Schwieter, J.W. *The Cambridge Handbook of Billingual Processing,*
Cambridge University Press 2015. ISBN 978-1-107-06058-6

18

Parallel language activation in bilinguals' word production and its modulating factors

A review and computer simulations

Annette M. B. de Groot and Peter A. Starreveld

18.1 Introduction

Witnessing fluent bilinguals holding a conversation can give the distinct impression that the language currently not in use does not interfere with the selected language, as if it is fully at rest, deactivated. Still, this common observation does not square with a substantial body of empirical evidence that suggests otherwise. In this chapter we will review this evidence. We will first show that during word recognition in bilinguals lexical elements in both sub-lexicons are activated in parallel, in other words, that bilingual word recognition involves language-non-selective lexical activation. From then we will focus on our main topic, bilingual word production, considering the results obtained with three versions of the picture-naming task: the picture–word interference task, simple picture naming, and phoneme monitoring. Meanwhile a number of variables that modulate the influence of activation in the other language will be identified: relative proficiency in the two languages, stimulus-set composition, stimulus repetition, sentence context, and sentence constraint.

A final section presents simple computer simulations of the main findings obtained in studies that used the simple picture-naming task. One of the simulated results is the cognate effect, the common marker of language-non-selective activation in these studies. Other simulated results are the reduction of the cognate effect with repeated picture naming and with increases in response-language proficiency, and the effect of language proficiency and sentence context on response-selection time. The most important message emerging from these simulations is that the absence of a measurable influence of the non-response language can still be fully compatible with the claim that the bilingual language system is profoundly language non-selective.

18.2 Recognizing single words

There is quite a bit of evidence to suggest that the presentation of a printed or spoken word to a bilingual results in the activation of representation units in both sub-lexicons. Most of this evidence has been gathered in studies in which the stimulus words were presented visually and in isolation, that is, not embedded in a larger linguistic unit such as a sentence or a paragraph. In the majority of these studies one of two special types of stimulus words were used, namely, stimulus words that share a resemblance with one particular word in the other language (interlexical homographs and cognates), or stimulus words that vary with respect to their numbers of neighbors (intralexical or interlexical). An interlexical homograph is a written word that has two totally different meanings in a bilingual's two languages. For instance, for a Dutch–English bilingual *brand* is an interlexical homograph (it means 'fire' in Dutch). A cognate is a word that not only shares meaning with its translation in the other language but also all, or a large part, of its form (e.g., for a French–English bilingual the two components of the translation pair *table–table* are cognates). Finally, intralexical neighbors are words from the same language that have totally different meanings but very similar, though not identical, word forms (e.g., English *band* and *bank*) whereas interlexical neighbors are words from two different languages that have different meanings but very similar forms (e.g., Dutch *rook*, 'smoke,' and English *book*).

18.2.1 Recognition of printed words

Earlier monolingual studies have shown that the time it takes speakers of just one language to recognize a printed word depends on the number of its neighbors, obviously all intralexical in this case (e.g., Andrews, 1989; Grainger, 1990). This finding suggests that the presentation of a word not only activates its own representation in the mental lexicon but also those of its neighbors and that all activated representations compete with one another during the recognition process. Bilingual studies have built on this finding by posing the question whether the set of activated representations includes interlexical neighbors as well, thus providing an indication that word recognition in bilinguals is language non-selective. These studies suggest that this is indeed the case (e.g., Bijeljac-Babic, Biardeau, & Grainger, 1997; Grainger & Dijkstra, 1992; van Heuven, Dijkstra, & Grainger, 1998).

Because, like interlexical neighbors, interlexical homographs and cognates share form between a bilingual's two languages, presenting the latter two types of words should also cause parallel activation in both sub-lexicons. To be able to infer the occurrence of parallel activation, recognition time for interlexical homographs is compared with the time it takes to recognize control words, that is, words that are not interlexical homographs but are matched to the latter on preferably all other stimulus aspects that are known to affect recognition time (e.g., length and frequency of use). Similarly, recognition times for cognates and matched control words are compared, the control stimuli in this case being non-cognates. Quite a few studies have shown an interlexical homograph effect, that is, a difference in recognition time between interlexical homographs and control words (e.g., de Groot, Delmaar, & Lupker, 2000; Dijkstra, Grainger, & van Heuven, 1999; Dijkstra, van Jaarsveld, & Ten Brinke, 1998; Jared & Szucs, 2002; von Studnitz & Green, 2002). This effect is attributed to co-activation of the homographs' representations in the non-target lexicon. Von Studnitz and Green furthermore found that the interlexical-homograph effect gets smaller over the course of the experiment, suggesting that task practice can modulate the effect.

Similarly, a number of studies have obtained a cognate effect, a difference in recognition time for cognates and non-cognates (e.g., van Hell & Dijkstra, 2002; Schwartz, Kroll, & Diaz, 2007). Interestingly, a cognate effect has also been obtained in a visual-word-recognition study that tested bilinguals speaking Hebrew and English, two languages that use completely different alphabets (Gollan, Forster, & Frost, 1997). The cognates used in this study were thus phonological cognates, translation equivalents that share sound between a bilingual's two languages. This finding suggests that bilingual visual-word recognition not only involves the parallel activation of orthographic representations in the two sub-lexicons, but that activated orthographic representations immediately send on their activation to phonological representations and that they do so in a language-non-selective way. Studies using other experimental techniques have provided converging evidence to support this claim (e.g., Duyck, 2005; Jared & Kroll, 2001; Nas, 1983; van Leerdam, Bosman, & de Groot, 2009). Accordingly, computational models of visual-word recognition in bilinguals (BIA+, Dijkstra & van Heuven, 2002; SOPHIA, van Heuven & Dijkstra, 2001, as summarized in Thomas & van Heuven, 2005) contain both orthographic and phonological memory units and assume activation to spread between these two types of units, both within and across languages.

18.2.2 Recognition of spoken words

That the bilingual word-recognition system operates in a language-non-selective manner is also supported by studies in which the recognition of spoken rather than written words was examined. Most of these have used the visual-world paradigm (e.g., Blumenfeld & Marian, 2007; Marian & Spivey, 2003a, 2003b; Weber & Cutler, 2004): Participants are presented with aural instructions to (mentally) carry out specific actions related to a display of objects on a computer screen while their eye movements are registered. Russian–English bilinguals may for instance hear the instruction *put the marker below the cross*. In addition to a marker, a couple of filler objects, and a cross sign in the middle of the display, the display may show an object, the *competitor*, with a name in Russian that shares phonology with the target object's English name (e.g., a stamp, called *marka* in Russian). The critical finding is that the competitor is looked at significantly more often than the filler objects (with names that do not resemble the target object's name). Earlier monolingual studies had already shown that participants more often look at objects with a name similar to the target object's name in the same language (e.g., to a candy when the target object is a candle) than to objects with completely different names (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). Bilingual participants show this same within-language effect (e.g., Marian & Spivey, 2003b). The combined results suggest that, just as written words, spoken words activate similarly sounding words in both the target- and the non-target language. However, while some of these studies have demonstrated the cross-language effect with both the native language (L1) and the second language (L2) as the target language (Marian & Spivey, 2003a; Spivey & Marian, 1999), in other studies it was only obtained when L2, the weaker language of the two, was the target language (Blumenfeld & Marian, 2007; Marian & Spivey, 2003b; Weber & Cutler, 2004). This suggests that language-non-selective bilingual word recognition might only hold for L2 or that it holds for both languages but that there are limits to *the effects of* language-non-selective bilingual word recognition.

