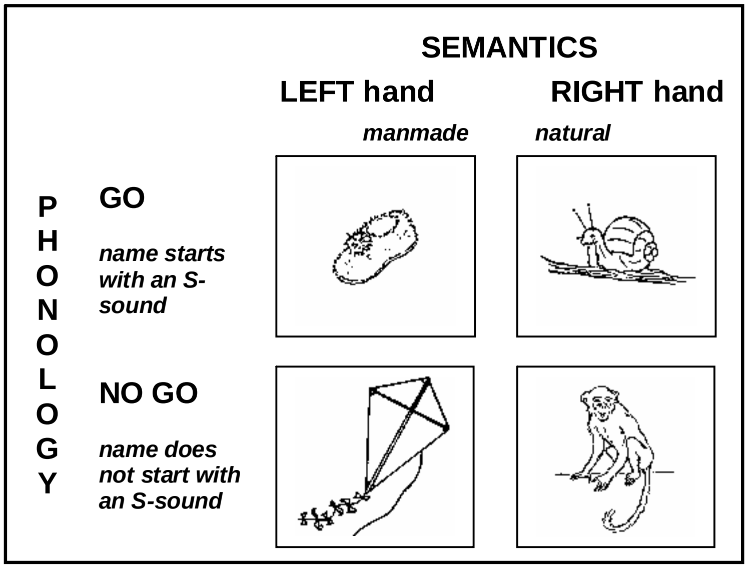


Figure 1. An example of the experimental design in the go-phonology/English version. The go/no-go response is contingent on the phonological decision and the choice of the response hand is contingent on the semantic decision. In the go-semantics version, these contingencies were reversed. (Depicted objects: slipper, snail, kite, monkey)

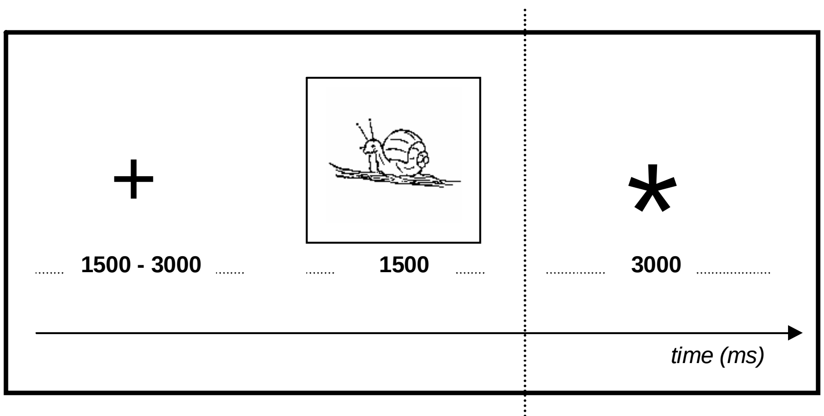


Figure 2*. A typical trial set-up: A fixation cross was presented in the centre of the screen for a randomized interval between 1500 – 3000 ms. Next, the picture appeared on the screen for 1500 ms, during which manual response was to be made. On one-third of the trials, a naming cue was presented for 3000 ms.*

