Journal of Experimental Psychology:
Human Perception and Performance
2004, Vol. 30, No. 5, 889–906

0096-1523/04/$12.00 DOI: 10.1037/0096-1523.30.5.889

Forward and Backward Number Translation Requires Conceptual Mediation in Both Balanced and Unbalanced Bilinguals

Wouter Duyck and Marc Brysbaert
Ghent University

It is much debated whether translation is semantically mediated or based on word–word associations at the lexical level. In 2 experiments with Dutch (L1)–French (L2) bilinguals, the authors showed that there is a semantic number magnitude effect in both forward and backward translation of number words: It takes longer to translate number words representing large quantities (e.g., *acht, huit* [eight]) than small quantities (e.g., *twee, deux* [two]). In a 3rd experiment, the authors replicated these effects with number words that had been acquired only just before the translation task. Finally, it was shown that the findings were not due to the restricted semantic context of the stimuli. These findings strongly suggest that translation processes can be semantically mediated in both directions, even at low levels of L2 proficiency.

Bilingual people are able not only to understand messages in two different languages but also to translate information from their first language (L1) to their second (L2) (forward translation) and vice versa (backward translation). An important question with respect to this ability is whether word translation relies on direct word–word associations in the lexicon or whether it requires activation of the meaning of the words. The aim of the present study was threefold: (a) to replicate semantic effects during forward translation, which almost all translation studies have reported (e.g., Cheung & Chen, 1998; Kroll & Stewart, 1994; La Heij, Hooglander, Kerling, & Vandervelden, 1996; Sánchez-Casas, Davis, & Garcia-Albea, 1992; Sholl, Sankaranarayanan, & Kroll, 1995), (b) to search for indications of semantic mediation in backward translation, and (c) to investigate whether the nature of the translation process depends on additional factors, such as L2 proficiency and semantic context. We report four experiments in which we addressed these questions by looking for a semantic number magnitude effect in a number-naming and translation task. The dominant view in the literature on this issue is provided by Kroll and Stewart's revised hierarchical model (RHM) of bilingual memory (Kroll & de Groot, 1997; Kroll & Stewart, 1994), which we describe in some detail below.

Conceptual Mediation in Word Translation

In a series of articles, Kroll and colleagues (Kroll & de Groot, 1997; Kroll & Stewart, 1994) developed a model of bilingual memory (see Figure 1) that can explain a broad range of findings (for reviews, see Kroll, 1993; Kroll & de Groot, 1997). First, in the model a distinction is made between lexical representations (representing word forms) and semantic representations (representing word meanings; for a recent review of evidence concerning this assumption, see Smith, 1997). Second, it is assumed that the conceptual representations are shared among the languages, whereas the lexical representations are language specific. So, there are two lexicons (one for L1 and one for L2) connected to a unitary semantic system. Finally, the connections between the different parts in the model have two interesting features: They are asymmetric, and their importance changes in the course of second language acquisition (the developmental hypothesis).

The asymmetry hypothesis states that the links between the L1 lexicon and the semantic system are stronger than the links between the L2 lexicon and the semantic system, because word meanings are more strongly activated by L1 words than by L2 words. In contrast, the direct word–word connections at the lexical level are stronger from L2 to L1 than the other way around. The reason for this is that L2 words are often learned by associating them with their L1 translations. Because of the asymmetries in the connections, forward translation is more likely to engage conceptual activation than backward translation. Backward translation in turn depends more on direct lexical connections. Support for this asymmetry assumption was reported by Sholl et al. (1995). On forward translation trials, they found facilitation when the concepts had been primed by the presentation of pictures in a first phase of the experiment. The priming effect was not present in the backward translation condition, suggesting a less conceptually mediated translation process. The asymmetries of the connections are believed to decline as L2 proficiency increases. This is the developmental aspect of the model. Support for this hypothesis was reported by Talamas, Kroll, and Dufour (1999), who found greater interference of semantically related false translations in a translation recognition task when the participants were highly proficient in L2, whereas less proficient bilinguals suffered more interference from form-related words. For a more detailed review of the findings supporting the different assumptions of the RHM, see Kroll

Wouter Duyck and Marc Brysbaert, Department of Experimental Psychology, Ghent University, Ghent, Belgium.

Marc Brysbaert is now at the Department of Psychology, Royal Holloway University of London, United Kingdom.

This research was made possible by the Fund for Scientific Research—Flanders, of which Wouter Duyck is a research assistant. We are indebted to Judith Kroll and Michael Thomas for excellent comments on an earlier version of this article. We also thank Jolien De Brauwer for successfully carrying out a replication of Experiment 3.

Correspondence concerning this article should be addressed to Wouter Duyck, Department of Experimental Psychology, Ghent University, Henri Dunantlaan 2, B-9000 Ghent, Belgium. E-mail: wouter.duyck@ugent.be

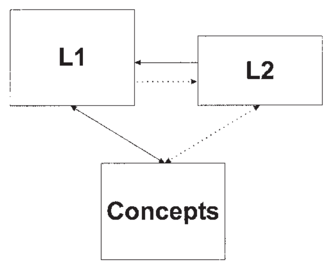

Figure 1. An illustration of Kroll and de Groot's (1997) revised hierarchical model of bilingual memory. Solid lines represent stronger links than dotted lines. L1 = first language; L2 = second language.

and de Groot (1997; see also Cheung & Chen, 1998; Kroll, Michael, Tokowicz, & Dufour, 2002; Kroll & Stewart, 1994; Sánchez-Casas et al., 1992).

In the experiments described below, we tested to what extent these assumptions and following predictions of the RHM hold in a number-word translation task. Before presenting these experiments, we briefly review what is currently known about number processing (for a more detailed review, see Brysbaert, in press).

Semantic Activation in Number Processing

Numbers provide a very appealing set of stimuli to study translation processes, as bilinguals have three sets of symbols to represent the same concept: Arabic digits (e.g., *3*), L1 number words (e.g., *drie* in Dutch), and L2 number words (e.g., *trois* in French). This makes it possible not only to study translation from L1 to L2 and from L2 to L1 but also from a common logographic symbol to either L1 or L2. Moreover, as we show below, activation of the underlying conceptual information depends on the symbol format and the nature of the task. In addition, Arabic digits do not take much longer to name than number words, contrary to the naming of object pictures (Ferrand, 1999).

The meaning of a number primarily refers to the magnitude represented by the number. These magnitudes can be understood quite well with the metaphor of a number line (e.g., Brysbaert, 1995; Dehaene, 1992). All integers (from 1 to at least 15) form an ordered continuum oriented from left (small) to right (large). So, when the magnitudes of two numbers have to be compared, this is easier when the distance between the numbers is large (e.g., to indicate the larger number of the pair 2–8) than when it is small (e.g., to indicate the larger number of the pair 7–8; Moyer & Landauer, 1967). Similarly, the processing of a number is primed when immediately before a number with a close magnitude has been presented than when a number with a more distant magnitude has been presented (Reynvoet & Brysbaert, 1999). In addition, participants react faster with their left hand to small numbers and with their right hand to large numbers than vice versa (the so-called *spatial–numerical association of response codes* [SNARC] effect; e.g., Dehaene, Bossini, & Giraux, 1993).

Further evidence indicates that the magnitude information is activated more rapidly for small numbers than for large numbers. So, it is easier to select the larger digit of the number pair 2–3 than of the number pair 7–8 (Moyer & Landauer, 1967). Using eye movement registrations, Brysbaert (1995, Experiment 1) found that this effect is due not entirely to the comparison process but also to the encoding speed of the numbers. In this experiment, participants had to read three Arabic numerals going from 0 to 99 and decide whether the middle number fell in between the two outer numbers (e.g., *23 41 65*) or not (e.g., *23 65 41*). The most important variable to predict the reading time of the first numeral turned out to be the (logarithm of the) number magnitude. More important, in a subsequent experiment (Brysbaert, 1995, Experiment 2), the same magnitude effect was replicated when participants simply had to read the three numerals and indicate whether a fourth, new numeral was one of the first three numbers. Because the fourth numeral could not be seen until all three numerals of the initial set had been read, this effect must have been due not to the comparison process but rather to the encoding of the numbers.

There are several lines of evidence indicating that the number magnitude effect is semantic in origin (i.e., related to the meaning of the numbers) and not due to associations between lexical representations. For a start, the distance-related priming effect is the same within notations as across notations. Thus, the target word *six* is primed as much by the prime *5* as by the prime *five* (Reynvoet, Brysbaert, & Fias, 2002). Similarly, the same distance effect in number comparison is found with digits, number words, and even random sets of dots as stimulus materials (e.g., Buckley & Gillman, 1974; Dehaene, Dupoux, & Mehler, 1990; Foltz, Poltrock, & Potts, 1984). Furthermore, the quantity priming effect is symmetrical (e.g., the target *6* is primed equally well by the primes *5* and *7*) and not asymmetric (with stronger priming in the forward prime-target direction), as an associative hypothesis would predict (e.g., Duyck & Brysbaert, 2002; Koechlin, Naccache, Block, & Dehaene, 1999). Also, Dehaene and Akhavein (1995) observed effects of numerical distance when the lexical distance between items (according to theories of number processing) was kept constant. Finally, brain imaging studies have shown that in all number comparison tasks, irrespective of the input format, a region in the parietal cortex is active. The brain area is not active in nonnumerical word processing tasks but is active in other analog magnitude estimation tasks (Fias, Lammertyn, Reynvoet, Dupont, & Orban, 2003; Naccache & Dehaene, 2001; Pesenti, Thioux, Seron, & De Volder, 2000; Pinel, Dehaene, Riviere, & Lebihan, 2001). For further evidence for the semantic origin of these effects, see Koechlin et al. (1999).

More important for the present study, magnitude information is not required for all number-processing tasks. There is quite some evidence that the processing of number words is only semantically mediated if the experimental task requires the activation of certain semantic information. For instance, Fias, Reynvoet, and Brysbaert (2001; see also Fias, 2001) showed that the word *five* was read equally fast when it was displayed together with the digit *3* as when it was presented together with the digit *5*. In contrast, responses to the word *five* were faster when the word was presented together with the digit *5* than when it was presented with the digit *3* in a parity judgment (odd–even) task. The finding that number words can be named without semantic mediation is in line with most models of visual word recognition, which assume the existence of nonsemantic routes for word naming. The situation is less clear for Arabic input, with authors claiming that digits can be named without semantic mediation (e.g., Campbell, 1994; Cipolotti & Butterworth, 1995; Dehaene, 1992), whereas others reject this possibility (e.g., Brysbaert, 1995; Fias, 2001; McCloskey, 1992). Both groups of authors agree, however, that number magnitude is activated more rapidly from Arabic input than from verbal input, as can be concluded from the finding that participants find it more difficult to select the physically larger number in the number pair 2–8 than in the number pair 2–8, whereas no such size congruency effect is observed in the pairs two–eight versus two–eight (e.g., Ito & Hatta, 2003).

