CHAPTER 9

Is bilingual speech production language-specific or non-specific?

The case of gender congruency in Dutch-English bilinguals

Niels O. Schiller and Rinus G. Verdonschot
Leiden University Centre for Linguistics / Leiden Institute for Brain and Cognition, The Netherlands

The present paper looks at semantic interference and gender congruency effects during bilingual picture-word naming. According to Costa, Miozzo & Caramazza (1999), only the activation from lexical nodes within a language is considered during lexical selection. If this is accurate, these findings should uphold with respect to semantic and gender/determiner effects even though the distractors are in another language. In the present study three effects were found, (1) a main effect of language, (2) semantic effects for both target language and non-target language distractors, and (3) gender congruency effects for targets with target-language distractors only. These findings are at odds with the language-specific proposal of Costa et al. (1999). Implications of these findings are discussed.

Keywords: Language Production, Psycholinguistics, Gender congruency, bilingual language control

1. Introduction

Mastering other languages besides one's mother tongue is a precious ability to have. People who can switch effortlessly between languages not only enjoy the awe of those who cannot, but also reap more benefits in terms of opportunities in life. Nowadays, we live in a world in which multilingual communication is used virtually everywhere. With the invention and implementation of new technologies, like mobile telecommunication and internet, it is becoming easier to get in touch with other people and media from different nationalities and backgrounds. Thus, it is often considered beneficial to be able to speak and understand more than one language.

DOI 10.1075/z.210.09sch

Although speaking in, for example, one's own mother tongue does not seem difficult to most of us, it is in fact a very complex activity. We must decide what we want to say, when and how to say it. We retrieve some two or three words per second from a lexicon that contains hundreds of thousands of items. We have to put these words in the correct order according to the appropriate syntax. Then we must transform these words into actual speech sounds by sending motor commands to our muscles and take correct gestures into account. All of this has to be done in a very short time and presented in a fluent manner. There is still much to be investigated to be able to unravel all the components of this wonderful ability.

Most research on speech production has been approached either from speech error data (such as: saying "shake a tower" instead of "take a shower") or from chronometric measurements (reaction times) in word production experiments (see Levelt, Roelofs & Meyer 1999 for a review). Most theories agree that there are at least two distinguishable processing stages in lexical access in language production. The first stage comprises the selection of a conceptually/semantically defined lexical representation and the second stage constitutes the selection of its phonological representation (Caramazza 1997, Dell 1986, Levelt 2001).

One of the most influential theories on speech production is described in Levelt, Roelofs and Meyer (1999; see also Figure 1). Here, the production of words is considered to be a staged process. The first step is *conceptual* preparation. This means the process leading up to the selection of a lexical concept, or, in other words, concepts for which there is an item in the mental lexicon. This selection depends on pragmatic and context-dependent decisions (Levelt et al. 1999). The model also assumes that during selection there is co-activation of related concepts. Each active lexical concept spreads activation to the corresponding lexical-syntactic representation (or lemma) in the speaker's mental lexicon (Levelt 2001). Syntactic features will be processed in this stage such as the distinction between mass and count nouns. Also, in gender-based languages such as Dutch, at this stage the syntactic gender would be processed (contrary to, for example, English, which has no syntactic gender).

Selection latency for the target lemma depends on the amount of co-activation of other lemmas. Lexical selection is complete as soon as the target lemma is selected (Levelt 2001).

There is a debate on whether concepts are complete/holistic semantic concepts like a "dog"-node (indivisible node), or rather, parts which constitute a whole concept (such as: it has paws, a tail, fur and it barks) and combined will evoke a semantic representation of a dog. The main difference between non-divisionalists (Levelt) and divisionalists (Caramazza) is whether the activation of a given concept would activate part of the semantic representation of other related concepts because some of their semantic features are shared (Costa, Colomé & Caramazza 2000).