18.3 Recognizing words in context

Unlike in the studies discussed above, during natural language comprehension words typically occur in meaningful sentences and discourse. Therefore, to properly evaluate the claim that bilingual word recognition involves the parallel activation of elements in both sub-lexicons, the above evidence in support of language-non-selective word recognition must be augmented by similar evidence from studies where the critical words are presented in veridical linguistic contexts. A number of these have been conducted. In most of them sentence contexts were used, both these contexts and the target words were presented visually, and cognate effects and/or interlexical homograph effects were used as the markers of language-non-selective activation (see Chambers & Cooke, 2009, for a related study that examines how bilinguals recognize spoken words in sentence context).

The dependent variables were reaction times, ERPs, eye-fixation measures, and combinations of these. These studies have produced mixed results. Cognate effects still showed up when the critical words were preceded by low-constraint sentence contexts, that is, sentence contexts that do not severely constrain the set of possible target words (e.g., *Across from the supermarket stood an old* tree *which was home to a lot of birds*, with *tree* as target word), especially when the experiment was conducted in L2, the weaker language (Duyck, Van Assche, Drieghe, & Hartsuiker, 2007; Libben & Titone, 2009; Schwartz & Kroll, 2006; Titone et al., 2011; Van Assche et al., 2011; van Hell & de Groot, 2008). When low-constraint sentences were used and the experiment was run in dominant L1 the effect is less clear-cut but has also been obtained (Titone et al., 2011; Van Assche et al., 2009). When high-constraint sentences were used (e.g., *The bird sat in the branches of the highest* tree *because cats could not reach him there*, again with *tree* as target word) often no cognate effect occurred (but see Van Assche et al., 2011).

Contrary to the cognate effects, homograph effects were generally absent (Elston-Güttler, Gunter, & Kotz, 2005; Schwartz & Arêas Da Luz Fontes, 2008; Schwartz & Kroll, 2006; Titone et al., 2011; but see Libben & Titone, 2009), although during the first part of Elston-Güttler et al.'s experiment, but not the second, such an effect *did* show up when before presenting the experimental sentences the activation level of the non-target language was boosted by showing a film fragment narrated in the non-target language. The finding that under these circumstances the effect is present at first but then disappears suggests again (see also von Studnitz & Green, 2002) that length of practice in the experimental task modulates the influence of the non-target language.

In conclusion, the joint findings from the bilingual word-recognition studies indicate that the presentation of a word, printed or spoken, to a bilingual induces parallel activation in both sub-lexicons, but they also point out that (the effect of) this phenomenon is modulated by three factors: relative proficiency in L1 and L2, stimulus type (cognates vs. inter-lexical homographs), and degree of sentence constraint. In addition, it appears that task practice can reduce or annihilate (the effect of) co-activation in the non-target language.

18.4 Producing single words

The evidence provided above that word recognition involves language-non-selective activation in the bilingual lexicon does not necessarily imply that the same holds for word production. Word recognition in fluent language users is an automatic process, and if a bilingual is fluent in both languages, word recognition will come about automatically in both languages. In other words, word recognition in fluent bilinguals will automatically activate representations in both sub-lexicons. As we claimed before (Starreveld, de Groot, Rossmark, & van Hell, 2014), for this reason evidence of language-non-selective lexical activation as obtained in word-recognition studies is in fact rather trivial. On the other hand, word production is a controlled, attention-demanding process. It is therefore conceivable that bilingual speakers can choose to focus attention exclusively on the sub-lexicon of the language they intend to speak, only activating representations in this sub-lexicon (see also Costa & Santesteban, 2004b).

Nevertheless, there is evidence to suggest that also during bilingual word production elements in the non-response (non-target) language get activated. This evidence has primarily been gathered in experiments that exploited one or another version of the task that is most frequently used in research on word production, both monolingual and bilingual: the picture-naming task. A crucial assumption underlying the use of this task in studying word production is that after the completion of the first stage of the picture-naming process, the computation of the visual percept, picture naming (invoked by an external stimulus) and word production (invoked by internal thought processes) involve the same processing stages: the activation of the appropriate concept, the selection of the target word from the mental lexicon, phonological encoding, phonetic encoding, and articulation (e.g., Levelt et al., 1998). Word-production research usually focuses on the first three of these five stages and the corresponding representations in the mental lexicon: the conceptual/semantic, lexical, and phonological representations, respectively.

18.4.1 The picture–word interference task

In one version of the picture-naming task, the picture–word interference task, each picture is accompanied by a spoken or printed distracter that shares a specific relation with the pictured entity (or its name) or is unrelated to it. Monolingual experiments in which this task version was used (e.g., Schriefers, Meyer, & Levelt, 1990; Starreveld & La Heij, 1995) have shown that, at specific intervals between the presentation of the picture and the distracter, semantic distracters (that share a semantic relation with the pictured object; e.g., a picture of a cat accompanied by the distracter word *dog*) slow down picture naming as compared with unrelated distracters (e.g., *mug*) whereas phonological distracters (sharing a phonological relation with the picture's name; e.g., a picture of a cat accompanied by the distracter word *cap*) speed up naming, again as compared with an unrelated distracter word. These two effects are known as the semantic-interference effect and the phonological-facilitation effect, respectively.

In explanation of the semantic-interference effect, models of monolingual word production hold the assumption that, following the activation of a conceptual representation (or, simply, concept) upon the presentation of a picture, not only the lexical representation associated with this concept (e.g., the cat concept) becomes activated next, but also those associated with semantically related concepts (e.g., the concepts for dog and tiger). Depending upon the type of word-production model that is adopted, with distributed or localist conceptual representations (see de Groot, 2011: 132–135 and 230–231, for details), these non-targeted lexical representations receive their activation directly from the concept activated by the picture, or indirectly, via a process whereby activation spreads within a semantic network from the concept activated by the picture to semantically related concepts. The latter then transmit activation to the corresponding lexical representations.

All activated lexical representations compete with one another during lexical selection, a process that, if all goes well, results in selecting the lexical representation associated with the target. The other activated ~~but~~ non-~~lexical~~ representations interfere with the selection process. The more highly they are activated the more interference they will cause and the longer lexical selection and, thus, responding will take. Both semantic and unrelated distracters will activate their corresponding lexical representations, through the automatic bottom-up word-recognition process elicited by the distracter. Consequently, the lexical representations of both types of distracters will act as competitors during lexical selection and delay the response. However, as described above, the lexical representation associated with a semantic distracter will receive additional, top-down, activation from the conceptual representation activated by the picture. The high level of activation resulting from these two sources combined will have the effect that the lexical representations of semantic distracters are stronger competitors than the lexical representations of unrelated distracters. This is assumed to be the reason why picture naming takes longer when the picture is paired with a semantic distracter than when it is paired with an unrelated distracter.

! the other activated lexical representations

A phonological distracter (e.g., the word *cap* accompanying a picture of a cat) will also initiate a bottom-up recognition process during which the corresponding lexical representation will be activated, but the level of activation in this representation is not increased further by the top-down activation process incited by the picture. But why, if both phonological and unrelated distracters give rise to bottom-up but no top-down activation from conceptual representations, does the phonological-facilitation effect occur at all? The reason is that a phonological distracter, but not an unrelated distracter, pre-activates part of the presented picture's phonological representation (the /ca/ part), thus facilitating phonological encoding of the picture's name.

Starreveld and La Heij (1996) implemented the assumptions underlying both explanations mentioned above in a connectionist model. Simulated word-production latencies indeed showed semantic effects, phonological effects, and the interactions of those effects (see also Roelofs, 1992, for a related neural network model that produced semantic effects), indicating that the assumptions are computationally plausible. Because in models of word production (and word recognition) memory representations are usually called nodes, we will henceforth use the terms representations and nodes interchangeably.