Experiment 1

From the previous section, it will be clear that number translation provides an interesting new paradigm to test Kroll and Stewart's (1994) RHM. Given that word naming does not explicitly require access to semantic information, L1 and L2 naming of, respectively, L1 and L2 number words is not expected to show a magnitude effect. The Dutch word *twee* (two) is not expected to be named faster than the word *acht* (eight). The same applies to the corresponding French number words *deux* and *huit.* It is less clear whether the naming of Arabic numbers will involve a magnitude effect: On the basis of rapid activation of magnitude information from Arabic input, one might expect to find such an effect both in L1 and in L2 naming of digits. So, the naming of the Arabic digit *2* in Dutch and French could be faster than the naming of *8.* Most important for the present study, however, the presence or absence of a semantic magnitude effect in the translation of number words allows us to directly test the asymmetry hypothesis of the RHM. Because number words have been shown to activate their underlying semantic information in certain (semantic) tasks and because forward translation implies conceptual mediation, a magnitude effect should be found when Dutch–French bilinguals translate Dutch (L1) number words into French (L2). Hence, the word *twee* (two) should be translated faster into French than the number word *acht* (eight). In contrast, no semantic number magnitude effect is expected for these bilinguals when French (L2) number words have to be translated into Dutch (L1), as this task is more likely to occur through word–word associations at the lexical level and therefore would not require access to semantic representations. So, for Dutch–French bilinguals the French number word *deux* (two) should be translated into Dutch as fast as *huit* (eight), except maybe for very proficient bilinguals (see the developmental hypothesis of the RHM).

To test these predictions of the RHM, we designed an experiment in which Arabic numbers and both L1 (Dutch) and L2 (French) number words had to be named in both L1 and L2. We also manipulated L2 proficiency to check for interactions with possible number magnitude effects. In addition, a short delayed naming task was administered after the actual experiment. In this task, participants were asked to delay responses for more than a second, so that semantic processing of the stimulus was finished before the response had to be given. This allows us to control for differences in voice key sensitivity to the response onset and other theoretically irrelevant variables that could confound the number magnitude effect that is of interest in this study.

Method

Participants

Twenty-two first-year university students participated for course requirements. All of them were native Dutch speakers and mainly used this language in everyday life. Eleven of them had started learning French at school between 10 and 13 years of age. We refer to this group as *unbalanced* bilinguals. They did not study their L2 at university, had no L2 speaking relatives, and did not practice their L2 on any other regular basis. The other 11 students had been raised in a Dutch–French bilingual setting from birth and were practically equally proficient in both languages (but indicated Dutch as L1). We refer to this group as *balanced* bilinguals.

Materials

All stimuli were presented on a standard 15-in. (38.1-cm) VGA color monitor as yellow characters on a black background. Stimulus presentation was computer driven by a PC equipped with a voice key, which was connected through the game port. Arabic digits and Dutch and French number words representing quantities from 1 to 12 served as stimuli.

Design

The experiment had a 2 (L2 proficiency: unbalanced vs. balanced) $\times$ 2 (naming language: L1 vs. L2) $\times$ 3 (stimulus format: Arabic numbers, L1 number words, and L2 number words) $\times$ 12 (number magnitude) full factorial design. Except for L2 proficiency, all variables were manipulated within subject.

Procedure

All participants completed two blocks (L1 naming and L2 naming) of 360 trials. The order of these blocks was counterbalanced with L2 proficiency. Within each block, 10 series of 36 randomly ordered trials were presented, corresponding to every number magnitude from 1 to 12 in each of the three stimulus formats (Arabic, L1, and L2). Hence, the participants did not know which stimulus format would appear before the beginning of each trial. Only naming language was blocked. Each trial started with the presentation of a fixation stimulus (an *asterisk*; the *plus sign* was not used as a fixation point because of its mathematical meaning) for 500 ms. After that time, the stimulus was replaced by the target, which remained visible until pronunciation of the target triggered the voice key. The intertrial interval (ITI) was 1,000 ms. The experiment lasted for about 50 min, including a short break.

As noted earlier, we also ran a delayed number-word naming task after the actual experiment to control for theoretically irrelevant variables such as voice key onset sensitivity. In this task, 12 participants (unbalanced bilinguals) were asked to delay responses for more than a second. All participants completed eight blocks of 24 trials. Within each block, all Dutch and French number words representing magnitudes from 1 to 12 were presented in a random order. In each trial, the target was presented centered on the screen for 1,000 ms, followed by a black screen for another 1,000 ms. Then, a question mark was presented, indicating that the participant had to name the target word just seen as fast as possible. The ITI was 1,000 ms. As soon as there was doubt concerning the accuracy of time registration (e.g., due to irrelevant noise), the trial was excluded from the data.

Results

Variance Analysis

The proportion of invalid trials in the immediate naming experiment due to naming errors or faulty time registration was 6.1%. These trials were excluded from all analyses. An analysis of variance (ANOVA) was performed with L2 proficiency as a between-subjects variable and naming language, stimulus format, and number magnitude as repeated measures factors. The dependent variable was mean response time (RT) across correct trials. Mean RTs for both L2 proficiency groups as a function of naming language, stimulus format, and number magnitude are presented in Figure 2. The backward translation condition can be found in the left part of the figure; forward translation is plotted in the right part.

The main effect of L2 proficiency reached significance, $F(1, 20) = 3.26$, $MSE = 160{,}709$, $p < .05$ (one-tailed). Mean RTs for balanced and unbalanced bilinguals were 510 ms and 546 ms, respectively. As the two, almost identical graphs in Figure 2 suggest, proficiency did not interact with any other factor in the design. The effect of naming language was not significant ($F < 1$). Naming in Dutch took 529 ms, whereas French naming took 527 ms. Unbalanced bilinguals were slightly slower for French (L2) naming ($M = 548$ ms) than for Dutch (L1) naming ($M = 544$ ms), but this difference was not significant ($F < 1$). Balanced bilinguals showed a tendency toward the reverse pattern (L2 naming [$M = 506$ ms] was slightly faster than L1 naming [$M = 513$ ms]), but again this difference was not significant ($F < 1$). A post hoc comparison using Tukey's honestly significant difference (HSD) test showed that backward translation was significantly slower ($M = 592$ ms) than forward translation ($M = 555$ ms; $p < .01$), as

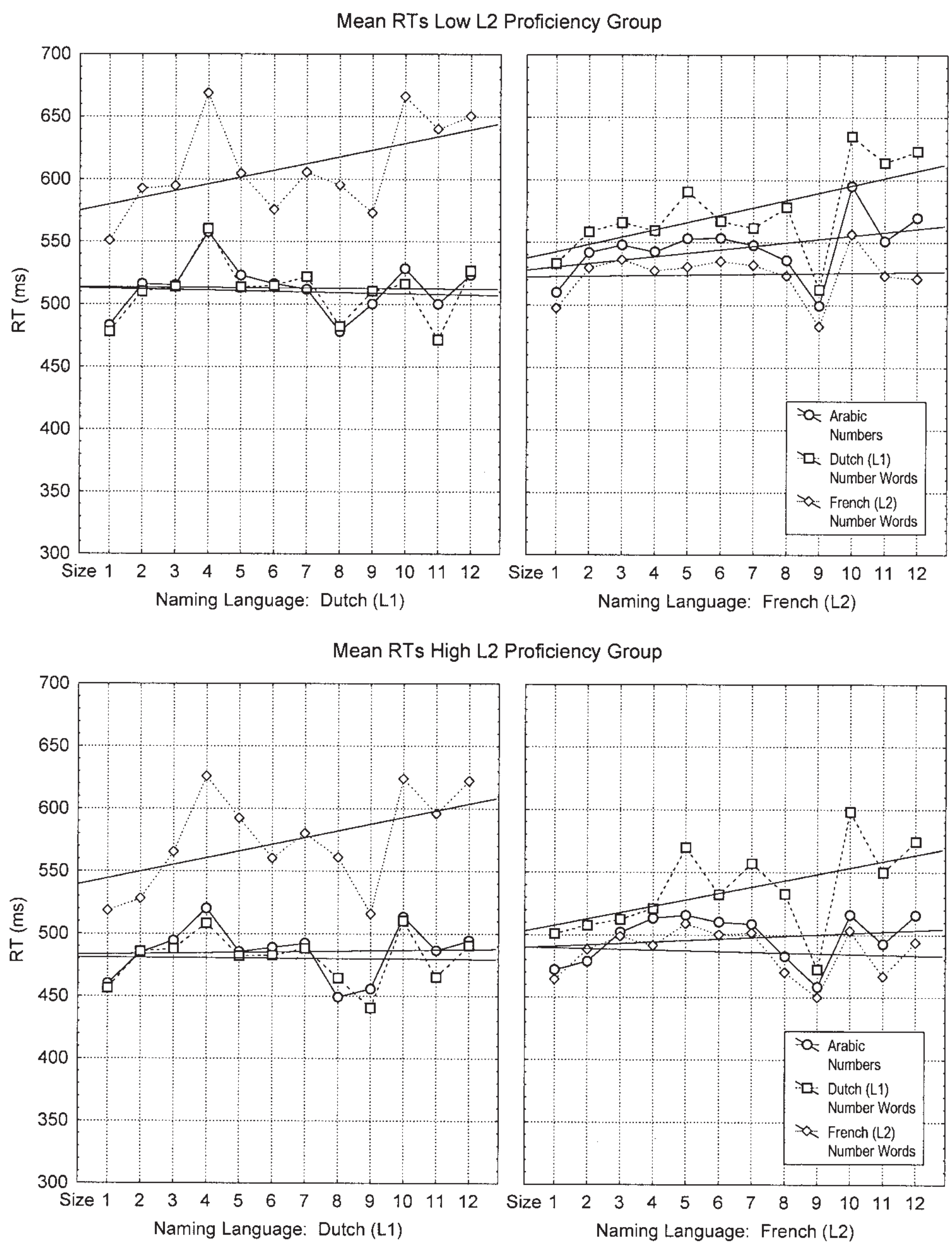

Figure 2. Mean naming response times (RTs) by naming language, stimulus format, and number magnitude (Experiment 1), plotted separately for both proficiency groups. Straight lines represent best linear fit according to a least-squares criterion.