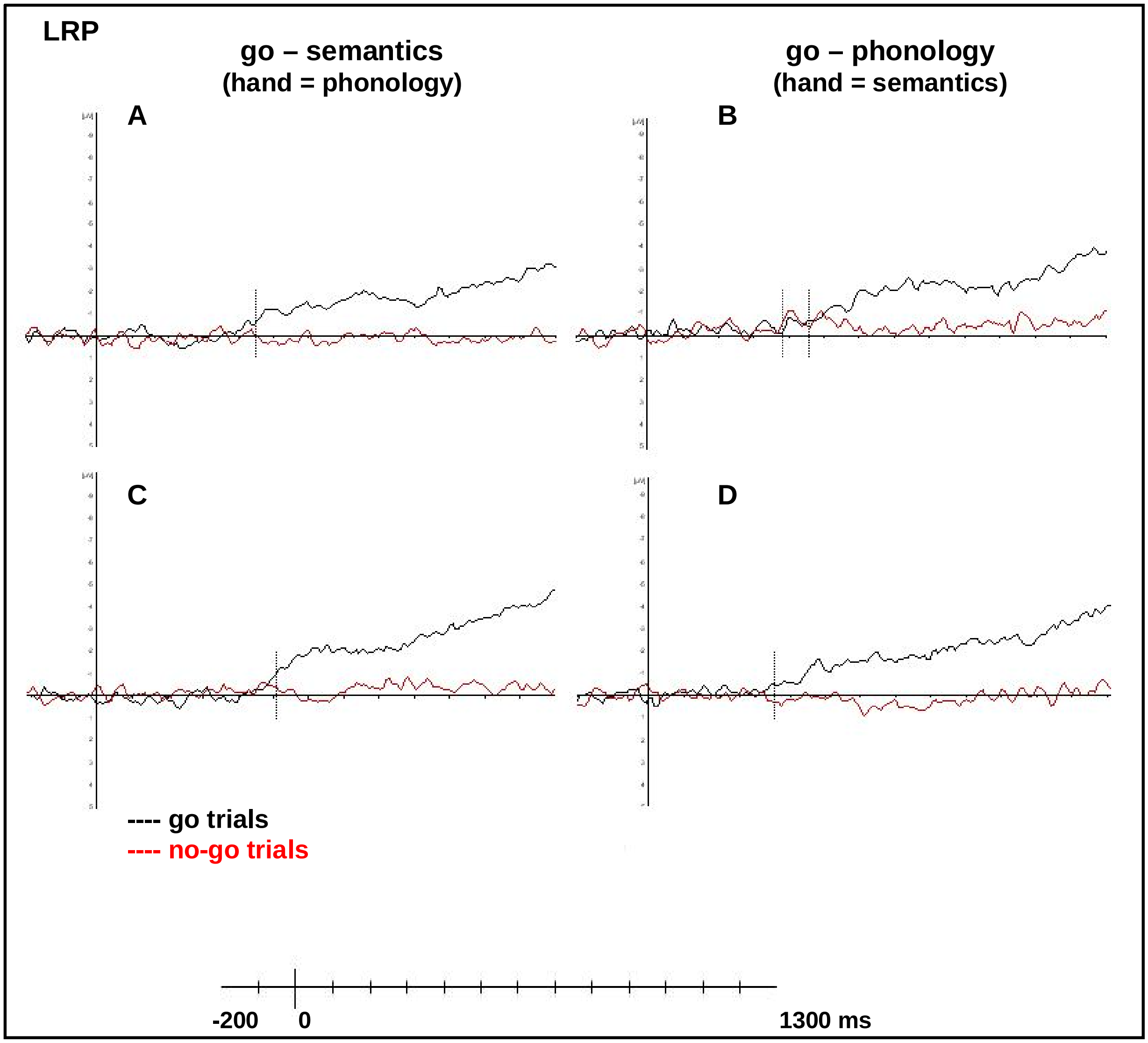


Figure 3. *Dutch (first row) and English (second row) LRPs in the go-semantics (A,C) and go-phonology conditions (B,D). Vertical lines indicate the points where the go-waveforms significantly diverge from the baseline. The second vertical line on 4.B indicates where the go and no-go LRPs significantly diverge from each other. Voltage ranges from +5/-9 μV. Negativity is plotted upwards. N=13*