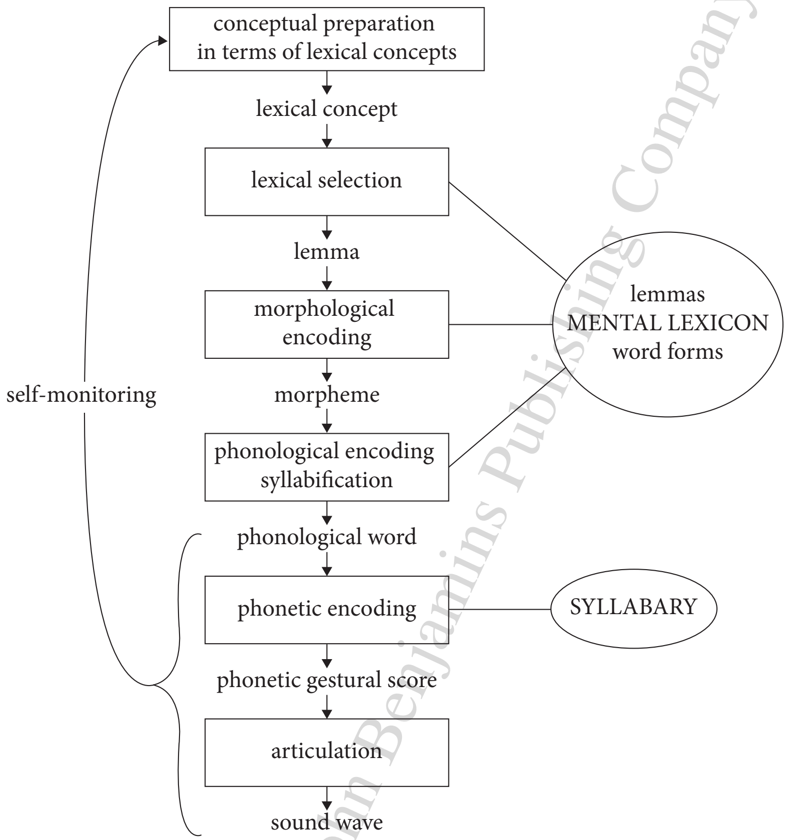

Figure 1. Levelt, Roelofs & Meyer's model of speech production (Levelt et al. 1999: 3)

After selection of the lemma, its syntactic properties are accessible for further processing. The construction of phrases, clauses, and whole sentences, depends on this syntactic information. If one were to produce an adjective + noun phrase in a gender-marked language, this type of information would be passed on with the lemma. In a gender-marked language like Dutch, the adjective inflects when the noun is neuter and becomes plural ("GROOT PAARD" $_{\text{singular}}$ [big horse] vs. "GROTE PAARDEN" $_{\text{plural}}$ [big horses]. In a genderless language such as English, however, the adjective will remain the same. Many verbs have also diacritic features for number and tense. It is critical that these values are respected in further encoding.

After selection of the lemma the next step is to continue to the phonological/articulatory domain. Each lemma node is connected to a lexical-phonological (or lexeme) node, which specifies the morpho-phonological structure of a word (Levelt 2001). Subsequently, the phonological word form will be created. This comprises activating the appropriate basic phonological units (e.g. phonemes in Dutch and English), and the metrical frame (i.e. the number of syllables and the stress pattern) and producing the segmental spell out (incremental). Once the phonological word is established, phonetic encoding takes place. In the phonetic encoding phase, gestural scores will be computed, i.e. phonetic encoding controls articulatory systems to invoke gestural intentions. Such control manipulates (among other systems) nasal, glottal and oral muscle systems and is a context-dependent process. The final process of the model in producing an utterance is called "articulation". This is the control and actual execution of the prepared gestural scores thus exerting them on the vocal cords and both the oral and nasal systems (Levelt & Indefrey 2000).

An interesting property of the language system concerns the dependencies between words. One such dependency, at least in Dutch, is the gender/determiner dependency. If a person wants to produce a noun phrase such as "DE KOE" [the cow], the gender of the word (in this case common gender) needs to be known before the appropriate determiner can be selected.

To investigate this process in more detail, Schriefers (1993) conducted an experiment, which looked into syntactic and lexical-semantic processes during the production of Dutch noun phrases (NPs). He found that gender incongruent distractors (such as "DE KOE" $_{\text{common}}$ [the cow] and "HET PAARD" $_{\text{neuter}}$ [the horse]) delayed the production of a noun phrase compared to gender congruent distractors. He interpreted these results in terms of the model proposed by Levelt (1989) (cf. also Levelt et al. 1999). As can be seen in Figure 2, the target (picture) will activate a concept and a lexical representation of the noun, for instance, "HUIS" [house]. This lexical representation will spread to its corresponding gender node. The model assumes that there is one universal gender node for each gender. Also, all lexical nodes are connected to one corresponding gender node. For instance, in Dutch it would be the case that all neuter-gender words share the same neuter gender node, and all common-gender words share the same common gender node.

Imagine that a possible distractor like "KERK" [church] activates its gender node and when, like in the case ("HET HUIS" $_{\text{neuter}}$ vs. "DE KERK" $_{\text{common}}$) the target and distractor nouns differ in gender (gender incongruent condition), these two activated gender nodes will compete for selection and delay selection of the correct node and naming time as a consequence (Schriefers 1993, Schriefers & Teruel 2000).