Table 1
The Regression Equations for the Six Naming Language × Stimulus Format Conditions (Experiment 1) According to the Procedure Described by Lorch and Myers (1990)

Naming language and stimulus format	Intercept	Number magnitude	Delayed naming RT	Frequency
L1 (Dutch)				
Arabic numbers	RT = 220	+0.22 NM	+0.70 D**	+0.03 F
L1 number words (Dutch)	RT = 183	+0.00 NM	+0.77 D**	+0.05 F
L2 number words (French)	RT = 80	+5.08 NM**	+1.26 D**	−0.01 F
L2 (French)				
Arabic numbers	RT = 231	+2.71 NM**	+0.75 D**	−0.04 F
L1 number words (Dutch)	RT = 215	+6.94 NM**	+0.73 D**	+0.04 F
L2 number words (French)	RT = 235	+0.71 NM	+0.72 D**	−0.02 F

Note. RT = response time.
** $p < .01$.

opposed to the predictions based on the RHM. This difference was significant for both unbalanced ($M = 610$ ms vs. $M = 575$ ms) and balanced bilinguals ($M = 574$ ms vs. $M = 536$ ms; $ps < .01$).

The main effects of stimulus format, $F(2, 40) = 76.12$, $MSE = 2{,}596$, $p < .01$, and number magnitude, $F(11, 220) = 27.55$, $MSE = 2{,}352$, $p < .01$, were significant, but these effects were embedded in an important three-way interaction with naming language, $F(22, 440) = 11.86$, $MSE = 705.6$, $p < .01$. Indeed, as can be seen in Figure 2, the effect of number magnitude appears to be present in only some of the Stimulus Format × Naming Language conditions. These effects of number magnitude are investigated in more detail by means of regression analyses in the next section.

Regression Analysis

To assess the importance of number magnitude independent of number frequency[1] and the delay to activate the voice key, we performed regression analyses according to the procedure for repeated measures data described by Lorch and Myers (1990, *Method 3,* p. 151), with number magnitude, number-word frequency, and mean delayed naming RTs as predictors.

The regression weights for the six conditions (i.e., 2 [naming language: L1 vs. L2] × 3 [stimulus format: Arabic, L1, and L2]) are displayed in Table 1. Most important, the regression weights of number magnitude differed significantly from zero in both the forward and the backward translation conditions, respectively, $t(21) = 5.135$, $p < .01$ (p values for two-tailed testing), and $t(21) = 7.940$, $p < .01$ (for a detailed statistical explanation of the computational procedure of these tests, see Lorch & Myers, 1990). These regression weights did not differ significantly from each other, $t(21) = 1.322$, $p > .20$. Hence, conceptual mediation was not larger in forward than in backward translation.

The size of the number magnitude effect did not differ between unbalanced and balanced bilinguals. This was the case for both forward and backward translation ($ts < 1$). To further increase the power of our analysis, we also tested whether the size of the magnitude effect correlated with the difference in mean RTs between L1 and L2 naming (as a measure of L2 proficiency).[2] Consistent with the previous results, this correlation was very weak and not significant for either direction of translation (backward translation: $r = -.09$, $p > .69$; forward translation: $r = -.13$, $p > .57$). Thus, the magnitude effect in number-word translation did not interact with L2 proficiency.

The number magnitude effect was not significant for within-language number-word naming (L1 [Dutch]: $t < 1$; L2 [French]: $t(21) = 1.218$, $p > .23$) or for L1 (Dutch) naming of Arabic digits ($t < 1$). In contrast, the regression weight of number magnitude differed significantly from zero for the L2 (French) naming of Arabic digits, $t(21) = 3.631$, $p < .01$. The regression weights of frequency never reached significance.

All regression weights of the delayed naming RTs were significant. This confirms that the RTs were indeed influenced by sensitivity differences in the triggering of the voice key by the different number names. Related studies in speech production have also acknowledged this problem. Jescheniak and Levelt (1994), for example, dealt with it by subtracting delayed response RTs from nondelayed RTs and using the resulting values as the dependent variable in the analysis. If this approach is followed (see Figure 3), the number magnitude effects observed in forward and backward translation remain clearly visible when this dependent variable is used. Indeed, regression analyses with this dependent variable yielded virtually identical results to the ones mentioned above.

Discussion

The results of Experiment 1 are quite clear. As predicted by the RHM (Kroll & Stewart, 1994), we obtained a reliable effect of number magnitude on forward translation. It took longer to trans-

[1] We wanted to control for frequency effects because Gielen, Brysbaert, and Dhondt (1991) found a significant correlation between number magnitude and number frequency ($r = -.621$, $p < .01$). So, because smaller numbers are more frequent and are thus processed faster, it is possible that any effect of number magnitude in the data is confounded by effects of number frequency. Therefore, we included the number-word frequency measures as reported by Gielen et al. in our analysis. Note that Dehaene and Mehler (1992) showed that the frequencies of numbers are very similar in different languages.

[2] We subtracted mean RTs for the Size 2 condition from mean RTs for Size 8 as a measure of the number magnitude effect. Note that different measures (e.g., the regression weights of number magnitude in the regression analyses) led to very similar results. We thank Michael Thomas for this suggestion.

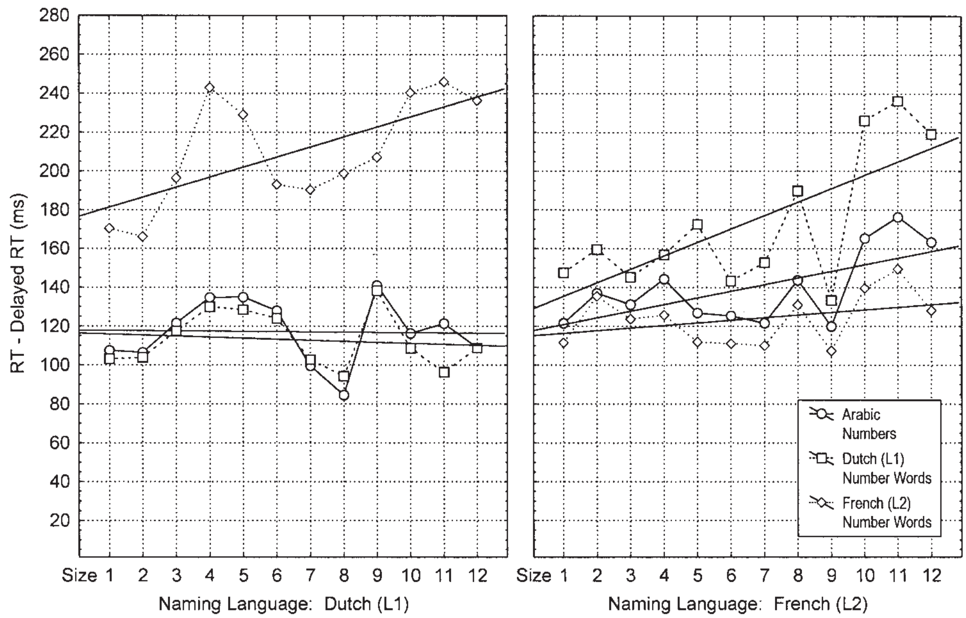

Figure 3. Mean naming response times (RTs) minus delayed naming RTs by naming language, stimulus format, and number magnitude (Experiment 1). Straight lines represent best linear fit according to a least-squares criterion.

late L1 number words representing large quantities (e.g., *acht* [eight]) than number words representing small quantities (e.g., *twee* [two]). This strongly suggests conceptual mediation, because magnitude information is not stored at the lexical level. However, contrary to the predictions of the RHM, such an effect was also obtained in backward translation: Translation was slower for large L2 number words (e.g., *huit* [eight]) than for small L2 number words (e.g., *deux* [two]). Moreover, the effect was equally strong for both directions of translation and did not interact with L2 proficiency. Hence, it seems that translation was conceptually mediated in both directions to the same degree for both balanced and less proficient bilinguals. Another finding that is not consistent with the predictions of the model concerns the speed of the translation processes: Backward translation was found to be significantly slower than forward translation. However, as is discussed later, this finding may be due to the mixed nature of the stimulus lists in this experiment.

An account of the observed number magnitude effects in terms of the correlated predictor word frequency (which has been shown to influence translation; e.g., de Groot, 1992; see also Footnote 1) cannot easily explain why the effect emerges only in the translation conditions and not in the naming conditions. Indeed, the regression analyses confirmed that the magnitude effects found were not due to effects of number-word frequency. Frequency did not have an effect in any of the conditions. As the frequency effect is usually situated at the lexical level, this is further indirect evidence that the translations were not based on direct word–word associations. A more detailed theoretical discussion of these results appears in the General Discussion.

As expected, we did not find a semantic effect for number-word naming, contrary to number-word translation, as this is not a semantic task. In contrast, we did find a magnitude effect in L2 naming of Arabic numbers. This is not surprising, as processing of Arabic numbers seems to trigger fast conceptual activation. However, we did not obtain an effect for L1 naming of Arabic numbers, contrary to Brysbaert (1995). The cause of the absence of such an effect could lie within the restricted range of digits used: Whereas Brysbaert presented numbers from 0 to 99, we used only digits from 1 to 12. An inspection of Brysbaert's data shows that the logarithmic number magnitude effect was not very clear for numbers smaller than 10. We believe this finding only adds further importance to the effects found in translation trials. Finally, the effects of the delayed naming predictor showed the importance of controlling for the speed with which a voice key is activated by different sequences of sounds.

Experiment 2

Although the results described above seem quite straightforward, one might object that the random presentation procedure used in Experiment 1 rendered the input stimulus format (but not the output language) unpredictable. It is possible that this introduced switching costs between trials (e.g., when an L1 target is followed by an L2 target). For example, Meuter and Allport (1999) showed in an Arabic number-naming task that switching naming language resulted in a time cost on the trial following the switch. The cost was larger when switching from L2 to L1 than in the reverse condition. Similar switching costs may have occurred in Experiment 1, even though there are theoretical reasons to believe that it is unlikely that such costs have influenced the obtained magnitude effects (see the *Discussion* section). This alternative switching cost account cannot be ruled out by a reanalysis of our data, as inclusion of the previous trial stimulus format as a variable leads to very few observations per cell (on average 3.33, if all trials were correct). Moreover, the random algorithm that steered stimulus presentation did not guarantee a minimum of one observation per cell. Therefore, we decided to repeat Experiment 1 with a procedure that blocked not only naming language but also stimulus format.

Method

Participants

Twelve first-year university students participated for course credit. They all belonged to the group of bilinguals labeled before as unbalanced.

Materials

All stimulus materials were identical to those in Experiment 1. The software used was adapted to make stimulus presentation blocked by stimulus format.