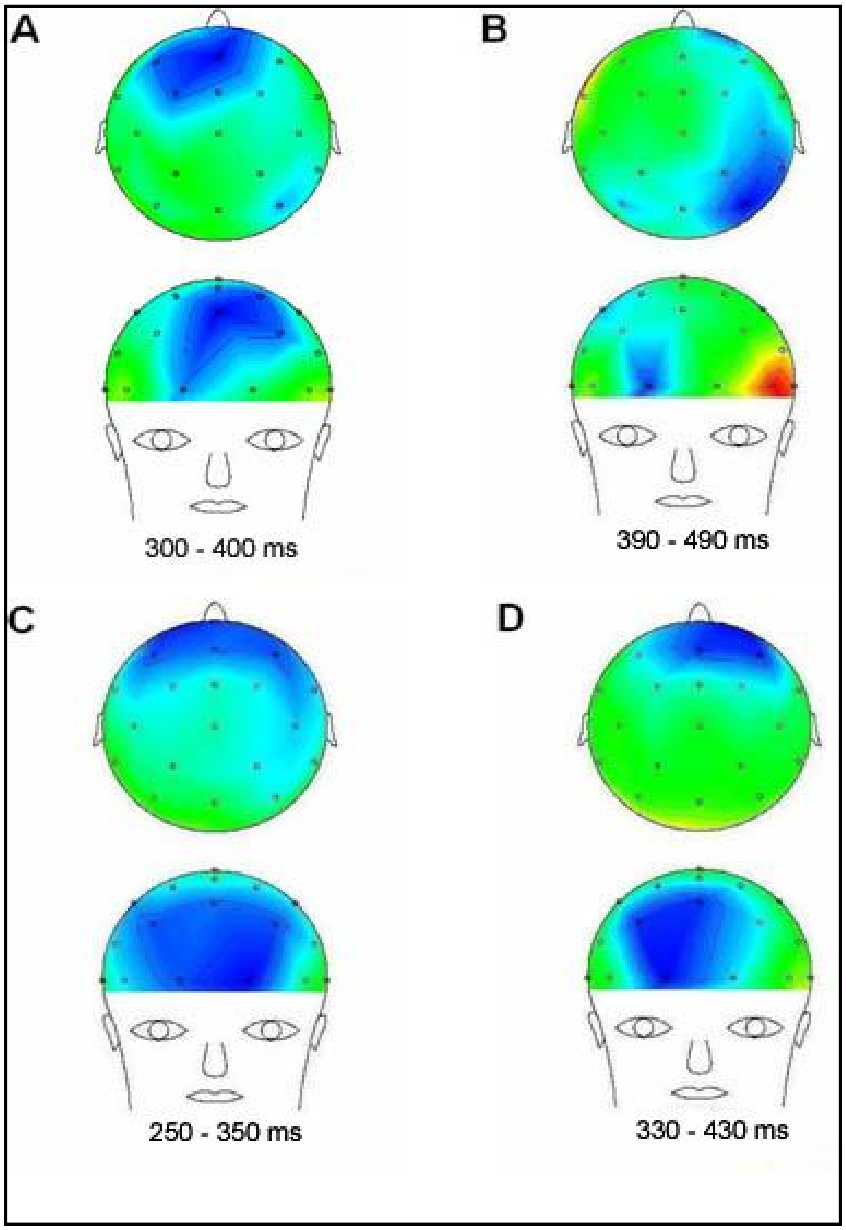


Figure 4. *Topographical maps of the no-go minus go difference averaged over 100 ms intervals starting at the onsets of the respective N200 effects. There is a frontal maximum visible in each condition (dark blue). Please note that the voltage is relative to the maximum and minimum voltage of each individual effect (i.e. the red color in 3.B corresponds to 0, not to a positivity.) The individual voltages are as follows (in microvolts):* ***A*** *(Dutch go-semantics) -1.4/0.8;* ***B*** *(Dutch go-phonology) -1.2/0;* ***C*** *(English go-semantics) -0.9/-0.1;* ***D*** *(English go-phonology) -1.2/-0.1*