The gender/determiner congruency effect has been demonstrated for Dutch (Schiller & Caramazza 2003, Schriefers 1993) and German (Schiller & Caramazza 2003, Schriefers & Teruel 2000). Costa, Miozzo & Caramazza (1999) pointed out, however, that there are other possibilities to explain this effect. In Dutch, the grammatical gender and number dictate the appropriate determiner and adjective inflection (early selection language). So, it is difficult to pinpoint whether the

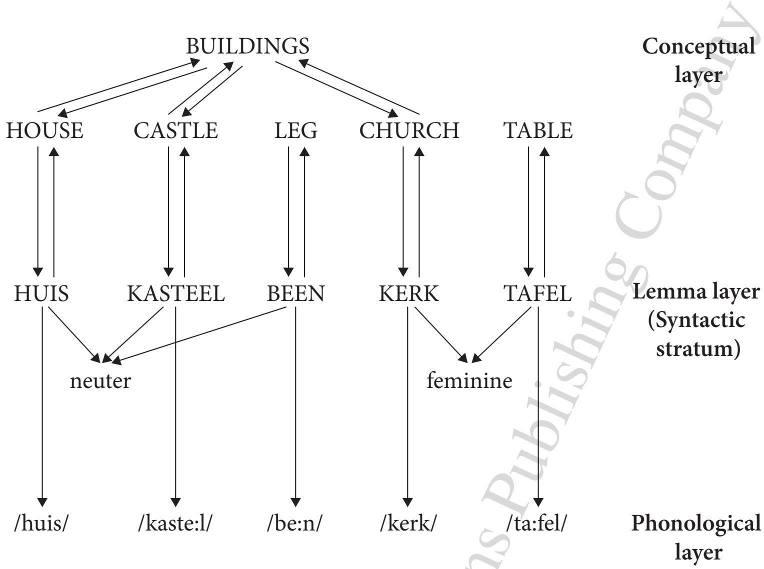

Figure 2. Speech production model (taken from Schriefers 1993: 842).

gender interference effect is caused by gender features or by competition between specific determiners ("DE"$_{\text{common}}$ vs. "HET"$_{\text{neuter}}$). However, evidence that the determiner might elicit this effect (and not gender) comes from research by Schiller and Caramazza (2003). They devised a paradigm, which distinguishes between singular and plural noun phrase production in German and Dutch. In the singular conditions, the to-be-named determiner was being specified by the gender, but in the plural condition the determiner was the same in every condition (irrespective of the word's gender). They found that the gender congruency effect only arises in the singular condition while the gender also differs in the plural condition; hence they called it a "determiner congruency effect" (Schiller & Caramazza 2003).

In Romance languages, like Italian and French (late selection languages), the determiner and adjective inflection are not solely determined by the grammatical gender. The selection also depends on the specific phonological form of the noun and the adjective to determine its definitive state. This means that in some cases the phonological information of the word following the determiner (i.e. the word onset) is needed for determiner selection (Miozzo & Caramazza 1999). In Italian the determiner for masculine nouns is *il* except when the next word starts with *s+-consonant*, *gn*, or a vowel, then it is *lo*. For example: *il treno* 'the train' but *lo sgabello* 'the stool'. In Italian, determiner selection is faced with conflict between the available information about the determiner. In 1999, Miozzo and Caramazza addressed the question whether the gender congruency results obtained by Schriefers (1993) could also be found in Italian. They failed to replicate Schriefers' results; however, an interesting phenomenon arose in one of their variants. In the last variant of their experiment participants were instructed to produce noun phrases with the word *grande* 'big' in between. For the masculine singular *lo* condition, the determiner changed from *lo* to *il* (for instance: *lo sgabello* becomes *il grande sgabello*) whereas in the *il* condition nothing changed (determiner stayed the same). The crucial finding was that only when the determiner changed under the influence of local phonological properties (e.g. from *lo* to *il*), the NPs were produced more slowly (Miozzo and Caramazza 1999). Miozzo and Caramazza proposed that the noun activates multiple determiners (for instance: *lo* and *il* for singular masculine) but that selection of a specific determiner will have to wait for the ordering and insertion of the noun and adjectives into a phonological phrase. In Dutch and German this is not the case. In these languages the correct determiner can be selected immediately. This between-language discrepancy provides a basis for a possible explanation of the contrasting results between Dutch/German and Italian.

As stated before, most models of language production assume there is competition between items during lexical selection. Most experiments have used the picture-word interference paradigm to look into the nature of this effect. In this type of experiment, participants are presented with line drawings of common objects (targets) and written words (distractors) printed on top of these pictures. Participants are instructed to name the object as quickly as possible while ignoring the distractor word. Naming latencies vary as a function of the relation between target picture and distractor word. Imagine the case where a participant is presented with a picture of a *pig* and the semantically related word *cow* is superimposed on the picture. Compared to a word, which is not semantically related, the word *cow* receives extra activation from the picture of the *pig* and as a result creates more competition than semantically unrelated words. As a consequence, naming times are delayed. This is called the "semantic interference effect" (Glaser & Düngelhoff 1984; but see Mahon et al. 2007 for an alternative account).