Design

The experiment had a 2 (naming language: L1 vs. L2) × 3 (stimulus format: Arabic numbers, L1 number words, and L2 number words) × 12 (number magnitude) full factorial design. All variables were manipulated within subjects. L2 proficiency was not manipulated because it did not interact with any variable in Experiment 1.

Procedure

All participants completed two series of three blocks, each consisting of 120 trials. Naming language was Dutch (L1) in one series and French (L2) in the other. The three blocks within a series corresponded to the three stimulus formats (Arabic numbers, L1 number words, and L2 number words). Within each block, all numbers from 1 to 12 were presented 10 times. Order of trials within blocks, blocks within series, and series was determined at random. The procedure for a trial was identical to that used in Experiment 1.

Results

Variance Analysis

The proportion of invalid trials due to naming errors or faulty time registration was 7.0%. These trials were excluded from all analyses. An ANOVA was performed with naming language, stimulus format, and number magnitude as repeated measures factors. The dependent variable was the mean RT across correct trials. Mean RTs as a function of naming language, stimulus format, and number magnitude are presented in Figure 4. The backward translation condition can be found in the upper left part of the figure; forward translation corresponds to the upper graph in the right part.

The effect of naming language was significant, $F(1, 11) = 7.71$, $MSE = 26{,}484$, $p < .02$. Respective means for Dutch and French naming were 516 and 547 ms. Forward translation ($M = 595$ ms) tended to be faster than backward translation ($M = 610$ ms), but Tukey's HSD test showed that this difference was not significant ($p < .69$). The main effects of stimulus format, $F(2, 22) = 25.86$, $MSE = 7{,}212$, $p < .01$, and number magnitude, $F(11, 121) = 24.91$, $MSE = 1{,}568$, $p < .01$, were significant, but these effects were embedded in a three-way interaction with naming language, $F(22, 242) = 10.67$, $MSE = 726.8$, $p < .01$. As can be seen in Figure 4, and similar to Experiment 1, the effect of number magnitude was present only in some of the Stimulus Format × Naming Language conditions. These effects are analyzed in detail in the following regression analyses.

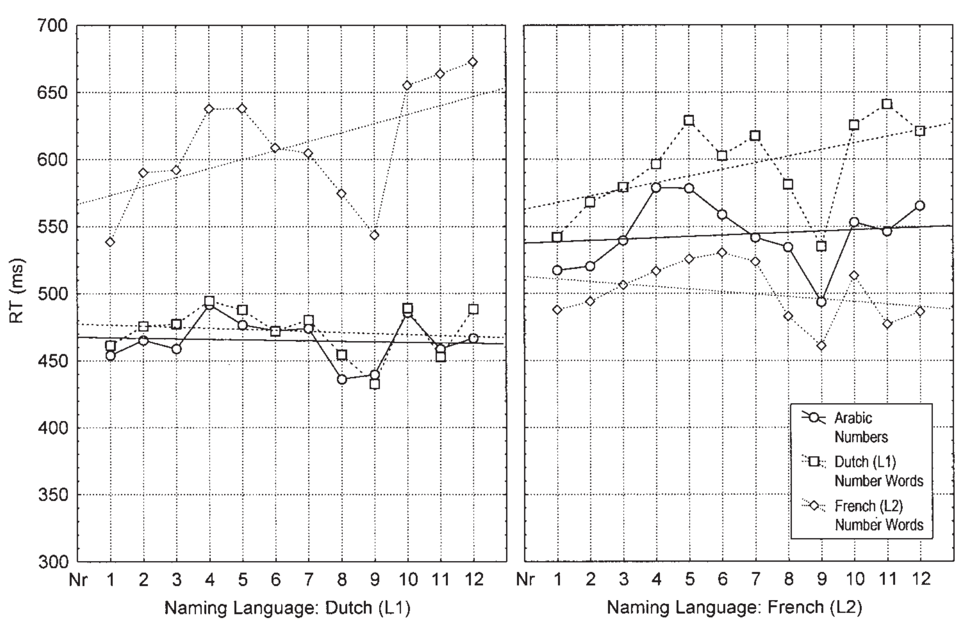

Figure 4. Mean naming response times (RTs) by naming language, stimulus format, and number magnitude (Experiment 2, blocked presentation). Straight lines represent best linear fit according to a least-squares criterion. Nr = number.

Regression Analysis

The regression analyses were again performed by the procedure for repeated measures data described by Lorch and Myers (1990, Method 3). Regression weights for the six Naming Language (L1 vs. L2) × Stimulus Format (Arabic, L1, and L2) conditions are displayed in Table 2. Similar to Experiment 1, the regression weights of number magnitude differed significantly from zero in both forward and backward translation conditions, respectively, $t(11) = 2.718$, $p < .02$, and $t(11) = 4.949$, $p < .01$. These regression weights did not differ significantly from each other ($t < 1$), although the effect of number magnitude tended to be somewhat larger in the backward translation condition. Note that an effect of frequency was found in the forward translation condition, whereas this effect was not present in backward translation.

A comparison of Tables 1 and 2 strongly suggests that the blocking of the stimuli made no difference for the number magnitude effect. This was confirmed by a statistical analysis (Lorch & Myers, 1990). There was no difference at all between the number magnitude regression weights for backward translation (Experiment 1: $B = 5.08$; Experiment 2: $B = 5.18$; $t < 1$). For forward translation, the regression weights were slightly higher in Experiment 1 (Experiment 1: $B = 6.94$; Experiment 2: $B = 3.88$), but this difference was not significant, $t(32) = 1.449$, $p > .15$. Also, the blocked presentation procedure had no effect on overall mean RTs (Experiment 1: $M = 528$ ms; Experiment 2: $M = 531$ ms; $F < 1$). If RTs from only the unbalanced bilinguals who participated in Experiment 1 were taken into account, mean RTs for Experiment 2 (only unbalanced participants) were slightly faster, but this difference again was far from significant (Experiment 1: $M = 546$ ms; Experiment 2: $M = 531$ ms; $F < 1$).

Discussion

The present experiment shows that the findings of Experiment 1 were not caused by switching costs due to the random presentation of different stimulus formats (e.g., Meuter & Allport, 1999). When stimulus format was blocked, exactly the same effects of number magnitude were found in both forward and backward translation. It took longer to translate L1 and L2 number words representing larger quantities than small quantities. The fact that mixed and blocked stimulus presentation yielded the same results is not inconsistent with the literature on switching costs, as previous studies on bilingualism have reported switching costs when the output language changed (e.g., Meuter & Allport, 1999) but not when the input language changed (e.g., Thomas & Allport, 2000). In addition, there is quite some evidence that the initial stages of visual word recognition in bilinguals are not language specific and consequently not subject to switching costs (Altenberg & Cairns, 1983; Brysbaert, Van Dyck, & Van de Poel, 1999; Dijkstra, Grainger, & Van Heuven, 1999; Dijkstra, Timmermans, & Schriefers, 2000; Dijkstra & Van Heuven, 2002; Duyck, Diependaele, Drieghe, & Brysbaert, 2004; Nas, 1983; Van Wijnendaele & Brysbaert, 2002; Von Studnitz & Green, 1997). More important, the fact that equivalent number magnitude effects were found for forward and backward translation in two different studies with different stimulus presentations adds credit to our claim that number translation is conceptually mediated in both translation directions. We go into further details about the theoretical implications of this finding in the General Discussion. Finally, it is important to note that forward translation was not significantly faster than backward translation. As hypothesized earlier, this suggests that the occurrence of such an effect in Experiment 1 is probably due to the mixed nature of the stimulus lists in that experiment.

Experiment 3

In the previous experiments, we found evidence for an equivalent number magnitude effect in forward and backward number translation, for both unbalanced and balanced bilinguals. As these findings are not entirely compatible with the predictions of the RHM (i.e., stronger semantic mediation in forward translation than in backward translation; more asymmetry at low levels of L2 proficiency), we decided to explore more in depth the limits of our findings and to eliminate some possible confounds in our stimulus materials.

A criticism of Experiments 1 and 2 might be that the unbalanced bilinguals were already too proficient. Indeed, mean RTs were reasonably fast for a translation task. This was probably partly due to the fact that the number words are among the first acquired and most frequent L2 words. Also, for political reasons, Belgian high school students receive rather extensive L2 teaching. Hence, it might be that the number magnitude effect in backward translation

Table 2
The Regression Equations for the Six Naming Language × Stimulus Format Conditions (Experiment 2) According to the Procedure Described by Lorch and Myers (1990)

Naming language and stimulus format	Intercept	Number magnitude	Delayed naming RT	Frequency
L1 (Dutch)				
Arabic numbers	RT = 217	+0.12 NM	+0.58 D**	+0.07 F
L1 number words (Dutch)	RT = 171	−0.43 NM	+0.74 D**	+0.06 F
L2 number words (French)	RT = 164	+5.18 NM**	+1.22 D**	−0.14 F
L2 (French)				
Arabic numbers	RT = 322	+0.02 NM	+0.81 D**	−0.23 F*
L1 number words (Dutch)	RT = 330	+3.88 NM*	+0.88 D**	−0.25 F*
L2 number words (French)	RT = 155	−1.10 NM	+1.00 D**	−0.07 F

Note. RT = response time.
* $p < .05$. ** $p < .01$.

was a manifestation of a very high L2 proficiency overall in our participant population, in accordance with the developmental hypothesis of the RHM. Therefore, we felt it would be interesting to investigate how the number magnitude translation effect manifests itself at much lower levels of L2 proficiency.

Because there is little conformity in the literature regarding the assessment of L2 proficiency, we decided to experimentally manipulate this variable, rather than making use of indirect measures of L2 proficiency. Therefore, we designed a learning experiment in which participants were told that they were going to learn the first 15 number words purported to be from the Estonian language. In reality, we used fabricated words, to exclude any inherent structure in the stimuli and any correspondences with other languages known to the participants (the actual Estonian words for 1–15 are *üks, kaks, kolm, neli, viis, kuus, seitse, kaheksa, üheksa, kümme, üksteist, kaksteist, kolmteist, neliteist,* and *viisteist*). Immediately after the participants had acquired these words, a number-word translation task similar to the previous experiments was administered. If the number magnitude translation effect does not occur in such an experiment, this would offer support for the developmental hypothesis of the hierarchical model. In contrast, if the magnitude effect manifests itself even though these words were acquired only minutes earlier, this would offer substantial evidence in favor of the hypothesis that number-word forms are mapped onto the number line from the very first encounters of these words.

Using a new number language also allowed us to better control our stimulus materials. As indicated above, we took out any inherent structure from the number "words" (as a matter of fact, each participant received a different number–word mapping). In addition, we controlled the number of times participants came across the various L2 words. Although it is common practice in L2 language acquisition to teach all number words up to 12 together, it is still possible that bilinguals in later reading more often encountered some numbers than others and that this accounted for (some of) our magnitude effect.