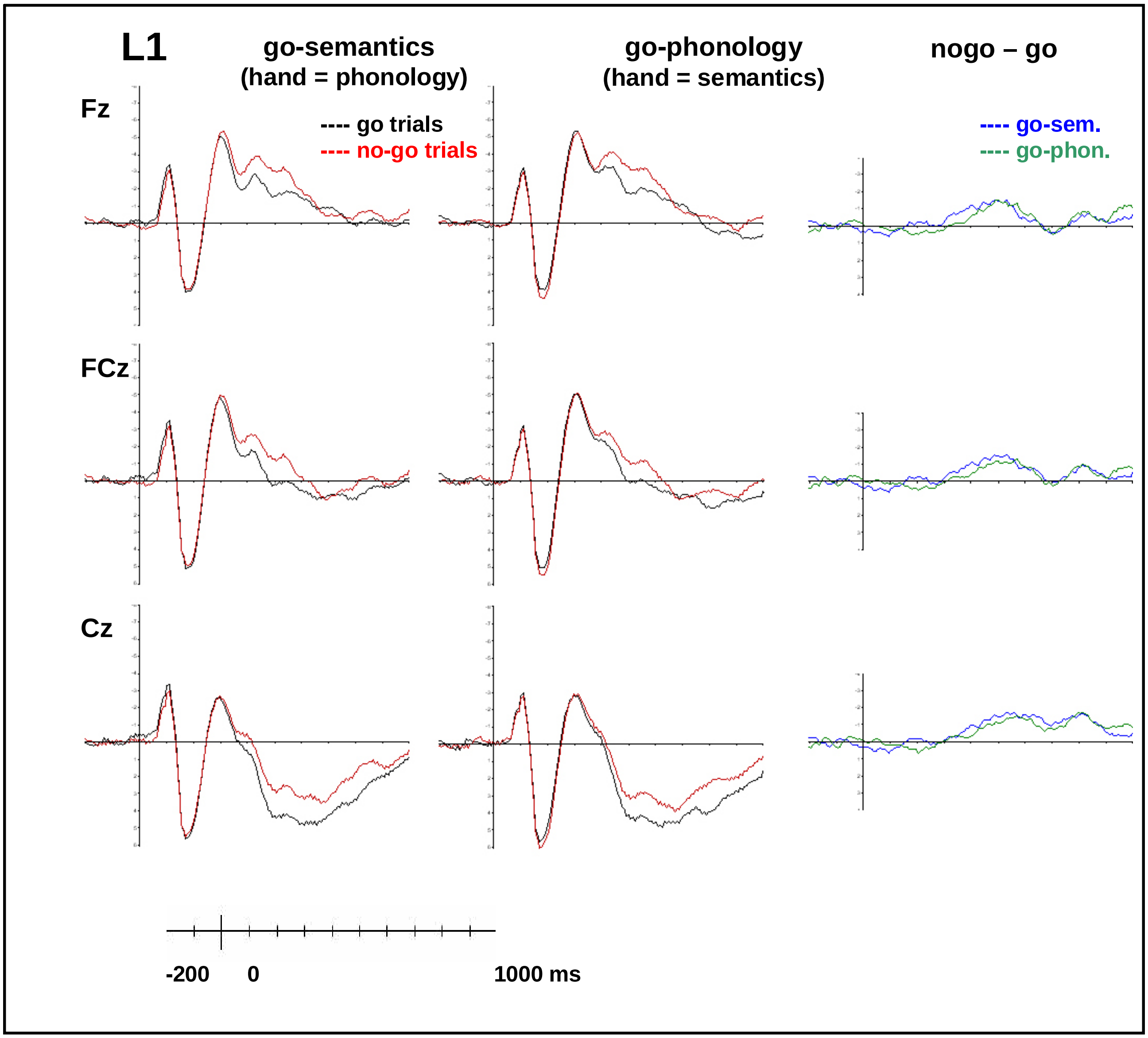


Figure 5A. *The Dutch N200 grand averages at midline electrode sites. (Left column = "go-semantics" condition; middle column = "go-phonology" condition; right column = "no-go minus go" difference waves.) Negativity is plotted upwards. Voltage ranges +6/-9 µV for N200; +4/-4 µV for difference waves.* *N=18*