18.4.1.1 Bilingual studies

While monolingual picture-word interference studies have shown that an activated concept excites the lexical representations of a set of semantically related words, bilingual studies have used this methodology to see whether the set of activated lexical representations also includes the one for the concept's name in the non-response language (i.e., its translation; Costa et al., 2003; Hermans et al., 1998). To be able to detect parallel activation in the non-response language, in addition to the semantic, phonological, and unrelated distracters used in the monolingual studies, one further critical type of distracter word was included in these studies: distracters that were phonologically related to the picture's name in the non-response language. For instance, for a Dutch–English bilingual asked to name pictures in English and presented with a picture of a frog, the word *kitchen* would be a *phonological-translation distracter* (a frog is called *kikker* in Dutch). If the activated frog concept activates the lexical nodes for both *frog* and *kikker*, the combined activation converging on the *kikker* node from the activated concept on the one hand and the phonological-translation distracter *kitchen* on the other hand will turn this node into a strong competitor for the *frog* response and delay it, again as compared with a condition with unrelated distracters. This interference effect has indeed been observed during L2 picture naming, both in unbalanced bilinguals (Hermans et al., 1998) and in proficient bilinguals (Costa et al., 2003). The effect was particularly robust when the distracters were words in the non-response language (which in both studies was the participants' L1) and when at the same time relatively many of the distracters were related to the picture's name (Hermans et al., 1998, Experiment 2; Costa et al., 2003, Experiment 1). When the distracters were words from the response language L2 (Hermans et al., Experiment 1; see the above example) or when the proportion of related to unrelated distracters was reduced by the inclusion of unrelated filler trials (Costa et al., 2003, Experiment 2), the effect was less robust, only showing up when the interval between the presentation of picture and distracter was relatively long.

Perusal of the reports of the original studies shows that the use of the picture-word interference task produces complex and not seldom equivocal data patterns. As pointed out by various researchers (e.g., Kroll, Sumutka, & Schwartz, 2005; Starreveld, 2000), this is probably due to the simultaneous presentation of two stimuli, the picture and the distracter, each of them triggering a separate process (of encoding and decoding, respectively) and the two unfolding in opposite directions. An effect of a distracter can only be observed if the top-down activation triggered by the picture and the bottom-up activation elicited by the distracter meet, and are measured, at a moment that both activation streams have had a chance to build up to a sufficiently strong level to be detected at all and neither of the two has decayed below some critical minimal level. Due to the sequential nature of the different stages involved in word production and word recognition, different types of distracters exert their effect at different intervals between the presentation of picture and distracter. For instance, to detect an effect of an aurally presented phonological distracter, a relatively long interval between the presentation of, first, the picture and, next, the distracter is required. This is because otherwise the processing of non-overlapping phonological representations of a spoken distracter might overwrite the processing of the overlapping representations completely, thus preventing the occurrence of the phonological effect (see Starreveld, 2000, for details). Therefore, the absence of a phonological-translation effect with some picture-distracter intervals can per se be perfectly consistent with a language-non-selective account of bilingual word production and even be predicted.

But what, given the complexity of the paradigm, may also happen is that one predicts a phonological-translation effect to occur at a particular picture-distracter interval and yet obtains a null-effect, not because the word-production process is in fact language-selective but because of inaccurate assumptions about the exact intersection point between the top-down and bottom-up streams of activation and the choice of picture-distracter interval based on these incorrect assumptions. Imagine, for instance, two bilingual studies in which the stimuli are presented with exactly the same interval between picture and phonological-translation distracter, that both ask for responses in the participants' L2, but that test participants with different levels of L2 proficiency. Even if, in actual fact, language-non-selective word production would hold for both participant groups, one of the groups might show a response pattern suggesting language-selective word production (equally long naming times for trials with phonological-translation distracters on the one hand and unrelated distracters on the other hand), because the chosen picture–word interval caused the two activation streams to miss each other.

Given the equivocal data patterns that picture–word interference studies may produce, converging evidence for language-(non)selective bilingual word production from other tasks is desirable. Other reasons why additional evidence is needed are the findings, mentioned above, that phonological-translation effects occurred most reliably when L2 was the response language and the distracters were L1 words, or when a relatively large proportion of related to unrelated picture-distracter pairs was used. The first of these two findings suggests that the effect might at least partly be due to the subjects being put in a bilingual mode by the presence of both languages during task performance (e.g., Grosjean, 1998), the non-response language of the distracters keeping this language activated as well. As pointed out by Costa et al. (2003), the second finding suggests that, when presented with a large proportion of related picture-distracter pairs, the participants may notice the relationships between picture names and distracters and develop a response strategy that deviates from the instruction, common in these experiments, to ignore the distracter words. Conscious attention to the distracter words may, again, have the effect of putting the participants in a bilingual mode.

18.4.2 Simple picture naming

In fact, additional evidence that word production in bilinguals involves the simultaneous activation of representations in both sub-lexicons already exists. Most of it has been gathered in studies that employed the simplest possible version of the picture-naming task, in which each picture is presented alone, unaccompanied by a distracter, and has to be named out loud (e.g., Christoffels, de Groot, & Kroll, 2006; Costa, Caramazza, & Sebastián-Gallés, 2000; Hoshino & Kroll, 2007; Starreveld et al., 2014). Crucially, in these studies the picture names were either cognates or non-cognates in the participants' two languages. Costa and colleagues, who were the first to use this paradigm in a Catalan–Spanish study, argued that the occurrence of a cognate effect, shorter naming times for pictures with cognate names than for those with non-cognate names, indexes parallel activation in the bilingual word-production system. What is more, they reasoned that it would indicate that language-non-selective activation holds for both main stages of word production following concept activation: the activation of the lexical representations and, next, the phonological representations. Figure 18.1 illustrates (with a Dutch–English example taken from Starreveld et al., 2014) Costa et al.'s view of what happens when bilinguals name pictures with cognate and non-cognate names and the underlying assumptions about the structure of the word-production system, containing three levels of representations.

After visual processing, the picture first activates the corresponding semantic representation or concept, which consists of a set of nodes in a semantic representational level that is shared between the two languages. The activated semantic nodes propagate activation forward to the associated lexical nodes in both languages and the latter send on activation to a level that stores sub-lexical phonological nodes. Like the level of semantic representations, the level containing the sub-lexical nodes is shared between the two languages. Consequently (most of) the sub-lexical nodes

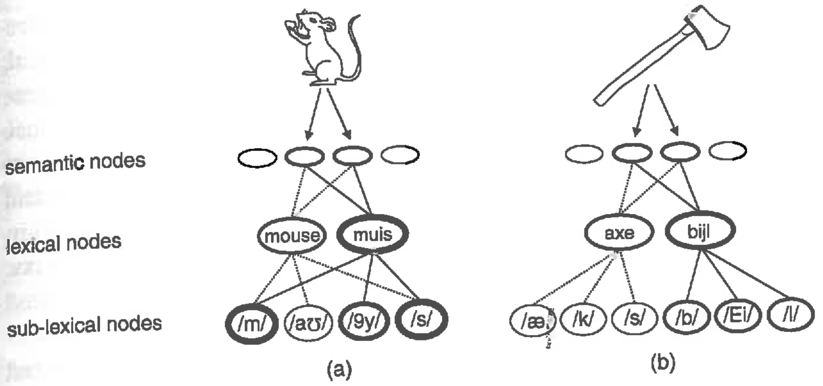

Figure 18.1 A model of picture naming in bilinguals (based on Costa et al., 2000). Note: All connections shown in the model are bidirectional; thicker ellipses represent more activation. Panels a and b illustrate the activation of the memory representations of the (English–Dutch) cognate pair *mouse–muis* and the non-cognate pair *axe–bijl* when the picture is named in Dutch and activation has just reached the sub-lexical nodes. Subsequently, the lexical node of the cognate *muis* will receive more feedback from the sub-lexical layer to the lexical layer than the lexical node of the non-cognate 'bijl.' As a result, with the continued spread of activation through the network, the complete set of sub-lexical nodes representing a cognate will receive more activation than the one representing a non-cognate.

that represent cognate names accumulate activation from two previously activated lexical nodes (Figure 18.1a), whereas those representing non-cognate names receive activation from one lexical node only (Figure 18.1b). Furthermore, the cognates' lexical nodes receive more activation than those of non-cognates due to feedback from the phonological level back to the lexical level. Because the lexical nodes continue to feed activation forward to the sub-lexical nodes, the differential activation in the lexical nodes for cognates and non-cognates caused by this feedback also results in a relatively high level of activation in the sub-lexical nodes representing cognates. The differential activation in the sub-lexical nodes for cognates and non-cognates resulting from these feed-forward and feed-backward activation processes results in faster naming for cognates.