Method

Participants

Twenty first-year university students participated for course requirements. They all belonged to the group of bilinguals labeled before as unbalanced.

Materials

Fifteen legal six-letter nonwords that followed the Dutch orthographic rules (*soltil, fidara, lacron, nelima, sipron, kodrim, sertir, badiks, dreksa, kummer, dorter, fistar, gabiro, meltir,* and *pridar*) were created and randomly assigned as the translation equivalents of one of the Dutch number words between *een* (one) and *vijftien* (fifteen). Each participant got a different, random mapping of numbers and words. Participants were told that the stimuli were Estonian number words. Using nonwords allowed us to control for word length (each had six letters), prior experience with the stimuli, and any correlation between the number words and the magnitudes they stand for (e.g., in real Estonian, there is a correlation of .93 between the length of the number words and the magnitudes they represent for the integers 1–15). We included number words ranging from 1 to 15 to slightly extend the magnitude range from the previous experiments.

Design

Similar to Experiment 1 and 2, the experiment had a 2 (naming language: L1 vs. L2) $\times$ 3 (stimulus format: Arabic numbers, L1 number words, and L2 number words) $\times$ 15 (number magnitude) full factorial design. All variables were manipulated within subjects.

Procedure

The experiment consisted of four parts: a learning phase, a test phase, an experimental phase, and a delayed naming task.

Learning phase. During each trial, a Dutch number word was presented for 6 s together with its purported Estonian translation. Participants were instructed to memorize these number words so that they would be able to recall a number word given its translation equivalent. No hints were given concerning possible memorization strategies. Each trial was preceded by a fixation stimulus for 500 ms. Number-word combinations were presented in random order. The following test phase started when each number-word combination was shown once.

Test phase. This phase aimed at measuring memorization performance after each learning phase. During each trial, a Dutch or "Estonian" number word was presented on the screen. Participants were instructed to enter the respective translation of the word by means of the keyboard within a 10-s timeframe. A recall test was chosen instead of a recognition test in order to avoid, as much as possible, learning during the test phase. If more than 85% of all trials were correct (i.e., 13 out of 15), a participant proceeded to the experimental phase, after a short break. If not, the learning phase was administered again. On average, participants needed 10 (range 6–16) exposures to the learning phase before they reached the criterion.

Experimental phase. This phase was identical to the number-word translation task of Experiment 1. The only difference concerned the amount of number magnitude conditions (i.e., 15 instead of 12, as described earlier).

Delayed naming task. The day after these phases, participants completed a delayed naming task of Dutch and "Estonian" words, identical to that of Experiment 1. This made it possible to include the same predictors in our regression model as in the previous analyses.

Results

Variance Analysis

The proportion of invalid trials due to naming errors, memorization failure, or faulty time registration was 25.3%. This number is quite high because we opted to use an 85% criterion (rather than a 100% criterion) during the word learning test phase to avoid overlearning and to keep L2 proficiency as low as possible. Of course, this caused a number of false trials in addition to the other errors generally committed in this type of experiment. The invalid trials were excluded from all analyses. An ANOVA was performed with naming language, stimulus format, and number magnitude as repeated measures factors. The dependent variable was the mean RT across correct trials (estimating the resulting few empty cells by means of a formula described by Winer, Brown, & Michels, 1991). Mean RTs as a function of naming language, stimulus format, and number magnitude are presented in Figure 5.

The main effect of stimulus format, $F(2, 38) = 2.66$, $MSE = 169{,}791$, $p < .08$, tended toward significance. The effect of naming language was significant, $F(1, 19) = 171.53$, $MSE = 579{,}952$, $p < .01$. Respective means for Dutch and "Estonian" naming were 869 ms and 1,339 ms. The interaction effect of naming language and stimulus format was also significant, $F(2, 38) = 541.74$, $MSE = 226{,}321$, $p < .01$. Tukey's HSD test showed that forward

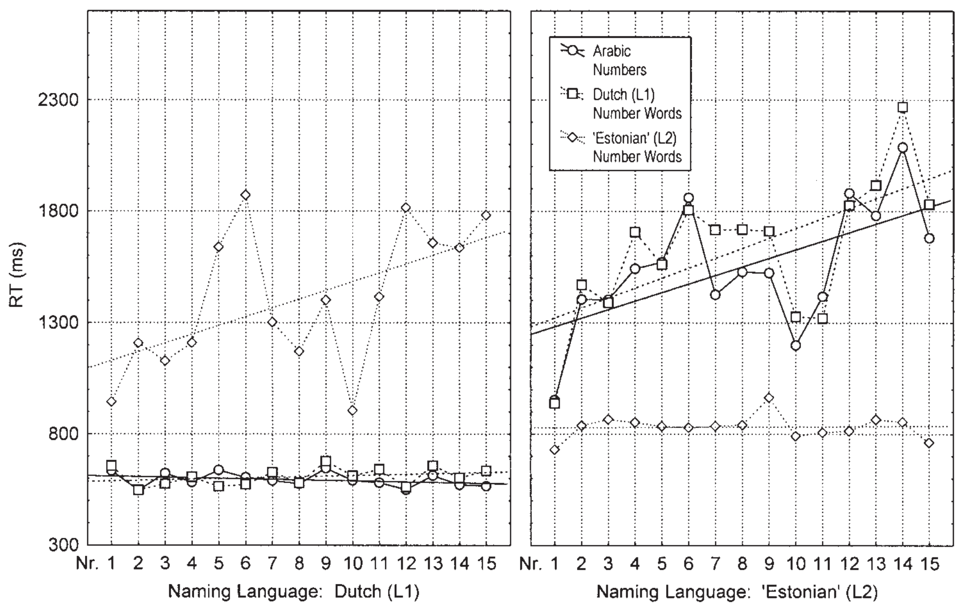

Figure 5. Mean naming response times (RTs) by naming language, stimulus format, and number magnitude (Experiment 3, "Estonian" learning experiment). Straight lines represent best linear fit according to a least-squares criterion. Nr = number.

translation (M = 1,634 ms) was significantly slower than backward translation (M = 1,406 ms), $p < .01$. Also, the effect of number magnitude, $F(14, 266) = 16.50$, $MSE = 149{,}204$, $p < .01$, was significant, but all of these effects were embedded in a three-way interaction of naming language, stimulus format, and number magnitude, $F(28, 532) = 12.41$, $MSE = 89{,}479$, $p < .01$. As can be seen in Figure 5 and similar to Experiments 1 and 2, the effect of number magnitude was present only in some of the Stimulus Format × Naming Language conditions. These effects are analyzed in detail in the following regression analyses.

Regression Analysis

The regression analyses were again performed by the procedure for repeated measures data described by Lorch and Myers (1990, *Method 3,* p. 151). This approach has the additional advantage that missing data in the design (resulting from the use of an 85% learning criterion) need not be estimated to carry out the analysis. For each participant, the multiple regression is calculated on the data that are available. Regression weights for the six Naming Language (L1 vs. L2) × Stimulus Format (Arabic, L1, and L2) conditions are displayed in Table 3. Note that the effect of frequency could not be computed for the conditions with "Estonian" stimulus words because the participants had not encountered the words prior to the experiment, and each word was presented an equal number of times during the learning phase. Consequently, any effect of number magnitude in these conditions cannot be a word frequency effect. Similar to Experiments 1 and 2, the regression weights of number magnitude differed significantly from zero in both forward and backward translation conditions, respectively, $t(19) = 2.427$, $p < .03$, and $t(19) = 2.157$, $p < .01$. These

Table 3
The Regression Equations for the Six Naming Language × Stimulus Format Conditions (Experiment 3) According to the Procedure Described by Lorch and Myers (1990)

Naming language and stimulus format	Intercept	Number magnitude	Delayed naming RT	Frequency
L1 (Dutch)				
Arabic numbers	RT = 634	−1.82 NM	−0.29 D	+0.22 F
L1 number words (Dutch)	RT = 748	+1.55 NM	−0.57 D*	+0.15 F
L2 number words ("Estonian")	RT = 963	+33.37 NM**	+0.26 D	
L2 ("Estonian")				
Arabic numbers	RT = 2,938	+16.03 NM†	+0.43 D	−4.33 F**
L1 number words (Dutch)	RT = 1,219	+18.69 NM**	+5.69 D†	−4.28 F**
L2 number words ("Estonian")	RT = 142	−0.30 NM	+2.10 D**	

Note. RT = response time.
* $p < .05$. ** $p < .01$. † $p < .05$ (one-tailed).

regression weights did not differ significantly from each other, $t(19) = 1.489$, $p > .15$, although the effect of number magnitude tended to be larger in the backward translation condition. Note that an effect of frequency, supplementary to the effect of number magnitude, was found in the forward translation conditions, $t(19) = -3.407$, $p < .01$.

Discussion

The results of Experiment 3 are quite surprising. Large semantic effects of number magnitude were found in both translation directions, even though the participants learned the "Estonian" number words only a few minutes prior to the translation task. The obtained pattern of regression weights was very similar to that of Experiments 1 and 2 (with slower overall mean RTs). A comparison of Figure 5 with Figures 2 and 3 clearly illustrates this. Just as in Experiment 1, there was an effect, albeit somewhat weaker, of number magnitude for L2 naming of Arabic digits. The only remarkable difference from the previous experiments concerns the finding that forward translation was significantly slower than backward translation, whereas the reverse pattern was observed in Experiment 1. As hypothesized earlier, this is probably due to the fact that Experiment 1 contained mixed-language stimulus lists. Note that we also found an effect of frequency in the forward translation condition, supplementary to the semantic number magnitude effect, suggesting a larger degree of lexical activation than in the previous experiments with more proficient bilinguals. However, this effect was not present in the backward translation condition. Also, this did not make the observed semantic number magnitude effects disappear.

Because "Estonian" L2 proficiency was extremely low in this experiment, with participants having seen each word only a few times, this strongly suggests that learned (number) word forms are mapped onto existing abstract (magnitude-related) semantic information very early in the L2 acquisition process. Moreover, this semantic information is activated to a certain degree when translating the presented word forms. This is not in line with a strong developmental hypothesis, which states that newly learned words are initially mapped onto the lexical representation of their translation equivalents. We go into further details about the theoretical implications of these results in the General Discussion.