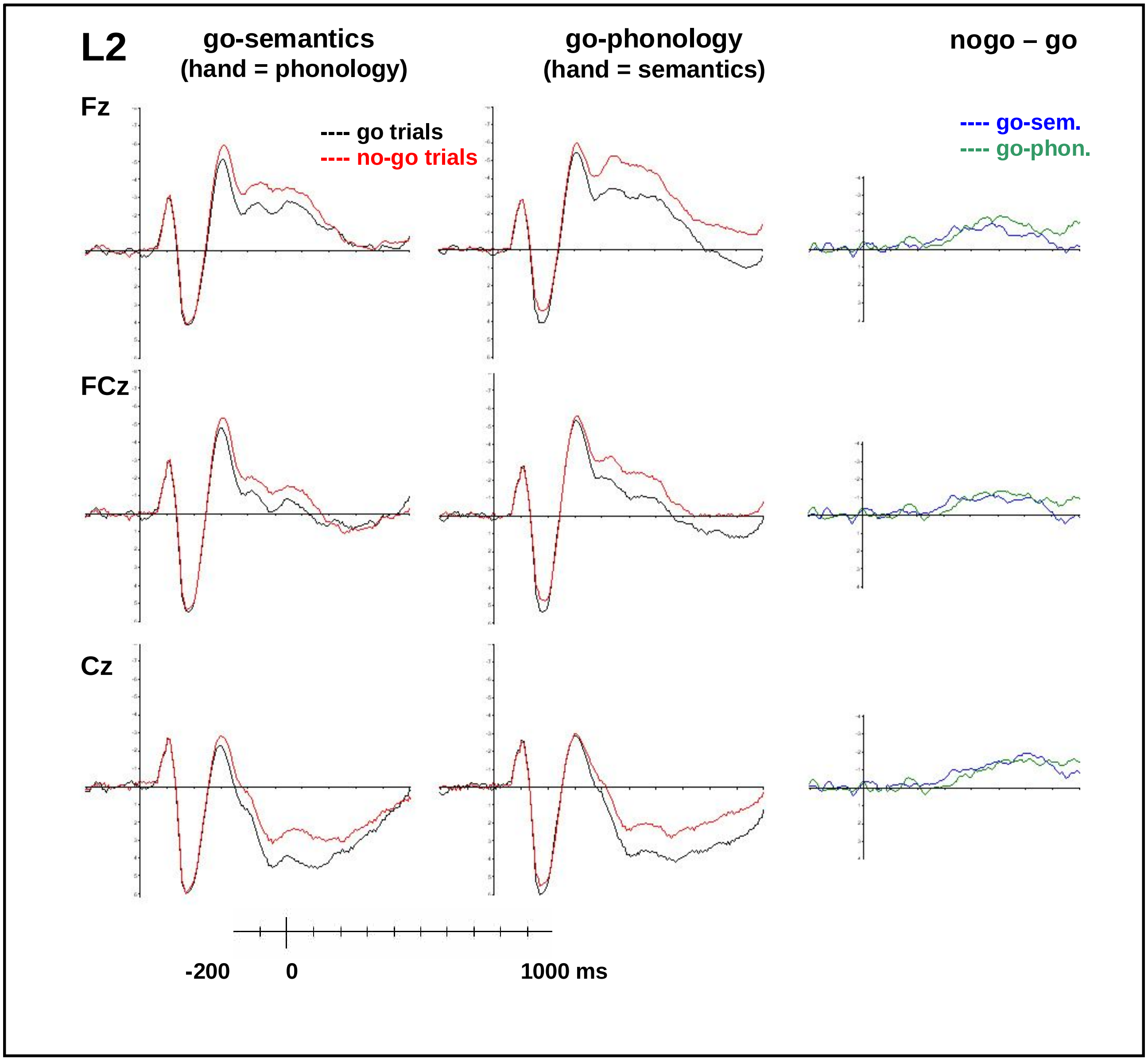


Figure 5B. *The English N200 grand averages at midline electrode sites. (Left column = "go-semantics" condition; middle column = "go-phonology" condition; right column = "no-go minus go" difference waves.) Negativity is plotted upwards. Voltage ranges +6/-9 µV for N200; +4/-4 µV for difference waves. N=18*