What do current theories have to say about semantic interference effects in bilingual language production? There are currently two main assumptions (Costa, Kovacic, Franck & Caramazza 2003), these are: (1) the two languages of a bilingual share the same conceptual system (Costa et al. 2000) and (2) this conceptual system activates corresponding lexical nodes in both languages. This means that the semantic system is connected to each representation in the corresponding language. The question is whether the language, which is not programmed for response, also receives activation and what kind of influence this activation will exert on the response language.

Recent theories assume that the activation spreads to the two systems whether they are programmed for response or not (Costa et al. 2000). How do bilinguals separate these languages so they will not interfere with each other? Two theories have emerged to provide an answer to this question. The first theory proposes that lexical selection is "language non-specific" (also called Inhibitory Control Model) (De Bot 1992, Green 1998). This means that all the nodes in all languages are taken into account, and that an inhibitory mechanism is necessary to suppress the undesired language. This suppression is proportional to the levels of activation of the non-target language (meaning highly activated nodes in the non-target language should require more suppression than lesser activated nodes). One could be tempted to say this is actually a language-specific model but this is incorrect because the activated nodes in the non-target language if not inhibited would cause an interference effect on the target language (they are all considered/non-specific, and therefore have to be inhibited). A contrasting approach by Costa and colleagues states that selection is language-specific (Costa et al 1999). This means that although both languages receive activation from the semantic representation, only one language is considered without requiring a suppression mechanism. The contrast between the language-specific model and the Inhibitory Control/non-specific model is that in the language-specific model only the nodes of the target language are considered and the non-target language is assumed not to exert an influence on the target language.

These theories make different predictions on a special case of the picture-word interference paradigm, namely with *identity* distractors. These are words, which are direct translations of a picture label in the language not to be named. For instance, if a picture of a horse has to be named in L1, Dutch, "PAARD", and an L2 distractor, "HORSE", is printed on top of it, the language non-specific selection hypothesis predicts longer naming latencies than unrelated words, like "CAR", because there is competition across languages and suppression is needed for "HORSE" to be able to select "PAARD". The language-specific selection hypothesis predicts faster naming latencies because the "PAARD" node receives activation (by translation) from the identity word (i.e. "HORSE") and the picture itself, and is therefore faster than an unrelated distractor. Costa and colleagues used the picture-word paradigm with Spanish-Catalan distractor pairs. They found *facilitation* effects for both same-language and different-language identity distractors. This is in contrast with the language non-specific account which predicts that the suppression mechanism would cause a delay with different-language identity distractors. These effects were replicated in several experiments and in a variety of conditions, e.g. blocked-language (persons knows in which language the response will be) vs. mixed-language naming (person is unaware of response language). Costa and colleagues concluded that lexical selection in the target language is achieved by a selection mechanism that considers only the activation of the lexical nodes, which belong to the target language without requiring inhibitory processes (Costa et al. 1999).

The present paper aims to look further into the semantic interference and gender congruency effects in bilingual picture-word naming. Most theories of bilingual lexical access assume that both lexicons are activated in parallel. For instance, an English distractor like "HORSE" will spread to its Dutch counterpart "PAARD". If Costa et al. (1999) are correct and only the activation from lexical nodes within a language is considered, it should be possible to replicate semantic and gender effects found in monolingual research though the distractors are in another language.

The current experiment modifies three variables: language, semantic relatedness, and gender/determiner congruency. According to most theories it is expected that semantically related distractors (e.g. target "TAFEL" [table] and distractor "STOEL" [chair]) lead to longer naming latencies than unrelated distractors (see Glaser & Düngelhoff 1984). We expect that this should hold when distractors are presented in L2, e.g. "TAFEL" and "CHAIR", because, if Costa and co-workers are correct, only one language is selected. Furthermore, the gender/determiner incongruency effect should also be found if only Dutch (L1) lexical nodes are considered. We used English as the L2, for, if a gender effect indeed is found with the (L2) distractors being in a language, which does not have syntactic gender, this is likely to be caused by L1 (gender-marked) distractor words. Effects with L2 distractors should preferably be of comparable magnitude to L1 distractors because one assumes there is only consideration of L1. This requires, however, exact translations, which enable the same L1 lexical nodes as in the monolingual condition. To summarize, we assume that both Dutch and English distractors will elicit semantic and gender/determiner congruency effects and that these effects are of comparable magnitude regardless of the language the distractor is in.