Although our findings are surprising, they are not without analogues in the literature. In a study on the development of automatic processing, Tzelgov, Yehene, Kotler, and Alon (2000) taught participants a sequence of nonsense symbols, which represented magnitudes from 1 to 9 (although this was never told to the participants). Participants learned about the sequence of the symbols by indicating for pairs of symbols which one was the larger. After a limited amount of training, participants showed the classical effects associated with the number line. In particular, they showed the distance effect (i.e., they were faster to indicate the larger symbol when the distance between the magnitudes was large than when it was small), and they showed the physical size congruity effect (i.e., it was easier to select the physically larger symbol when this symbol represented a large magnitude than when it represented a small magnitude). Similarly, Logan and Klapp (1991) reported that participants could fluently verify equations of the form A + 2 = C (in which the digit indicated how many letters one had to go down the alphabet) after they had memorized six facts in a single session by means of rote rehearsal.

Finally, the fact that we obtained magnitude effects in number translation with stimulus words that were strongly controlled for their structure and the amount of prior exposure is further evidence that the magnitude effect originates from the activation of semantic information and is not a confound of the frequency with which different number words have been translated in the past, or any type of inherent structure that may be present in real-language number words.

Experiment 4

Now that we have shown that number translation is semantically mediated in both directions for a very wide range of L2 proficiency levels, one remaining question is to what extent the number magnitude effect is dependent on the restricted semantic context created by the experimental procedure itself. In the previous experiments, the same—limited—set of stimuli was presented several times in order to obtain reliable RTs in all conditions. In addition, all stimuli came from the same semantic category. This repeated presentation may have caused strong activation of the mental number line, which may have increased the impact of the semantic system on the translation process.

There are reasons to doubt this possibility. For instance, if the number line had been excessively activated, this should also have caused a magnitude effect in the naming of number words, or indeed in the L1 naming of Arabic digits. The semantic route can be used for correct performance in these conditions as well (e.g., Fias, 2001; Reynvoet et al., 2002). Also, an analysis of the initial trials of Experiment 1 showed that the number magnitude effect was already present from the first few trials on.[3]

Nevertheless, as the hierarchical model has mainly been developed to explain out-of-context translation performance in experimental settings, we wanted to investigate empirically whether our findings would hold in a less restricted semantic context. Therefore, we designed a translation experiment in which each of the 12 number words was presented only once. Moreover, these number words were scattered among 230 other middle- to high-frequency filler words that had to be translated as well. To keep the amount

[3] We ran an analysis of only the first 2 (out of 10) trials of the first (out of two) naming block of Experiment 1. Hence, we used only the first 10% of all trials for each subject in our analysis (estimating the resulting few empty cells by means of a formula described by Winer et al., 1991). Despite the small number of data points included in this illustrative analysis, the regression weights of number magnitude were significant for both forward, $B = 6.94$, $t(21) = 5.135$, $p < .01$, and backward translation, $B = 5.08$, $t(21) = 7.940$, $p < .01$. Thus, the magnitude effect in number-word translation was already present during the first trials of the experiment, at which point such long-term priming by semantic context may be argued to have not yet taken effect. Note that the used blocked stimulus presentation in Experiment 2 could not result in sufficient data points to perform a similar analysis. Because the participants in that experiment completed all 10 trials of a particular Naming Language × Stimulus Format condition before proceeding to the next block, only data from the first block of each participant could be used in an analysis of the early trials. Because there are six possible Naming Language × Stimulus Format conditions and 12 participants, such an analysis would be based on only 2 participants per condition.

of number presentations as low as possible, and because the previous experiments showed consistent results for these conditions, we did not include within-language naming and Arabic digit naming. Consequently, the present study consisted of two subsequent blocks of forward and backward translation with, respectively, Dutch (L1) and French (L2) stimulus words. Thus, as in Experiment 2, the present experiment contained blocked stimulus language presentation.

Method

Participants

Twenty-nine first-year university students participated for course requirements. They all belonged to the group of bilinguals labeled before as unbalanced.

Materials

All materials were identical to those in the previous experiments. The stimuli for each language consisted of the number words ranging from 1 to 12 and 230 other middle- to high-frequency filler words. Mean log CELEX frequency per million of the filler words (Baayen, Piepenbrock, & Van Rijn, 1993) was 1.06 ($SD = 0.72$).

Design

The experiment had a 2 (direction of translation: forward vs. backward) × 12 (number magnitude) full factorial design. All variables were manipulated within subjects.

Procedure

All participants completed two blocks, each consisting of 242 trials. Hence, each (number word) stimulus was presented only once. The 12 number-word trials were randomly scattered among the 230 filler trials. Naming language was Dutch (L1) in one block (French L2 stimuli) and French (L2) in the other (Dutch L1 stimuli). Order of blocks was determined at random. The procedure for a trial was identical to that used in Experiments 1, 2, and 3.

Results

Variance Analysis

The proportion of invalid number-word trials due to naming errors or faulty time registration was 17.0%. Because the nature of the design allowed only one observation per condition, we excluded trials from all analyses online as soon as there was the slightest doubt concerning accurate functioning of the voice key, in order to get mean RTs that were as reliable as possible with such a low number of observations. These trials were excluded from all analyses. An ANOVA was performed with direction of translation and number magnitude as repeated measures factors. The dependent variable was the mean RT across correct trials (estimating the resulting few empty cells by means of a formula described by Winer et al., 1991). Mean RTs as a function of direction of translation and number magnitude are presented in Figure 6.

The effect of direction of translation was nearly significant, $F(1, 28) = 3.84$, $MSE = 144{,}547$, $p < .06$. Respective means for forward and backward translation were 1,053 ms and 996 ms. As in the previous experiments, there was also a significant effect of number magnitude, $F(11, 308) = 10.32$, $MSE = 43{,}349$, $p < .01$. This number magnitude effect just failed to interact with direction of translation, $F(11, 308) = 1.67$, $MSE = 42{,}817$, $p < .08$. These effects are analyzed in detail in the following regression analyses.

Regression Analysis

The regression analyses were again performed by the procedure for repeated measures data described by Lorch and Myers (1990,

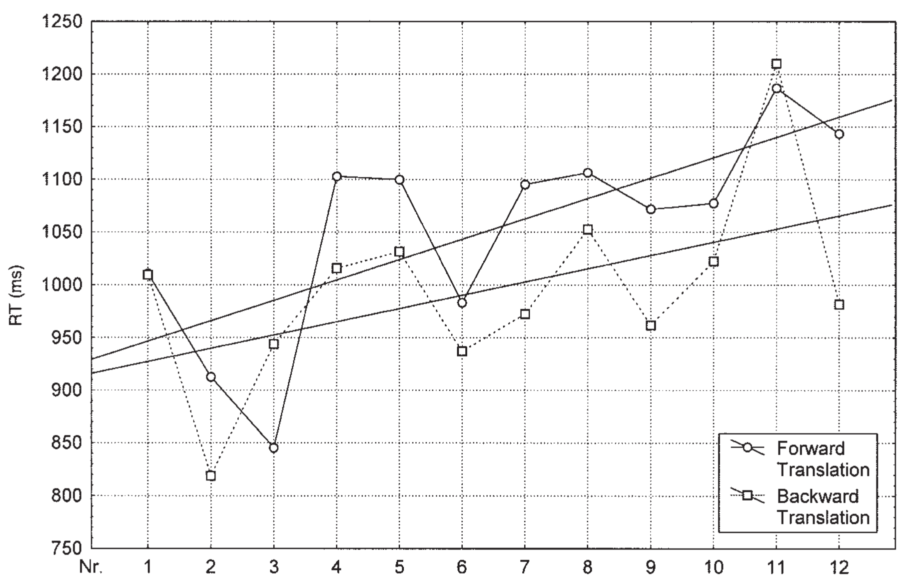

Figure 6. Mean naming response times (RTs) by direction of translation and number magnitude (Experiment 4, mixed-context stimuli). Straight lines represent best linear fit according to a least-squares criterion. Nr = number.

Method 3). This approach has the advantage that the empty cells in the design, resulting from false trials combined with the use of a single observation per condition, need not be estimated to carry out the analysis. Regression weights for the two direction of translation conditions are displayed in Table 4. Similar to the previous experiments, the regression weights of number magnitude differed significantly from zero in both forward and backward translation conditions, respectively, $t(28) = 2.097$, $p < .05$, and $t(28) = 2.403$, $p < .03$. The effect of number magnitude tended to be larger for forward translation, but this difference was not significant, $t(28) = 1.146$, $p > .25$. No lexical effects of word frequency were found.

Discussion

Again, we found an effect of number magnitude for both directions of translation. Consequently, it can be concluded that the translation of number words is semantically mediated also when these words are presented in a diversified semantic context. The fact that we obtained the number magnitude translation effect with only one observation per participant per condition is a further indication that the effect is quite robust. The effect of number magnitude tended to be larger for forward translation (as predicted by the RHM), but this difference failed to reach significance. No effect of word frequency was found, suggesting a limited degree of lexical activation. Finally, it is interesting to note that, just as in Experiment 3 (very low L2 proficiency) but in contrast to Experiment 1 (mixed-language stimulus lists), backward translation was faster ($p < .06$, two-tailed) than forward translation. However, this did not substantially affect the degree of semantic activation during translation.

General Discussion

The RHM of Kroll and Stewart (1994; Kroll & de Groot, 1997), depicted in Figure 1, postulates that forward translation is more likely to be conceptually mediated than backward translation, because links between the lexicon and the semantic system are stronger for L1 than for L2. Backward translation will be semantically mediated only for high levels of L2 proficiency. As discussed in the introduction, these assumptions have been supported by a number of studies (e.g., Cheung & Chen, 1998; Kroll & Stewart, 1994; Sholl et al., 1995; Talamas et al., 1999; for a review, see Kroll & de Groot, 1997; Kroll et al., 2002).

However, the results of our number-word translation experiments were not completely in line with some of the model's predictions. In four experiments, we obtained clear semantic effects of number magnitude, not only when number words were translated forward from L1 to L2 but also when they were translated backward from L2 to L1. Thus, for Dutch–French bilinguals, it took less time to forward translate *twee* (two; L1) into *deux* (L2) than *acht* (eight; L1) into *huit* (L2) but also to backward translate *deux* into *twee* than *huit* into *acht.* There was no statistically reliable difference in the number magnitude effect between backward and forward translation. The observed difference slightly tended toward the expected direction for two of the four experiments (i.e., a smaller magnitude effect in backward translation than in forward translation was found for Experiments 1 and 4) but tended toward the opposite direction for the other two experiments.