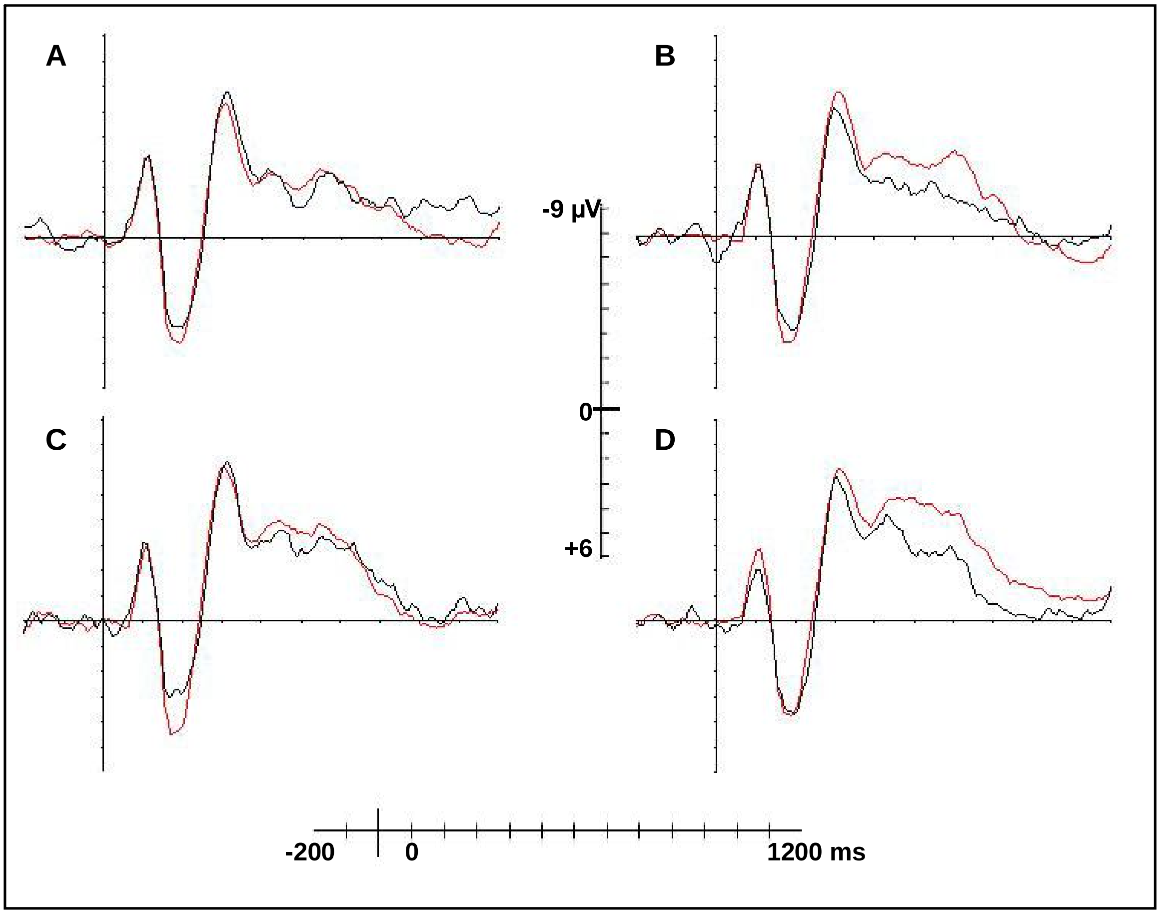
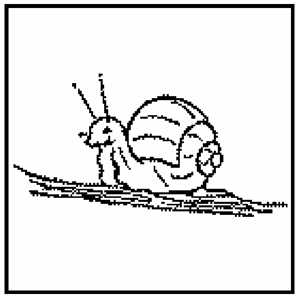