2. Experiment – semantic interference and gender congruency effects in a bilingual picture-word interference paradigm

2.1 Methods

2.1.1 *Participants*

Twenty-eight bilingual Dutch-English students from Maastricht University and the Maastricht Interpreter Academy took part in this study. Participants were between 19 and 30 years old (mean age: 24 ± 3 years). Participants gave informed consent and received payment of €5,00 (gift-certificate) for their participation in the experiment. Two participants were excluded from the analysis because their Dutch was not native-like presumably because they were not born in the Netherlands. All participants had normal or corrected-to-normal vision. Table 1 lists the self-assessment scores of the included participants.

Table 1. Description of self-assessment English (L2) language skills and history (production to reading are self-assessed markings from one to ten, with ten being the highest mark).

	Onset of L2 acquisition	Using L2 in years	Production	Comprehension	Writing	Reading
Average	9.68	7.36	6.84	7.72	6.72	7.64
SD	4.34	5.12	1.03	0.60	1.07	0.72

2.1.2 *Materials*

The experiment was divided into 2 blocks. In both blocks, participants were required to produce determiner-noun phrases in Dutch (L1), but the distractors were either in L1 (Dutch block) or L2 (English block). Pictures to be named were taken from the picture database from the Max Planck Institute for Psycholinguistics in Nijmegen. Special attention was paid to avoid the use of ambiguous words, cognates, phonologically related distractors, large frequency differences (taken from the CELEX database; Baayen, Piepenbrock & Gulikers 1995) and a non-coordinate relationship. Distractor words were translated by means of a standard Dutch/English dictionary (Van Dale – *Groot woordenboek Nederlands-Engels*. Utrecht, Van Dale Lexicografie Bv. 1999). Before the start of the present experiment a pilot experiment (13 participants) was conducted in order to establish that these stimuli would elicit a (monolingual) effect and were adequate for the present experiment. Please see Appendix A for an overview of the stimuli used.

2.1.3 *Design*

A 2×2 (semantic vs. gender) design was implemented such that each target had two semantically related distractors and two semantically unrelated distractors. The semantically related distractors were designed to be the semantically unrelated distractors from another target (and vice versa). Figure 3 shows an example of a target picture with its corresponding distractor words. The order between blocks was counterbalanced across participants. Half of the participants was presented with the Dutch block first and vice versa. Sequence of targets and distractors was randomized within every block and for every participant.

First, all stimuli were shown to participants with the correct names written below them (without determiners). This was done in order to maximize name agreement among all participants. Some people say, for instance, "BOOT"$_{\text{common}}$ [boat] where others say "SCHIP"$_{\text{neuter}}$ [ship]. Name agreement is important because a gender/determiner might differ if name agreement is not 100% (see the above-mentioned example). Subsequently, participants were shown all pictures a second

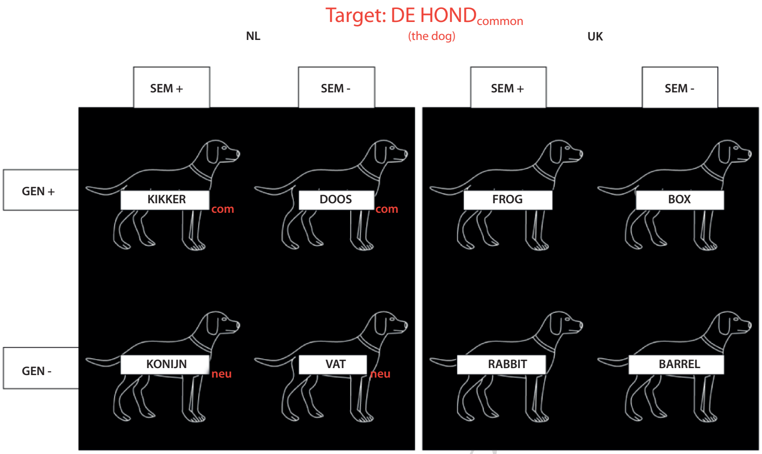

Figure 3. Example of a target picture and its corresponding distractor words in both languages (com/neu = common/neuter gender; gender information is not shown in the actual experiment).

time but this time without the names written underneath. Participants were asked to name the picture plus the correct determiner. If any mistakes occurred during this period, participants were told what the mistakes were and given a chance to redo this phase. After correctly naming the pictures, participants started the final phase. They were requested to name the pictures in Dutch with the distractors written on top of them. The first two items of each series were practice trials. Reaction times were recorded using a voice-key. The final phase consisted of two blocks (Dutch and English) and each block consisted of four repetitions. In each repetition another condition of a target was presented and the order of stimuli within each block was randomized using a Latin square design.

2.1.4 *Procedure*

Participants were tested individually while seated in a soundproof booth in front of a computer screen. Distance from the computer screen was approximately 90 cm. A voice-key consisting of a Behringer microphone connected to a Soundblaster AWE 64 sound card was used to record the reaction times. The software package ERTS (BeriSoft Cooperation 1999) was used to present the stimuli. Participants were subjected to a total number of 72 trials per block. These trials consisted of 18 visually presented target pictures with four distractors each (superimposed on the picture). This resulted in a total amount of 18 targets × 4 conditions × 2 blocks (Dutch or English) = 144 trials and lasted approximately 22 minutes.