In addition, the effect of number magnitude in the translation conditions did not interact with L2 proficiency. It was equally strong in unbalanced and balanced bilinguals in Experiment 1, despite the reliable difference in mean RTs between both groups. Also, we replicated the number magnitude translation effect in participants with extremely low L2 proficiency with number words that had been acquired only a few minutes before the translation experiment (Experiment 3). This suggests that number-word forms are mapped onto existing semantic information (the mental number line in this particular case) very early in the L2 acquisition process. Moreover, the mappings are strong enough to exert an influence on both backward and forward translation. These findings are not in line with a strong developmental hypothesis, which states that L2 word forms have connections only with the abstract concepts that they represent at high levels of L2 proficiency. On the other hand, our observation of rapid semantic connections between new symbols and number magnitudes is in line with the literature of numerical cognition (see Logan & Klapp, 1991; Tzelgov, Yehene, Kotler, & Alon, 2000).

There are a few other studies that failed to confirm the predictions of the RHM (Kroll & Stewart, 1994). Using a bilingual Stroop task, La Heij et al. (1996, Experiment 1) found that congruent color words (for which the ink color corresponded to the word) were translated faster than incongruent color words. This was true for both directions of translation. In further experiments (La Heij et al., 1996, Experiments 3, 4, and 5), they found a facilitation effect (which was larger on translation trials than on naming trials) of distractor pictures depicting an object (e.g., a table) belonging to the same semantic category as the target word to be translated (e.g., *chair*), compared with unrelated pictures. This was also true for both directions of translation. Similar effects of context words and pictures in backward translation were recently reported by Bloem and La Heij (2003). But, as Kroll and de Groot (1997, pp. 183–184) argued, context availability induced by the accompanying pictures may have provided the semantic support for L2 to access the meaning system, whereas the RHM was designed to account for out-of-context translation performance. It can be argued that this criticism also applies to our first three experiments. Although no explicit semantic cues (such as the distractor stimuli of La Heij et al. and Bloem and La Heij) were used in our experiments, repeated presentation of the same—limited—stimulus set may have induced sufficient semantic activation for L2 to access the semantic system (similar to results reported by Sholl et al., 1995). However, the increased semantic activation account is a much less plausible explanation for Experiment 4, where the number words were presented only once, amid

Table 4
The Regression Equations for the Two Direction of Translation Conditions (Experiment 4) According to the Procedure Described by Lorch and Myers (1990)

Translation	Intercept	Number magnitude	Delayed naming RT	Frequency
Backward	RT = 1,721	+9.91 NM*	−1.53 D	−0.55 F
Forward	RT = 2,938	+24.37 NM*	−0.16 D	+0.23 F

Note. RT = response time.
* $p < .05$.

a whole range of other words from different semantic categories. Also, this account rests uneasily with the observation that the magnitude effect was present in translation conditions only and not in direct word naming or in the L1 naming of Arabic digits, although the semantically mediated route also plays a role in these tasks (e.g., Fias, 2001; Reynvoet et al., 2002).

Another recent study suggesting early L2 lexicosemantic links, even after a few hours of artificial L2 learning, is that of Altarriba and Mathis (1997). After training a group of monolinguals on a set of English–Spanish word pairs, they found more errors on both lexically and semantically related false translations than on unrelated words in a translation recognition task (Altarriba & Mathis, 1997, Experiment 1). In a more recent study by von Pein and Altarriba (2003), similar findings were reported for English participants learning noniconic American Sign Language gestures. In line with our findings in the "Estonian" L2 learning study (Experiment 3), these findings suggest that links between L2 and the conceptual system can be established quite early. However, Altarriba and Mathis also reported more form-related interference for nonfluent than for fluent bilinguals, supporting the developmental hypothesis of the RHM. In further experiments (Altarriba & Mathis, 1997, Experiment 2), they reported a bilingual Stroop effect similar to that found by La Heij et al. (1996) using the same L2 training procedure as in their first experiment. Talamas et al. (1999) attributed these findings to the fact that the involved semantic representations may have been primed by the semantic procedures used during the training phase. Again, this alternative explanation by Talamas et al. would seem to be less applicable to our third experiment, given the fact that no semantic memorization strategy was used during the word-learning phase. On the contrary, the associative word-learning procedure used was more likely to elicit L2 word learning by means of lexical connections.

Finally, evidence against translation asymmetries has been reported by de Groot and colleagues (e.g., de Groot, 1992; de Groot & Comijs, 1995; de Groot, Dannenburg, & van Hell, 1994; de Groot & Poot, 1997; Van Hell & de Groot, 1998b). Their translation studies showed that semantic word variables (e.g., imageability or context availability) affected RTs in both backward and forward translation. Although there were some (although not always convincing) indications of asymmetries in the translation processes in some of the studies, this was not the case in the studies of de Groot and Comijs (1995) and de Groot and Poot (1997), where semantic variables affected forward and backward translation to an equal degree in three groups of bilinguals, differing in their level of L2 proficiency.

To conclude this discussion, we would like to point to some theoretical implications of the present findings. In line with the above studies, our results are clearly not compatible with a strong asymmetric model of bilingual memory or with a strong developmental lexical hypothesis. Although the RHM has done a good job in explaining a wide range of findings and in producing new research hypotheses, the incompatible findings described above (as well as those of this study) indicate that some updating may be warranted. In our view, two amendments could entail a considerably useful extension of the model.

First, the model, like many models of that time, is implicitly based on the horse-race metaphor. There are two routes, and the fastest wins (i.e., completely determines the output). So, translation follows either the lexical route (in backward translation) or the semantic route (in forward translation). There is no influence from the slower route. Increasingly, however, horse-race models are being replaced by connectionist-type models (the most famous example being Coltheart's dual-route model of visual word naming; see Coltheart, 1978, vs. Coltheart, Curtis, Atkins, & Haller, 1993). Following this approach, the central question should no longer be whether the output comes from one *or* the other route but *how much* each of the routes contributes to the buildup of the overall output activation. In this view, one route is not faster than the other (the processing cycle is the same throughout the model); it only may have stronger connection weights and, therefore, influence the activation of the output units to a larger degree. If this line of reasoning is applied to the RHM, depicted in Figure 1, this would mean that the connection weights are stronger from L2 lexical units to L1 lexical units than the other way around. Similarly, the connection weights between L1 units and semantic units would be stronger than those between L2 units and semantic units. So, even though both the direct lexical–lexical and the indirect lexical–semantic–lexical routes change the activation level of the units at each processing cycle, their relative contributions can differ (and maybe appear even nonexistent, if the respective weights are very small) as a function of the translation direction.

A second proposal to improve the current theoretical framework of the RHM concerns the fact that the asymmetry depends only on the proficiency level of the bilingual. No distinction is made between different types of words. Because of this, the model has to predict the same semantic involvement for the translation of all types of words, including number words, abstract words, and even syntactic function words. For this reason, too, our findings with number words are critical for the model as a whole, even though there is independent evidence that the linking between new symbols and meanings is particularly fast for numerical stimuli (Logan & Klapp, 1991; Tzelgov et al., 2000). It would seem to us that the general framework of the RHM can easily be adapted to include influences at the word level as well as at the subject level (e.g., proficiency), certainly if a connectionist-type model is used. For a start, the connection weights between the lexical units and the semantic units would depend on the consistency of the mappings between the words and the meanings. For each language, they would be bigger (and grow much faster in the acquisition phase) for words that always have the same meaning independent of the context (e.g., *two*) than for words that have different meanings as a function of the context (e.g., *great*). Second, it does not seem unreasonable to assume that some words have a richer semantic representation than others, which could be implemented by the number of semantic features to which the word units are connected. Finally, the impact of the semantically mediated route on translation times would also depend on the degree of semantic overlap between two translation equivalents.

Note that some of these ideas are already partly present in the distributed feature model of de Groot and colleagues (e.g., de Groot, 1992; de Groot, 1993; de Groot et al., 1994; Van Hell & de Groot, 1998a; Van Hell & de Groot, 1998b). According to this model, the overlap in meaning, indexed by the number of shared semantic features, depends on word concreteness (with concrete translation equivalents sharing more features than abstract words).[4] These shared features become active during translation and facilitate the translation process, resulting in faster translation of concrete words (e.g., de Groot, 1992; de Groot et al., 1994; Van Hell & de Groot, 1998b). This line of reasoning might also explain the strong semantic effects obtained in this study, as the meaning of number words is virtually completely overlapping across languages. At this point, it is important to note that the findings of Duyck, Szmalec, Kemps, and Vandierendonck (2003) suggest that the early lexicosemantic mapping observed in Experiment 3 is probably not restricted to words from which the meaning is so clearly defined and overlapping, as is the case for number words. Using a selective interference paradigm, they showed that new word forms are mapped onto available existing semantic (visual) information during associative word–new word learning (e.g., *auto–plornam* [car–legal Dutch nonword]), provided that such a visual representation is available. This shows that visual information—if possible—is coded early during the word acquisition process, which is compatible with the work of de Groot and colleagues.[5]

Figure 7 gives the broad outline of how such a model might look. Note, however, that the model described above is a hypothetical description, which of course must be implemented before all of the intricacies become clear.[6] In the model, the semantic overlap is different for certain types of words (just as in the distributed feature model), with the overall weights of the lexicosemantic connections being stronger for L1 (just as in the RHM). This results in the activation of a smaller amount of semantic nodes, feeding activation into the L1 lexicon during backward translation of certain L2 words (but not for words with a very large semantic overlap between languages, e.g., *eight*).

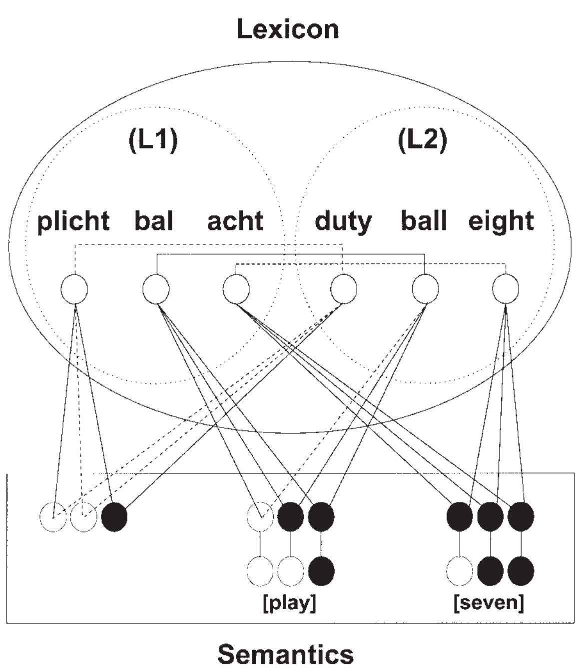

Figure 7. The revised hierarchical model of bilingual memory with varying semantic overlap and differently weighted lexicosemantic and intralexical connections. Solid lines represent stronger links than dotted lines. Depicted words and semantic representations are illustrative examples for Dutch–English bilinguals. L1 = first language; L2 = second language.