The cognate effect has shown up with both dominant L1 and weaker L2 as response languages, though it was generally larger in the latter case (Christoffels et al., 2006; Costa et al., 2000; Poarch & van Hell, 2012; Starreveld et al., 2014). In other words, relative language proficiency in the two languages modulates the non-response language's influence on producing words in the target language, just like it has been shown to influence the non-response language's role in word recognition (e.g., Marian & Spivey, 2003b; Weber & Cutler, 2004). The different magnitude of the cognate effect in the participants' two languages is attributed to a difference in the strength of the connections between the conceptual, lexical, and sub-lexical phonological nodes in the two language-subsystems, both the links between the conceptual and lexical nodes and those between the lexical and sub-lexical nodes being stronger for dominant L1 than for L2 (Costa et al., 2000; Starreveld et al., 2014; weaker and stronger links are represented by dotted and solid lines, respectively, in Figure 18.1). Stronger links transmit more activation than weaker links. Consequently, when dominant L1 is the response language the targeted sub-lexical nodes receive less activation from the lexical node in the non-response language than when weaker L2 is the response language. Stated differently, cognate naming in weaker L2 benefits more from activation in stronger L1 than vice versa.

Although only fleetingly alluded to in a footnote, Starreveld at al. (2014) found that picture-naming repetition also modulates the cognate effect in word production. The cognate effect gradually became smaller over four presentation series that all included the same set of pictures to name. When the pictures had to be named in L2 English, the participants' weaker language, the effect was 151 ms, 113 ms, 90 ms, and 65 ms in Series 1, 2, 3, and 4, respectively, and significant in all cases. When the pictures were named in L1 Dutch, the cognate effect was 35 ms, 15 ms, 1 ms, and 7 ms in Series 1, 2, 3, and 4, respectively, and only significant in Series 1 and 2. Correlated with the cognate effect, naming time decreased considerably in the successive presentation series: overall naming time was 924 ms, 798 ms, 750 ms, and 728 ms, respectively, when the pictures were named in weaker English, and 737 ms, 653 ms, 638 ms, and 635 ms, respectively, when they were named in dominant Dutch.

The correlation between naming speed and the size of the cognate effect, as shown in Starreveld et al.'s study for both the effects of relative language proficiency and stimulus repetition, appears to be a general phenomenon (see below for two further demonstrations). Importantly, we believe that the null-effects obtained when the pictures were named in dominant Dutch and the participants had already named the stimuli a couple of times before do not point at language-selective processing, because how could the bilingual word-production system have evolved from one operating in a language-non-selective manner (the system illustrated in Figure 18.1) to a language-selective system during just one half hour or so? An alternative interpretation of these null-effects, which we consider to be more likely, is that stimulus repetition brings about temporary changes in components of the word-production system, for instance, in the strength of the connections between the memory nodes and/or the baseline level of activation in the memory nodes that are exploited during naming. These temporary changes may then influence both the size of the cognate effect and naming time. Similarly, it is unlikely that the smaller vs. larger cognate effects associated with naming in the

S //

stronger and weaker language, respectively, indicate that word production is language-non-selective in both cases but less so with naming in the stronger language. Given the word-production system shown in Figure 18.1, the activation pattern that follows picture presentation is *always* language non-selective. However, *the effect of* the activation of the non-response language, as manifested in the size of the cognate effect, varies with, for instance, the strength of the connections between the various types of nodes and the associated difference in activation transmission across these connections.

We are aware of the drastic theoretical consequences of these claims because if even a zero cognate-effect is not regarded as an unequivocal marker of language-selective word production, what is? Nonetheless, simple computer simulations to be presented later demonstrate how, indeed, varying parameters such as the connection strength of the links between nodes and the nodes' activation level affect both the magnitude of the cognate effect and naming time. But before presenting these data we will provide evidence of parallel language activation in bilinguals' word production as obtained in studies using one further version of the picture-naming task, starting that section by expressing a concern, hitherto ignored, related to using the cognate effect as a marker of language-non-selective activation. In addition, a couple of studies that examined word production in sentence context will also be discussed first.

18.4.3 Phoneme monitoring

As described above, the relatively fast naming for pictures with cognate names as observed in the simple picture-naming task is attributed to activation converging on the sub-lexical phonological nodes from two sources: the concept's lexical nodes in both languages. In other words, the effect is explained in terms of the simultaneous activation of two lexical nodes that benefits cognates but not non-cognates while, implicitly, the representation structures are assumed not to differ between cognates and non-cognates. Yet, it is disputed whether this implicit assumption is a legitimate one. In fact, indications that the representations of cognates and non-cognates differ from one another have been gathered in more than an occasional study and the authors of these studies have suggested a number of possible representational differences between the two types of words. For instance, a pair of cognates may share a morphological representation in memory whereas a pair of non-cognates does not (Sánchez-Casas & García-Albea, 2005), or cognate pairs may share a larger part of their meaning representations than pairs of non-cognates do (van Hell & de Groot, 1998). If such representational differences indeed exist, instead of a process of parallel activation that affects cognates and non-cognates differently, these might somehow underlie the cognate effect in picture naming.

Given the concerns raised earlier regarding the picture-word interference task and the present concern about the use of cognates in the simple picture-naming task, additional evidence for parallel activation in bilinguals' word production is desirable. In a study with fluent Catalan–Spanish bilinguals as participants, Colomé (2001) introduced a clever new version of the picture-naming task by which the problems associated with the use of the other task versions can be circumvented: it does not require the use of cognate stimuli, nor the use of distracters that, if they involve words from the non-response language, may put the participants in a bilingual mode. In this study each picture was accompanied by a phoneme (more precisely, by a letter that represented this phoneme) and the participant had to indicate, by pushing one of two response keys, whether or not that phoneme occurred in the picture's Catalan name, Catalan being the participants' L1. All pictures in the experimental stimuli had non-cognate names and care was taken that the participants had no reason to suspect that their bilingualism was being tested so that deliberate activation of the non-response language was unlikely. (Unfortunately though, some pictures in a set of filler stimuli had cognate names; we will return to this point below.)

To detect co-activation in the non-response language that might nevertheless occur, in addition to stimuli inviting a "yes" response (e.g., a picture of a chair, *cadira* in Catalan, accompanied by the letter *c*), two types of stimuli requiring a *no* response were presented, labeled *no-translation* and *no-unrelated* stimuli here. The phoneme that accompanied the picture in a no-translation stimulus did not occur in the picture's Catalan name but was the first phoneme of its Spanish name (e.g., the same picture of a chair but now paired with *s*, a chair being called *silla* in Spanish). Finally, the phoneme in a no-unrelated stimulus occurred in neither the picture's Catalan name nor in its Spanish name (e.g., the same picture again but now paired with a totally unrelated letter, for example, the letter *m*).