Figure 6. *The go-waveforms for the coincidence (**red**) and noncoincidence (**black**) stimuli in the go-semantics (A) and go-phonology (B) conditions. Graphs C and D show the no-go coincidence and non-coincidence stimuli in the go-semantics (C) and go-phonology (D) conditions. Negativity is plotted upwards.*



#	**instruction**	**response**
1	Press the right button if depicted object starts with the sound „S“ and left button if it does not start with the sound „S“. Press only for manmade objects.	--------
2	Press the right button if depicted object does not start with the sound „S“ and left button if it starts with the sound „S“. Press only for manmade objects.	--------
3	Press the right button if depicted object does not start with the sound „S“ and left button if it starts with the sound „S“. Press only for natural objects.	LEFT
4	Press the right button if depicted object starts with the sound „S“ and left button if it does not start with the sound „S“. Press only for natural objects.	RIGHT
5	Press the right button if the depicted object is natural and left button if it is manmade. Press only if its name starts with the sound „S“.	RIGHT
6	Press the right button if the depicted object is manmade and left button if it is natural. Press only if its name starts with the sound „S“.	LEFT
7	Press the right button if the depicted object is manmade and left button if it is natural. Press only if its name does not start with the sound „S“.	--------
8	Press the right button if the depicted object is natural and left button if it is manmade. Press only if its name does not start with the sound „S“.	--------

Appendix 1. *Example of all 8 instructions in one language together with model responses for the target item “snail“. Items 1 – 4 belong to the “go-semantics“ and item 5 – 8 to the “go-phonology“ version. Stimulus list “A“. The order of instruction presentation for the other 3 stimuli lists is as follows: B (8,7,6,5,4,3,2,1), C (3,4,1,2,7,8,5,6) and D (6,5,8,7,2,1,4,3).*

nonshared onsets		*shared onsets*	
man-made	**natural**	**natural**	**man-made**
<u>sjaal (scarf)</u>	ober (waiter)	pauw (peacock)	potlood (pencil)
raam (window)	vleugel (wing)	pinda (peanut)	kaars (candle)
horloge (watch)	<u>aardbei (strawberry)</u>	papegaai (parrot)	schaar (scissors)
<u>rok (skirt)</u>	<u>vleermuis (bat)</u>	slang (snake)	schroef (screw)
<u>lepel (spoon)</u>	<u>vlinder (butterfly)</u>	regen (rain)	schommel (swing)
<u>emmer (bucket)</u>	<u>varken (pig)</u>	slak (snail)	stuur (steering wheel)
<u>fles (bottle)</u>	<u>ananas (pineapple)</u>	spin (spider)	tandenborstel (toothbrush)
<u>riem (belt)</u>	**kreeft (lobster)**	traan (tear)	beha (bra)
<u>knoop (button)</u>	hond (dog)	teen (toe)	bezem (broom)
dak (roof)	oog (eye)	bot (bone)	bank (bench)
<u>kussen (pillow)</u>	**kikker (frog)**	bruid (bride)	brug (bridge)
<u>doos (box)</u>	**bloem (flower)**	koe (cow)	kooi (cage)
<u>schilderij (painting)</u>	meisje (girl)	koning (king)	kopje (cup)
sleutel (key)	**spook (ghost)**	ei (egg)	kam (comb)
brief (letter)	**boom (tree)**	draak (dragon)	pakje (parcel)
bureau (desk)	aap (monkey)	soldaat (soldier)	kruis (cross)
blik (can)	wortel (carrot)	regenboog (rainbow)	kroon (crown)
ketting (chain)	paard (horse)	naald (nail)	leeuw (lion)
stoel (chair)	**blad (leaf)**	lippenstift (lipstick)	walvis (whale)

Appendix 2. *Stimuli. The two left columns contain the originally non- coincidenctal stimuli (with initial phonemes overlapping in the two languages); the two right columns contain the originally coincidental stimuli. In bold are those noncoincidental stimuli which were included in Dutch the "coincidence" analysis, underlined are those used in the English coincidence analysis. Nore that some stimuli were used in both Dutch and English coincidence analysis (see Results).*