A trial began with the presentation of a fixation cross in the middle of the screen. The fixation cross was replaced by a picture and distractor (SOA = 0 ms) after a variable delay of between 300 and 750 ms. After each recording there was a delay of 1,000 ms before the next trial started. Targets and distractors were white pictures and words presented on a black background. On all trials, participants were instructed to respond by naming the picture as fast and accurate as possible. Participants were requested to name pictures in Dutch (L1) by producing determiner-noun phrases such as: "DE KAT" $_{\text{common}}$ [the cat]. Distractors were related to the picture semantically (related vs. unrelated) and were either gender congruent or incongruent. Participants were instructed to be relatively quiet because undesired noise could trigger the voice key. Also, they were instructed to speak at the same volume level throughout the whole experiment and to maintain a clear pronunciation (e.g. they should not say "'T" $_{\text{neuter}}$ but rather "HET".)

Participants were asked at the end of the experiment to judge their L2 language history skills. Also, after the experiment, a list of all the English distractors was presented and participants had to check a box denoting whether they (1) knew and could translate the word, (2) knew, but could not translate the word, or (3) did not know the word.

2.2 Results

Errors were recorded as (1) gender error (for example: saying "DE" $_{\text{common}}$ instead of "HET" $_{\text{neuter}}$) or (2) voice key error (i.e. technical errors with the recording system). Reaction times (RTs) below 350 ms or above 1,500 ms were discarded. For the Dutch block a total amount of 3.3% among all participants was discarded (62 trials) leaving 96.7% to be analyzed. For the English block 2.6% of the data was rejected leaving 97.4%. From the list participants were required to fill in regarding knowledge of the L2 distractor words it was determined per participant per word whether or not the translation was known. For 13.3% (250 items) measured across all participants this was not the case and these data were discarded leaving a total number of 86.7% to be analyzed (see Appendix B for an overview of the unknown words).

First of all, a three-way analysis (language × semantics × gender) was conducted to investigate whether the presence of a particular pattern for one variable depended on another experimental variable (e.g. whether semantic or gender effects would only be found in Dutch or not). None of the interactions were significant except for the language × gender interaction in the subject analysis, $F1(1, 25) = 6.85$, $p < .05$, MSE = 886; $F2(1, 17) = 3.56$, $p = .076$. This interaction indicates that in order to investigate the gender congruency effect, separate analyses per language block are necessary (as the distractor language influences the gender congruency effect in some way). Second, an overall effect of language was found (22 ms), $F1(1, 25) = 7.49$, $p < .05$, MSE = 3323, $F2(1, 17) = 6.80$, $p < .05$, MSE = 9355, indicating that participants were overall faster when L2 (English) distractors were printed on the pictures. Third, an overall effect of semantic relatedness (14 ms) was found in the subject analysis, $F1(1, 25) = 10.55$, $p < .01$, MSE = 965; $F2(1, 17) < 1$, indicating that semantically related distractors slowed down naming latencies irrespective of language and gender congruency.

To investigate the effect of gender congruency (because of the interaction between language and gender) two separate analyses were conducted (i.e. one per language). When using Dutch distractors, a significant difference in the subject analysis was obtained for gender (13 ms) in the subject analysis, $F1(1, 25) = 4.65$; $p < .05$, MSE = 909; $F2(1, 17) = 3.31$, $p = .083$. However, when using English distractors no significant difference was obtained for gender congruency, $F1(1, 25) = 2.40$, $p = .134$; $F2(1, 17) < 1$.

Table 2. Mean naming latencies (in ms) and standard errors (in parentheses) per block.

Block	Semantic relatedness	Gender congruent	Gender incongruent
Dutch	related	779 (16.7)	796 (14.9)
	unrelated	766 (14.7)	775 (18.2)
English	related	768 (17.0)	757 (14.2)
	unrelated	755 (16.6)	749 (14.5)

3. Discussion

This study addressed the semantic interference and gender/determiner congruency effect in bilingual picture word naming. Three effects were found, (1) a main effect of language, (2) semantic effects for both target-language and different-language distractors, and (3) gender congruency effects for targets with target-language distractors but not for targets with different-language distractors.