To perform this *phoneme-monitoring* task, a participant must first tacitly generate the picture's name and then monitor the internally generated name for the presence of the specified phoneme. Longer response times for no-translation stimuli than for no-unrelated stimuli (the *no-translation effect*) serve as the marker of parallel activation in both sub-lexicons during word production, on the assumption that co-activation of the picture's name in the non-response language is particularly harmful for the former of these two types of no-trials. One reason might be that, given parallel activation in both sub-lexicons, the phoneme presented in a no-translation stimulus and occurring as part of the co-activated memory units in the non-response language, would create a tendency in the participants toward the *yes*-response; overcoming this tendency would consume additional processing time. Alternatively, and again given parallel activation in both sub-lexicons, the incongruence between a tendency to respond *yes* on the one hand (because the picture's name in the non-response language contains the specified phoneme) and a simultaneous tendency to respond *no* (because the picture's name in the response language does not contain this phoneme) causes the delay in the no-translation condition (note that the no-unrelated condition produces two congruent *no*-response tendencies).

In three experiments, in which different intervals between the presentation of picture and phoneme were used, Colomé (2001) found that the no-translation stimuli indeed took longer to respond to than the no-unrelated stimuli. The magnitude of this effect varied between 41 ms and 55 ms over the three experiments. The overall response times for the two types of no-trials varied between 952 ms and 1132 ms across the three experiments. The longer response times for the phoneme-monitoring task in comparison with the times required for simple picture naming (cf. the response times provided above for the latter task) plausibly reflect the relative complexity of the phoneme-monitoring task, which requires tacit naming, keeping both the generated name and the phoneme available in working memory, conscious monitoring of the internally generated name for the presence of the phoneme, converting the outcome of this matching process into a *yes* or *no* response and, finally, response execution.

Several studies have followed up on Colomé's (2001) study. In a Dutch–English study testing unbalanced bilinguals in their weaker L2, Hermans et al. (2011) wondered whether the set of filler materials that Colomé used in her study might have contained pictures with cognate names in Catalan and Spanish and, if so, how this might have affected the results. After personal communication with Colomé had confirmed that, indeed, there were a few pictures with cognate names among her filler materials, Hermans et al. examined the influence of the proportion of pictures with cognate names among the filler materials on the no-translation effect. As in Colomé's study, the pictures used in the experimental stimuli all had non-cognate names.

In a first experiment in which the pictures in the filler stimuli, ~~also 24 in all;~~ all had non-cognate names, no no-translation effect occurred. In a second experiment in which pictures in the filler stimuli all had cognate names (so that, overall, 50% of the pictures had cognate names and the remaining 50% had non-cognate names) a significant no-translation effect was obtained: Responding took 37 ms longer in the no-translation condition than in the no-unrelated condition. Interestingly, in a third experiment wherein only 25% of the filler stimuli had pictures with cognate names (so that, overall, 87.5% non-cognate pictures and 12.5% cognate pictures were used) a statistically equally large (and numerically even larger) no-translation effect of 53 ms was obtained. The authors concluded that, depending on the composition of the stimulus set, "the bilingual language production system ... can operate in different language activation states" (Hermans et al., 2011: 1696) or, in other words, that "language activation in bilinguals' speech production is dynamic" (the article's title). Furthermore, from the fact that the experiment with only 12.5% cognate pictures produced the same data pattern as Colomé's (2001) study, despite the fact that the two studies tested different types of bilinguals (fluent, early bilinguals vs. unbalanced, late bilinguals), Hermans and colleagues concluded that their findings appear to be "impervious to variations in language learning history and proficiency" (p. 1702).

Hermans et al.'s (2011) results are consistent with studies on visual word recognition that have shown the inclusion of words from the other language to influence the response patterns (e.g., de Groot et al., 2000; Dijkstra et al., 1998; von Studnitz & Green, 2002) and with the earlier indication that in the picture-word interference studies with L2 as the response language and L1 distracter words (Costa et al., 2003; Hermans et al., 1998, Experiment 2) the distracters may have put the participants in a bilingual mode. The present results add the important suggestion that for this effect to occur the words from the non-response language do not have to be present in the form of an external stimulus but that it suffices if they are internally generated. Apparently, if a cognate name instead of a non-cognate name is generated on only a small proportion of trials, this (or, perhaps, noticing this) boosts the activation of the non-response language and a no-translation effect ensues.

While the conclusion that bilingual speech production is dynamic is thus substantiated by other studies, Hermans et al.'s (2011) suggestion that the data are resistant to variations in language learning history and proficiency is nuanced by de Groot, Starreveld, and Geambaçu (in preparation). We drew our participants from the same population as Hermans et al. did, but unlike them we asked one group of participants to respond in their weaker L2 English while for a second group dominant L1 Dutch was the response language. The same picture set was used in both language conditions. Despite a vigorous attempt to only have pictures with non-cognate names as stimuli, it could not be prevented that the names of 6 out of the 22 filler pictures (13.6% of the total stimulus set) embedded a (non-identical) cognate as part of a polysyllabic word (e.g., *seahorse-zeepaardje*, *sea* and *zee* being cognates). (The set of 520 pictures from which we selected our stimuli did not contain any more pictures that met all our selection constraints.)

When L2 English was the response language, we observed a no-translation effect of 50 ms in a first presentation series and a statistically equally large effect of 38 ms in a second series in which all stimuli were presented a second time. These effects are comparable in size to those in Colomé (2001) and in Hermans et al.'s (2011) two experiments that included cognate pictures among the filler materials. However, despite the fact that the same picture materials, including the embedded-cognate filler stimuli, were used in both language conditions, on neither of the two presentation rounds was a no-translation effect obtained when dominant Dutch was the response language. In other words, the presence of cognate stimuli is no guarantee that the effect occurs. To be absolutely sure that this null-effect was real, we selected a further group of participants from the same population and repeated the experiment with dominant Dutch as response language. Again, both presentation rounds showed a null-effect. It thus appears that relative language proficiency modulates the critical effect in the phoneme-monitoring task, just like it modulates the cognate effect in simple picture naming and the markers of parallel activation in bilingual word recognition.

Rodriguez-Fornells et al. (2005) used a modified version of the phoneme-monitoring task that, like the original task, requires phonological encoding and mental inspection of the pictures' names. These researchers asked Spanish–German bilinguals and monolingual German speakers to push a response button if the picture's name in the response language started with a vowel and to not respond if it started with a consonant, or vice versa, using behavioral measures, ERPs and fMRI data as dependent variables. Crucially, in a *coincidence* condition, the pictures' names started with a vowel or a consonant in both of the bilinguals' two languages (e.g., *asno–Esel*, 'donkey'; *vela–Kerze*, 'candle'), whereas in a *non-coincidence* condition the pictures' names started with a consonant in one language and with a vowel in the other (e.g., *embudo–Trichter*, 'funnel'). All response measures showed an effect of the coincidence manipulation for the bilinguals, but not for the monolinguals. For instance, only the bilinguals made fewer errors in the coincidence condition than in the non-coincidence condition and only the bilinguals showed a different ERP pattern for these two conditions. These results point, again, at parallel phonological encoding in both languages, the activation in the non-response language affecting the response differently in the coincidence and non-coincidence conditions. For example, the opposite response tendencies incited by the pictures' names in both languages when one starts with a vowel and the other with a consonant likely led to the relatively large number of errors in the non-coincidence condition. Interestingly, the fMRI data revealed that the bilingual participants exploited brain areas that are also used to control behavior in non-verbal tasks, probably in order to cope with the interference caused by activation of the non-response language. These results accord with those of studies on bilingual language control as examined by means of the language-switching paradigm (e.g., Abutalebi et al., 2008; Hernandez & Meschyan, 2006).

18.5 Producing words in context

Just like language comprehension normally involves the recognition of words presented in a larger linguistic context, during normal speech production words are spoken in the context of larger linguistic utterances (and non-linguistic sources of information, ignored here). This larger linguistic context may modulate the influence of the non-response language in word production, as it does in word recognition. To examine this possibility, Starreveld et al. (2014) and de Groot et al. (in preparation) not only studied picture naming out of context (as discussed above), but also included experiments wherein the pictures to name were preceded by a sentence fragment presented visually, word by word. In Starreveld et al. both high-constraint and low-constraint sentence fragments were used (see section 18.3) whereas de Groot et al. only presented low-constraint sentence fragments. As described before, in Starreveld et al. the participants named the pictures aloud and the cognate effect served as marker of parallel activation whereas de Groot et al. used the phoneme-monitoring task, involving tacit naming, and parallel activation was indexed by the no-translation effect. The two studies tested participants from the same population: late Dutch–English bilinguals with a relatively high level of L2 English but clearly dominant in L1 Dutch. In these subjects the dominance of Dutch over English is more obvious for production than for comprehension, presumably because they use English actively less often than passively.

Recall that Starreveld et al. (2014) obtained a cognate effect when the pictures were named out of context, both with L1 Dutch and L2 English as response language, though the effect was substantially larger in the latter case (Dutch: 35 ms; English: 151 ms). Furthermore, in both language conditions the cognate effect decreased with picture repetition and in the Dutch condition it was annihilated after the first two presentation rounds (see section 18.4.2). In out-of-context phoneme monitoring (de Groot et al., in preparation; see section 18.4.3), we observed a no-translation effect when English was the response language and this effect was (statistically) equally large in two subsequent presentation rounds (50 ms and 38 ms). When Dutch was the response language no no-translation effect occurred.

When in Starreveld et al.'s (2014) study the pictures were preceded by low-constraint sentence fragments, cognate effects still turned up in both language conditions. The size of these effects (17 ms in Condition Dutch; 44 ms in Condition English) was comparable to the effect of 35 ms observed in the Dutch out-of-context condition (the statistical analysis showed no difference between these three effects), but the effects were substantially smaller than the 151 ms effect obtained in the English out-of-context condition. Interestingly, response times were considerably shorter when the pictures were preceded by context (704 ms overall) than when presented alone (830 ms overall). When the pictures were preceded by high-constraint sentence fragments, response time reduced even further to 464 ms and 613 ms overall with Dutch and English as response language, respectively. A cognate effect (of 28 ms) now still turned up when the pictures were named in English, but not when Dutch was the response language. This general pattern of results once again illustrates the close relation between the size of the cognate effect and naming speed. In section 18.4.2 we demonstrated this relation to hold for the effects of picture repetition and relative language proficiency. Furthermore, our finding that degree of sentence constraint modulates the cognate effect agrees with, and extends, a similar result from studies on bilingual word recognition in context (see section 18.3).

Finally, in de Groot et al. (in preparation) we obtained an equally large (39 ms) no-translation effect in both language conditions during phoneme monitoring in context. Unlike in Starreveld et al. (2014), in this study no clear relation between the size of the critical effect, here the no-translation effect, and response time was observed.

To summarize, the joint set of studies on bilingual word production reviewed above indicates that the production of single words involves the parallel activation of representations in both sub-lexicons but that several factors modify the influence that the activated representations in the non-target language's sub-lexicon exert on performance. First, like in the studies on bilingual word recognition reviewed in the first part of this chapter, relative language proficiency in L1 and L2 turns out to be one of the relevant factors. This variable affects the response pattern both in simple out-of-context picture naming and in out-of-context phoneme monitoring, the cognate effect in simple picture naming being smaller in L1 than in L2 and the no-translation effect in phoneme monitoring only occurring in L2. Second, the influence of the other language depends on the composition of the stimulus set, as is shown from the effect of having pictures with cognate names among the stimulus materials in out-of-context phoneme monitoring. Third, stimulus repetition is a relevant factor, as is indicated by the reduction of the cognate effect in successive presentations of the same pictures in the out-of-context simple-picture-naming studies. Interestingly, in out-of-context phoneme monitoring the no-translation effect was not reduced by stimulus repetition. Fourth, the in-context overt-picture-naming study (but not the phoneme-monitoring study) revealed that sentence context tempers the influence of the non-target language, in particular during word production in the weaker language. This can be concluded from the substantial reduction of the cognate effect when the picture was preceded by a sentence fragment and named in L2. Fifth, the type of sentence context has been identified as a relevant variable, as can be concluded from the fact that no cognate effect occurred when the picture was preceded by a high-constraint sentence fragment and named in dominant L1.

In the next section we will redeem our earlier promise to show by means of simple computer simulations that variations in the strength of the connections between nodes in the bilingual lexicon and in the nodes' activation levels can, in principle, account for fluctuations in the size of the cognate effect in picture naming despite the fact that the system always operates in a language-non-selective manner. In this section we will not only simulate the cognate effect in picture naming itself, but also how it is modulated by relative language proficiency, picture repetition, and sentence context.

18.6 Computer simulations

18.6.1 Models and assumptions

In order to show that our theoretical interpretations of the effect of relative language proficiency, sentence context, the cognate effect, and the modifying effect of picture repetition on the size of the cognate effect are computationally realistic, we performed a number of computer simulations. To keep the simulations as insightful as possible, we kept them as simple as possible. We simulated the time course of activation of only one of the phonological output nodes of both a cognate and of a non-cognate under various conditions.

We used two simple models that both consisted of two layers. The output layer of both models contained only one output node. For each simulation the time course of activation of this node is reported below. For the simulations involving the output node of a cognate, the input layer contained two nodes, one representing the lexical node in the target language (T) and the other representing the lexical node in the non-target language (nT, see Figure 18.2a). For the simulations involving the output node of a non-cognate, the input layer contained just one node, representing the lexical node in the target language (T, see Figure 18.2b). The resulting implemented

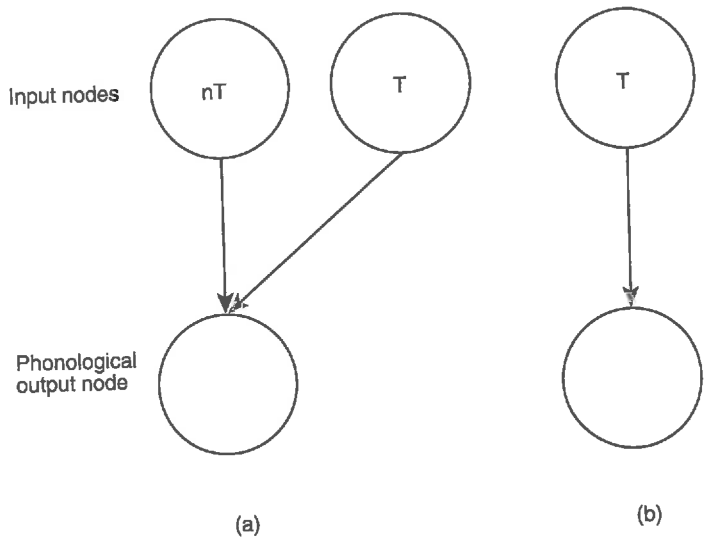

Figure 18.2 The basic layout of the networks used in our simulations of the activation of a phonological node of a cognate (a) and a non-cognate (b). T = target language; nT = non-target language