The main effect of language shows an interesting pattern. Participants took between 11 to 40 ms (mean: 22 ms) longer to name targets with L1-distractors than targets with L2-distractors. This is at odds with the language-specific proposal of Costa and Caramazza (1999). In their view, target and non-target distractors should typically evoke an equal amount of interference. According to Finkbeiner, Gollan & Caramazza (2006), a possible solution this difference in magnitude between blocks could be that written distractors activate their corresponding nodes directly. This means their activation is not mediated by the semantic system and as a consequence same language distractors would produce more activation in the target lexicon than do different language distractors. This solution predicts an interaction between the semantic interference effect and language (Finkbeiner, Gollan & Caramazza 2006). Because the interaction of language by semantics was not significant in the present experiment, it was statistically not permitted to explore this difference any further, although it can be seen from Table 2 that the semantic effects differ in magnitude per block (18 ms for Dutch and 10 ms for English).

Problematic for the Costa and Caramazza (1999) proposal that the non-target language is not considered when distractor words were in English is the absence of a gender effect. In their view, a cross-language and within-language semantic interference effect is in fact the same effect (Costa et al. 1999, Finkbeiner et al. 2006). It would naturally follow from this statement that the observed gender/determiner effect in the Dutch block would also be present in the English block as they are considered to reflect the same process. But this is not in line with the present data. Our findings could indicate that translation is absent and perhaps the semantic effects are occurring at a different level.

A different possibility to explain the absence of the gender/determiner effect in the English block would be that there is a single integrated syntactic gender system for the two languages (the so-called "Language Integrated View"). In this case, it could be that because the English distractors have no syntactic gender no competition will arise because either there is no node present or that a present node is not in conflict with the target gender/determiner node. However, evidence in favor of the view that different languages have different gender systems (the so-called "Language Autonomy View") comes from research conducted by Costa, Kovacic, Franck and Caramazza (2003). In their experiments, highly proficient bilinguals (Croatian/Italian and Italian/French) were asked to produce noun-phrases in a picture-word interference paradigm. Their reasoning was that if gender-systems of a bilingual are somehow shared, one should be able to observe faster reaction times for same-gender translation words (such as "MELA"$_{\text{Italian/feminine}}$ vs. "JABUKA" $_{\text{Croatian/feminine}}$ [both meaning apple]) compared to different gender translation words ("POMODORO" $_{\text{Italian/masculine}}$ vs. "RAJĈICA" $_{\text{Croatian/feminine}}$ [both meaning tomato]). Their main result was in fact a null effect, meaning that there was no difference between conditions for different/same gender features. This led Costa et al. (2003) to the conclusion that whether L1 and L2 have the same or different grammatical features is irrelevant for the grammatical organization of both languages (as they are independent).

Finkbeiner and colleagues have proposed an interesting alternative way of thinking which perhaps in future experiments could explain effects found in the present research. Their reasoning is that one of the assumptions underlying most (bilingual) speech production models could be incorrect, namely the assumption that lexical selection is *competitive*. Suppose this were not the case, and a simple "selection-by-threshold" mechanism would select the first node, which reaches threshold. The speaker would determine by means of intention to speak in a specific language which language receives more activation. Then, it would simply be a matter of increasing activation to the target-language nodes relative to the non-target language node (Finkbeiner et al. 2006).

In short, our results are problematic for the language non-specific proposal of Costa and colleagues (1999). More experiments between other languages need to be conducted to provide better insights into the mechanisms of the bilingual language production system.

References

Baayen, R. H., R. Piepenbrock & L. Gulikers. 1995. The CELEX lexical database (CD-ROM). Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

Caramazza, A. 1997. How many levels of processing are there in lexical access? *Cognitive Neuropsychology* 14. 177–208. doi: 10.1080/026432997381664

Costa, A. & A. Caramazza. 1999. Is lexical selection in bilingual speech production language-specific? Further evidence from Spanish-English and English-Spanish bilinguals. *Bilingualism: Language & Cognition* 2. 231–244. doi: 10.1017/S1366728999000334

Costa, A., M. Miozzo & A. Caramazza. 1999. Lexical selection in bilinguals: Do words in the bilingual's two lexicons compete for selection? *Journal of Memory and Language* 41. 365–397. doi: 10.1006/jmla.1999.2651

Costa, A., A. Colomé, A. Caramazza. 2000. Lexical access in speech production: The bilingual case. *Psicológica* 21. 403–437.

Costa, A., D. Kovacic, J. Franck, A. Caramazza. 2003. On the autonomy of the grammatical gender systems of the two languages of a bilingual. *Bilingualism: Language and Cognition* 6. 181–200. doi: 10.1017/S1366728903001123

Costa, A., F.-X. Alario & A. Caramazza. 2005. On the categorical nature of the semantic interference effect in the picture–word interference paradigm. *Psychonomic Bulletin and Review* 12. 125–131. doi: 10.3758/BF03196357

De Bot, K. 1992. A bilingual processing model: Levelt's 'Speaking' model adapted. *Applied Linguistics* 13. 1–24.