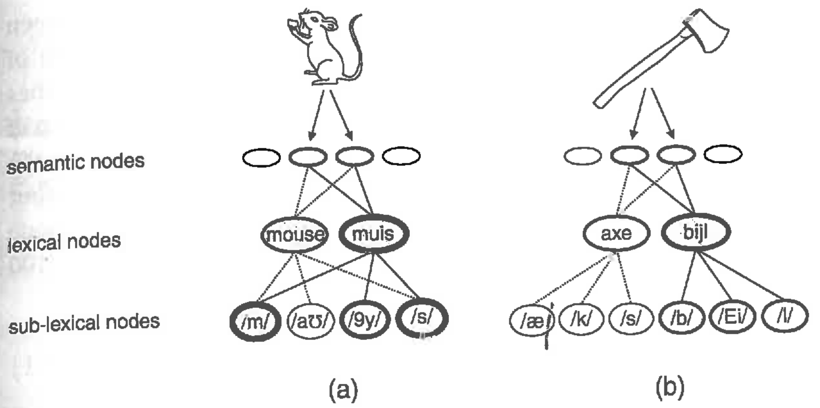

Figure 18.3 A model of picture naming in bilinguals. The nodes indicated with ellipses correspond to the nodes of the implemented models depicted in Figure 18.2

shaded !

models correspond to the patterned light gray nodes of the model depicted in Figure 18.3.

Two assumptions were made to keep processing within the models simple. First, we kept the activation level of the input nodes constant during each simulation. A slow accrual of activation of these nodes, which would reflect lexical activation processes better than a constant activation level, would just decrease the speed of activation accrual at the output layer but would not change the general behavior of the model. Second, connections were assumed to be one-way only, from the input nodes to the output nodes. In the model depicted in Figure 18.3 (see also Figure 18.1), connections are reciprocal, so our present implemented models are simplified versions of that model.

All connections in our models had strengths, expressed as weights, referred to with parameter w. At the start of the simulation, input activation (A_{input}) was applied to the input nodes and activation was propagated through the network in successive steps, called iterations. An iteration can be conceived of as a time step and any number of iterations can directly be linked to processing time by multiplying the iteration number with a constant. For example, if it is assumed that one iteration takes 4 ms, an effect of 10 iterations corresponds to a 40 ms effect in RT data.

At each iteration, an activation function was applied to calculate the activation of the output node. In order to simulate the spreading activation process, we used a basic activation function that calculates the amount of activation that reached the output node at each iteration by just multiplying the activation of each input node with the weight of the connection between that input node and the output node. We added the resulting amounts of activation, both within an iteration and between iterations. These calculations were performed by the repeated use of formula (1) in which i is the iteration number, j is an index that identifies the input nodes, n is the number of input nodes (two for the simulation of an output node of a cognate, one for the simulation of an output node of a non-cognate), and w_j is the weight connecting input node j to the output node. In all simulations it was assumed that selection of the output node occurred when its activation level reached a critical threshold of 100 activation units.

$$A_{output}(i+1) = A_{output}(i) + \sum_{j=1}^{n} A_{input}(j) * w_j \qquad (1)$$

18.6.2 Simulations of the effects of relative language proficiency and sentence context on selection times

Using these very basic assumptions, we first simulated the effect of relative language proficiency in L1 and L2 on the selection times of a non-cognate output node (the corresponding model is depicted in Figure 18.2b). If it is assumed that language proficiency involves different weights between nodes in the language system, changing the weights should affect processing time. In our simulations, we set the input activation to 1 and kept it constant. For the simulation of a high level of language proficiency (as in L1), we set the weight between the input and output node to 3. For the simulation of a lower level of language proficiency (as in L2), we (arbitrarily) set the weight between the input and output node to 2. The results of the simulations showed that the threshold of 100 activation units was reached in 34 and 50 iterations, respectively, for the simulation of picture naming in dominant L1 and in weaker L2 (Figure 18.4a, dotted and solid line, respectively), an effect of 16 iterations. Note that the size of the effect is linearly related to the strength of the weights, so the model can simulate any effect size by adjusting the strength of the weights. Therefore, the fit of the model to actual data is trivial, but we note that the model naturally accounts for the effects of relative language proficiency on selection times in terms of weight differences.

Next, we simulated the effect of a low-constraint sentence context on the selection times of a non-cognate output node. If it is assumed that a sentence context increases the amount of activation in the corresponding language sub-system (e.g., Starreveld et al., 2014), an increase of the input activation in the model should mimic the effects of sentence context we obtained empirically.

Interestingly, models of the kind we used here exhibit the feature that any effect of a change in the strength of the weights can also be obtained by keeping the weights constant, but changing the amount of input activation to the model. This can easily be understood by realizing that

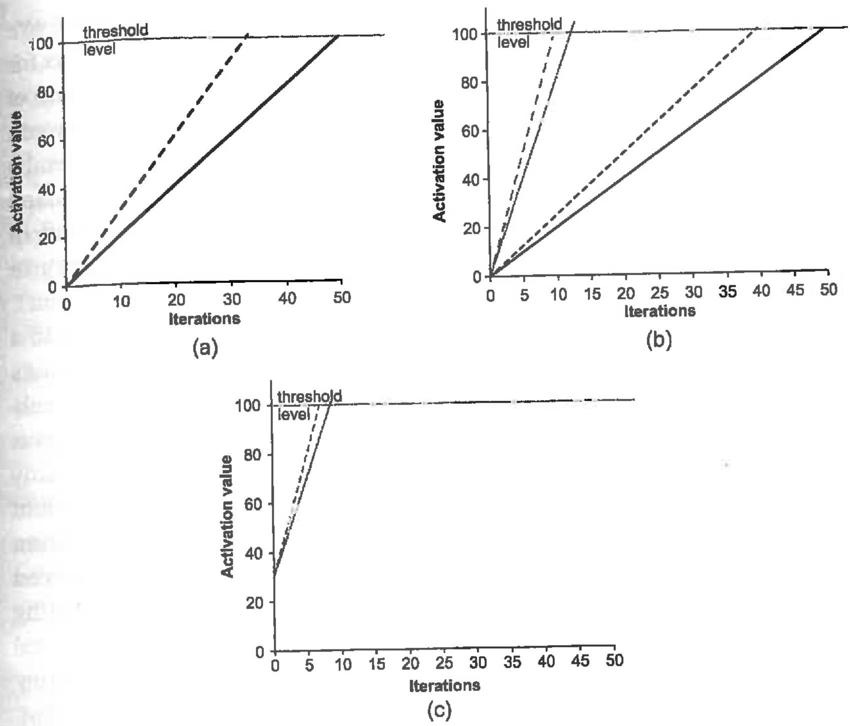

Figure 18.4 Computer simulation results. Panel (a) shows the results of a simulation of the influence of language proficiency and sentence context; panel (b) shows the results of a simulation of a regular cognate effect and the results of a simulation of the influence of repeated picture naming on the size of the cognate effect when it is assumed that repetition only involves the increase of the weights between input nodes and output nodes; panel (c) shows the results of a simulation of the influence of repeated picture naming on the size of the cognate effect when it is assumed that repetition involves both an increase of the weights between input nodes and output nodes and an increase in the resting level of activation of the output nodes. See the text for further explanation

the amount of activation that reaches a node at each time step is the result of a multiplication of weights and activations. To increase or decrease the result of the multiplication, both the strength of the weights or the amount of activation can be adjusted. This feature explains why we can use the same figure to show the results of the simulation of sentence context (to be described next) as we used for the simulation of relative language proficiency in L1 and L2.

In our simulation of the effects of sentence context, we used the same model as in the previous simulation. We first simulated the time course of activation of the output node of a non-cognate for the situation in which pictures were named in isolation. In this simulation, we set the activation of the input node T to 2 activation units. In order to reflect our assumption that adding a sentence context results in the increase of activation in the corresponding language sub-system, we simulated the effect of a sentence context by increasing the amount of activation of the input node T. We (arbitrarily) set the activation of the input node T to 3 activation units for the simulation of picture naming in context. In both simulations, of picture naming with and without sentence context, the weight between the input and output nodes was set to 1 and was kept constant. The results showed that the threshold of 100 activation units was reached in 50 and 34 iterations for the simulation of picture naming in isolation and in context, respectively (Figure 18.4a, solid and dotted line, respectively), an effect of 16 iterations.

In conclusion, the model is able to simulate that picture naming in a sentence context is faster than picture naming in isolation. These results were caused by additional activation in the appropriate language sub-system induced by a sentence context. Note that the size of the effect is linearly related to the amount of input, so the model can simulate any effect size by adjusting the amount of input activation. Therefore, the fit of the model to actual data is trivial, but the numerical demonstration that adding activation to the appropriate language sub-system can indeed cause a decrease of selection time is not. The simulation results concerning both relative language proficiency and sentence context are a natural consequence of the models' processing assumptions.

18.6.3 Simulations of cognate effects and the modulating effect of picture repetition

Our next step was to simulate a cognate effect. To that end we first simulated the activation accrual of a non-cognate output node (the corresponding model is depicted in Figure 18.2b) by setting the activation of the input node T to 2 activation units and the weight between the input node and the output node to 1. Using these parameters, the selection threshold was reached in 50 iterations (Figure 18.4b, black solid line). Next we simulated the activation accrual of a cognate output node (the corresponding model is depicted in Figure 18.2a) by using identical parameters for the T node, but, in addition, we set the activation of the input node nT to 1 and its weight to the output node to 0.5. Using these parameters, the selection threshold was reached in 40 iterations (Figure 18.4b, black dotted line). Thus, these simple models simulated a cognate effect of 10 iterations, demonstrating that additional activation from the non-target language that reaches a phonological output node from the target language can indeed cause a decrease of selection time.

In our third pair of simulations, we simulated the effect of picture-naming repetition on the response times, using the same two models. In order to explain the shorter response times with each repeated presentation of a picture, we assumed earlier in this chapter that the weights of the links connecting lexical nodes to phonological nodes get stronger with each presentation. As an example of this principle, the weights in the present simulations were multiplied by the number of presentations. So, at the fourth presentation, the weights between the T nodes and the output nodes was set to 1 * 4 = 4 and the weight between the nT node and the output node was set to 0.5 * 4 = 2. All other parameters were kept identical to those in the previous cognate simulations. Using these parameters, the selection thresholds for non-cognate and cognate output nodes were reached in 13 (Figure 18.4b, gray solid line) and 10 (Figure 18.4b, gray dotted line) iterations, respectively (effect size = 3). Thus, the simulations show a clear decrease of the size of the cognate effect as a result of picture-naming repetition (from 10 to 3 iterations for the first and fourth presentation, respectively). The models naturally account for this decrease in terms of an increase in the strengths of weights.

In the previous simulation, it was assumed that repetition only affected the strengths of the model's weights. However, we argued earlier in this chapter that it is reasonable to assume that also the resting level activation of the output nodes is affected by repetition. Therefore, in our fourth pair of simulations, we used the exact same parameters as in the previous simulations, but added 10 units of activation to the resting level of the output nodes after each repetition. Therefore, at the start of the fourth repetition, the resting level activation of the output nodes was set to 30 units. Using these parameters, the selection threshold for non-cognates and cognates was reached in 9 (Figure 18.4c, solid line) and 7 iterations (Figure 18.4c, dotted line), respectively (effect size = 2). Thus, when in addition to an increase in the strength of the weights we also assumed an increase in the resting level activation of the output nodes due to repetition, the simulated cognate effect reduced even further, from 3 iterations (see above) to 2 iterations.

18.6.4 Simulations of cognate effects and the modulating effect of relative language proficiency

If it is assumed that proficiency in a language is reflected, among other things, in the strength of the links between the lexical nodes and the sublexical nodes, the results of the simulations of the effect of picture repetition on the size of the cognate effect can also be used to computationally show that an increase in language proficiency causes a decrease of the cognate effect (a finding that has often been reported, see section 18.4.2, which shows larger cognate effects in weaker L2 than in stronger L1). The reason is that we simulated the effect of repetition by increasing the weights in our models. If it is assumed that an increase in proficiency also causes an increase in those weights, these simulations thus also show the influence of relative language proficiency on the size of the cognate effect. In that case, Figure 18.4b should be read as showing the results of the simulation of a bilingual naming pictures in L1 (gray lines) and in L2 (black lines). For both types of line color, the solid lines portray non-cognate output-node activation and the dotted lines portray cognate output-node activation. The cognate effect for naming in L1 is much smaller (3 iterations) than the one for naming in L2 (10 iterations). Note that our simulated models will produce a less severe reduction of the cognate effect if smaller increases of the weights are used than the multiplication by 4 that we chose for this particular simulation.

18.6.5 Implications of the simulations

The results of the simulations of the cognate effects clearly show that a cognate effect of 10 iterations obtained in a simulation of picture naming during the first presentation might reduce to a very small cognate effect of only 3 or 2 iterations in a simulation of picture naming during the fourth presentation. In real experiments, such a small cognate effect might be very hard to detect. Therefore, these simulations suggest that the fact that we failed to find a cognate effect when the pictures were named for the third and fourth time in the stronger language (L1 Dutch) need not indicate that the language system was language-selective at that time. Instead, the simulations show that the size of a cognate effect caused by an intrinsically non-selective bilingual language system might have been too small to be detected.

On a broader level, the simulations suggest that the absence of any marker of language non-selectivity in RT data does not necessarily imply selectivity in the language system. For example, when a marker of language non-selectivity (e.g., a cognate effect) is obtained in L2 whereas it is not obtained in L1, it might be argued that strong connections within the L1 language sub-system and/or strong activation of the L1 language sub-system could have reduced the marker to an undetectable size, despite the fact that the bilingual language system in itself is fundamentally language non-selective.

18.7 Conclusion

In an early study on language switching in bilinguals, Macnamara, Krauthammer, and Bolgar (1968) likened the bilingual's linguistic performance to that of a musician "who observes the notation for key at the beginning of a piece of music and then forgets about it though in his playing he performs the actions appropriate to the key." They continued noting that: "Similarly, the bilingual once started in one language can forget about which language he is speaking and yet obey the rules of that language" (p. 213). These authors thus suggest that initially choosing the language to use requires effort but that subsequently staying in the chosen language does not; in other words, that the non-selected language does not interfere with the language in use. The empirical data and the computer simulations presented in this chapter suggest that this observation at least holds true for fluent bilinguals involved in authentic monolingual discourse. We have seen that, even in experiments that do not truthfully mimic real discourse, the influence of the non-target language is modest if the participants' stronger language serves as the target language. Furthermore, it was found that sentence context and stimulus repetition mitigate the influence of the other language. In veridical discourse word production is not only supported by sentence context but also by the wider linguistic context and by extra-linguistic contextual information. In addition, in realistic discourse specific words (those central to the discourse's topic) are likely to be repeated. These ingredients of rich contextual information, repeated word use, and speaker fluency may combine into a solid firewall against an influence of the other language in discourse. Importantly though, this conclusion does not imply that under these circumstances the language not in use is deactivated, because our computer simulations have shown it need not be. These simulations furthermore indicated that immunity from activation in the language not in use likely emerges from simple structural and processing features of the bilingual word-production system, such as (temporarily) strong connections between specific representations in the word-production system and a high level of activation in these representations.