Dell, G. 1986. A spreading activation theory of retrieval in sentence production. *Psychological Review* 93. 283–381. doi: 10.1037/0033-295X.93.3.283

Finkbeiner, M., T. Gollan & A. Caramazza. 2006. Lexical access in bilingual speakers: What's the (hard) problem? *Bilingualism: Language and Cognition* 9. 153–166. doi: 10.1017/S1366728906002501

Glaser, W. R. & F. J. Düngelhoff. 1984. The time course of picture word interference. *Journal of Experimental Psychology: Human Perception and Performance* 10. 640–654.

Green, D. W. 1998. Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition* 1. 67–81. doi: 10.1017/S1366728998000133

Levelt, W. J. M. 1989. *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Levelt, W. J. M., A. Roelofs & A. S. Meyer. 1999. A theory of lexical access in speech production. *Behavioral and Brain Sciences* 22. 1–75. doi: 10.1017/S0140525X99001776

Levelt, W. J. M. 2001. Spoken word production: A theory of lexical access. *Proceedings of the National Academy of Sciences* 98. 13464–13471. doi: 10.1073/pnas.231459498

Levelt, W. J. M. & P. Indefrey. 2000. The speaking mind/brain: where do spoken words come from? In R. Port (ed.), *Image, Language, Brain: Papers from the First Mind Articulation Project Symposium*, 77–93. Cambridge, MA: MIT Press.
Mahon, B. Z., A. Costa, R. Peterson, K. A. Vargas & A. Caramazza. 2007. Lexical selection is not by competition: a reinterpretation of semantic interference and facilitation effects in the picture-word interference paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 33(3). 503.
Schiller, N. O. & A. Caramazza. 2003. Grammatical feature selection in noun phrase production: Evidence from German and Dutch. *Journal of Memory and Language* 48. 169–194. doi: 10.1016/S0749-596X(02)00508-9
Schriefers, H. 1993. Syntactic processes in the production of noun phases. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 19. 841–850.
Schriefers, H. & E. Teruel. 2000. Grammatical gender in noun phrase production: The gender interference effect in German. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 26. 1368–1377.

Appendices

Appendix A. Overview of Stimuli used

Pair	Target Picture	Semantically related Gender of Distractor Word	
		common	**neuter**
1	Geweer (rifle) $_{\text{neuter}}$	Bijl (axe)	Zwaard (sword)
	Ekster (magpie) $_{\text{common}}$	Mus (sparrow)	Varken (pig)
	Hond (dog) $_{\text{common}}$	Kikker (frog)	Konijn (rabbit)
2	Mand (basket) $_{\text{common}}$	Doos (box)	Vat (barrel)
3	Spek (bacon) $_{\text{neuter}}$	Karbonade (cutlet)	Gehakt (mince)
	Dak (roof) $_{\text{neuter}}$	Muur (wall)	Raam (window)
	Gebit (teeth) $_{\text{neuter}}$	Kaak (jaw)	Verhemelte (palate)
4	Broek (trousers) $_{\text{common}}$	Trui (sweater)	Hemd (shirt) $_{\text{neuter}}$
	Been (leg) $_{\text{neuter}}$	Wang (cheek)	Oog (eye)
5	Kaas (cheese) $_{\text{common}}$	Groente (vegetable)	Vlees (meat)
6	Jurk (dress) $_{\text{common}}$	Rok (skirt)	Pak (suit)
	Bos (forest) $_{\text{neuter}}$	Woestijn (desert)	Moeras (swamp)
7	Vest (waistcoast) $_{\text{neuter}}$	Das (tie)	Colbert (jacket)
	Vergiet (colander) $_{\text{neuter}}$	Rasp (grater)	Mes (knife)
8	Beitel (chisel) $_{\text{common}}$	Spijker (nail)	Houweel (mattock)
	Prei (leek) $_{\text{common}}$	Ui (onion)	Witlof (chicory)
	Stoel (chair) $_{\text{common}}$	Kist (chest)	Buro (desk)
9*	Paard (horse) $_{\text{neuter}}$	Eend (duck)	Hert (deer)

* A detailed example of word pair 9 is described below

Target Picture	Distractor			
	Distractor Semantically related		Semantically unrelated	
	Gender of Distractor		Gender of Distractor	
	common	neuter	common	neuter
Stoel (chair) $_{common}$	Kist (chest)	Buro (desk)	Eend (duck)	Hert (deer)
Paard (horse) $_{neuter}$	Eend (duck)	Hert (deer)	Kist (chest)	Buro (desk)

Appendix B. Frequency of words reported to be unknown in all subjects

Distractor	% of total distracter repetitions discarded
AXE	4
SPARROW	23
BARREL	19
MINCE	58
CUTLET	73
PALATE	58
CHEEK	4
SKIRT	4
SWAMP	8
GRATER	65
MATTOCK	100
ONION	4
CHICORY	58
DEER	4