Notice that the model in Figure 7 also provides a straightforward interpretation of the semantic activation account, previously used by Kroll and de Groot (1997) to explain the data of the Stroop experiments (La Heij et al., 1996), for example. In a connectionist-type model, the change of activation caused in a unit by a particular connection is a function not only of the weight of that connection but also of the activation level of the unit from which the connection originates. So, enhancing the activation level of the semantic units (e.g., by the presentation of a picture) would already suffice to increase the impact of the semantic route in a model like the one depicted in Figure 7.

Finally, it might be argued that connection weights differ as a function of word characteristics not only between the lexical and the semantic level but possibly also between the L1 and the L2 lexicon. It could be that these connections are stronger for words with a large form overlap (e.g., in Figure 7, *ball–bal* for an English–Dutch bilingual) than for words with a small form overlap (e.g., in Figure 7, *duty–plicht* for an English–Dutch bilingual) (e.g., Van Hell & de Groot, 1998a). This could provide an explanation for the fact that translation equivalents with a large form overlap (so-called cognates) are easier to translate and show less evidence for semantic mediation in the translation process than words with no form overlap (e.g., de Groot, 1992; see also Sánchez-Casas et al., 1992).[7] However, in the model described above, such strong lexical links do not necessarily exclude all semantic influences. This is compatible with Kroll and Stewart (1994), who demonstrated that cognate translation may be affected by the semantic organization of list context. Note that the intralexical connections, depicted in Figure 7, between two word nodes are bidirectional (just as the lexicosemantic connections in the RHM and the intralexical connections in the BIA+ [bilingual interactive activation] model; see Dijkstra & Van Heuven, 2002). This does not mean that

[4] It is important to note here that recent research (see Tokowicz & Kroll, 2003) has shown that certain word concreteness effects may alternatively be explained by competition effects resulting from cross-language ambiguity (i.e., the number of translation equivalents in each language for a given concept, which correlates with concreteness). Note that this alternative explanation of these effects is also quite compatible with the model depicted in Figure 7.

[5] Note that recent work from Izura and Ellis (2004) on effects of L1 and L2 age of acquisition illustrates that word concreteness is not the only word variable that may affect lexicosemantic organization. In this view, it is interesting to note that the stimuli used in this study (number words) are words that are acquired very early (both in L1 and L2).

[6] Notice that the model in Figure 7 contains one integrated lexicon in line with recent developments in the literature on bilingual lexical organization. As this discussion is outside the scope of this article, see Dijkstra and Van Heuven (2002) for a recent review of evidence supporting this assumption.

[7] Whereas it is reasonable to assume strong facilitatory, intralexical connections for words that share both form and meaning (cognates), this may be less plausible for words that are similar in form but have different meanings in both languages (interlingual homographs). However, de Groot, Delmaar, and Lupker (2000), for example, argued that slower processing of interlingual homographs in a translation recognition task can be explained in terms of competition between semantic (and not lexical) representations.

the impact of these lexical connections will always be equally large for both directions of translation. On the contrary, the asymmetry (which is also present in the RHM) follows from the lexicosemantic weights, which are weaker for certain L2 words (e.g., *duty, ball*) than for their L1 counterparts. This may cause smaller incoming semantic activation in the L1 word node during backward translation, resulting in a relatively larger impact of the intralexical activation (even with a bidirectional weight). However, at present we do not exclude the possibility of unidirectional connections with differing weights (as is the case in the RHM at the lexical level). Future modeling will have to show the necessity of this assumption. In general, we believe the strength of the model sketched above lies partly in the fact that it may be useful to explain or predict in which cases semantic activation during translation performance will be asymmetrical and in which cases it will not. Overall, it is plausible to assume that bilingual language organization in the model will become more symmetrical as connection weights increase and approach their final state (e.g., through increasing L2 experience and proficiency). This will hold especially true for highly frequent words, which were used as stimuli in a number of studies (e.g., this study; Altarriba & Mathis, 1997; La Heij et al., 1996) that failed to demonstrate translation asymmetries.

In summary, we found a semantic effect of number magnitude when number words were translated from Dutch (L1) to French (L2) and vice versa. Number words representing smaller quantities (e.g., *twee, deux* [two]) were translated faster than number words representing larger quantities (e.g., *acht, huit* [eight]). The effect was replicated using different procedures, semantic contexts, and levels of L2 proficiency (including a very low level). These results strongly suggest that, at least for certain types of words, the mappings between L2 words and their meaning are more important than the intralexical mappings between the L2 words and their L1 equivalents, already from the first stages of L2 acquisition on. On the basis of these findings, we have concluded that translation should not be viewed as an all-or-none semantic or lexical process but rather as the simultaneous buildup of activation from both the lexical and the semantic route. Furthermore, we have suggested that the contribution of each route depends not only on the translation direction and the L2 proficiency but also on the characteristics of the words involved in the translation. As such, we proposed a model in which the overall architecture of Kroll and Stewart's RHM is preserved but in which we reviewed the way the different components interact, by dropping the implicit horse-race assumption.

References

Altarriba, J., & Mathis, K. M. (1997). Conceptual and lexical development in second language acquisition. *Journal of Memory and Language, 36,* 550–568.

Altenberg, E. P., & Cairns, H. S. (1983). The effects of phonotactic constraints on lexical processing in bilingual and monolingual subjects. *Journal of Verbal Learning and Verbal Behavior, 22,* 174–188.

Baayen, R., Piepenbrock, R., & Van Rijn, H. (1993). *The CELEX lexical database* [CD-ROM]. Philadelphia: University of Pennsylvania, Linguistic Data Consortium.

Bloem, I., & La Heij, W. (2003). Semantic facilitation and semantic interference in word translation: Implications for models of lexical access in language production. *Journal of Memory and Language, 48,* 468–488.

Brysbaert, M. (1995). Arabic number reading: On the nature of the numerical scale and the origin of phonological recoding. *Journal of Experimental Psychology: General, 124,* 434–452.

Brysbaert, M. (in press). Number recognition in different formats. In J. I. D. Campbell (Ed.), *Handbook of mathematical cognition.* Hove, England: Psychology Press.

Brysbaert, M., Van Dyck, G., & Van de Poel, M. (1999). Visual word recognition in bilinguals: Evidence from masked phonological priming. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 137–148.

Buckley, P. B., & Gillman, C. B. (1974). Comparisons of digits and dot patterns. *Journal of Experimental Psychology, 103,* 1131–1136.

Campbell, J. I. D. (1994). Architectures for numerical cognition. *Cognition, 53,* 1–44.

Cheung, H., & Chen, H. C. (1998). Lexical and conceptual processing in Chinese-English bilinguals: Further evidence for asymmetry. *Memory & Cognition, 26,* 1002–1013.

Cipolotti, L., & Butterworth, B. (1995). Toward a multiroute model of number processing: Impaired number transcoding with preserved calculation skills. *Journal of Experimental Psychology: General, 124,* 375–390.

Coltheart, M. (1978). Lexical access in simple reading tasks. In G. Underwood (Ed.), *Strategies of information-processing* (pp. 151–216). London: Academic Press.

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review, 100,* 589–608.

de Groot, A. M. B. (1992). Determinants of word translation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 1001–1018.

de Groot, A. M. B. (1993). Word-type effects in bilingual processing tasks. Support for a mixed-representational system. In R. Schreuder & B. Weltens (Eds.), *The bilingual lexicon: Vol. 6. Studies in bilingualism* (pp. 27–57). Amsterdam: John Benjamins.

de Groot, A. M. B., & Comijs, H. (1995). Translation recognition and translation production: Comparing a new and an old tool in the study of bilingualism. *Language Learning, 45,* 467–509.

de Groot, A. M. B., Dannenburg, L., & van Hell, J. G. (1994). Forward and backward word translation by bilinguals. *Journal of Memory and Language, 33,* 600–629.

de Groot, A. M. B., Delmaar, P., & Lupker, S. J. (2000). The processing of interlexical homographs in translation recognition and lexical decision: Support for non-selective access to bilingual memory. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 53*(A), 397–428.

de Groot, A. M. B., & Poot, R. (1997). Word translation at three levels of proficiency in a second language: The ubiquitous involvement of conceptual memory. *Language Learning, 47,* 215–264.

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition, 44,* 1–42.

Dehaene, S., & Akhavein, R. (1995). Attention, automaticity, and levels of representation in number processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 314–326.

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General, 122,* 371–396.

Dehaene, S., Dupoux, E., & Mehler, J. (1990). Is numerical comparison digital: Analogical and symbolic effects in two-digit number comparison. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 626–641.

Dehaene, S., & Mehler, J. (1992). Cross-linguistic regularities in the frequency of number words. *Cognition, 43,* 1–29.

Dijkstra, T., Grainger, J., & Van Heuven, W. J. B. (1999). Recognition of cognates and interlingual homographs: The neglected role of phonology. *Journal of Memory and Language, 41,* 496–518.

Dijkstra, T., Timmermans, M., & Schriefers, H. (2000). On being blinded by your other language: Effects of task demands on interlingual homograph recognition. *Journal of Memory and Language, 42,* 445–464.

Dijkstra, T., & Van Heuven, W. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition, 5,* 175–197.

Duyck, W., & Brysbaert, M. (2002). What number translation studies can teach us about the lexico-semantic organisation in bilinguals. *Psychologica Belgica, 42,* 151–175.

Duyck, W., Diependaele, K., Drieghe, D., & Brysbaert, M. (2004). The size of the cross-lingual masked phonological priming effect does not depend on second language proficiency. *Experimental Psychology, 51*(2), 1–9.

Duyck, W., Szmalec, A., Kemps, E., & Vandierendonck, A. (2003). Verbal working memory is involved in associative word learning unless visual codes are available. *Journal of Memory and Language, 48,* 527–541.

Ferrand, L. (1999). Why naming takes longer than reading? The special case of Arabic numbers. *Acta Psychologica, 100,* 253–266.

Fias, W. (2001). Two routes for the processing of verbal numbers: Evidence from the SNARC effect. *Psychological Research, 65,* 250–259.

Fias, W., Lammertyn, J., Reynvoet, B., Dupont, P., & Orban, G. A. (2003). Parietal representation of symbolic and nonsymbolic magnitude. *Journal of Cognitive Neuroscience, 15,* 47–56.

Fias, W., Reynvoet, B., & Brysbaert, M. (2001). Are Arabic numerals processed as pictures in a Stroop interference task? *Psychological Research, 65,* 250–259.

Foltz, G. S., Poltrock, S. E., & Potts, G. R. (1984). Mental comparison of size and magnitude: Size congruity effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 442–453.

Gielen, I., Brysbaert, M., & Dhondt, A. (1991). The syllable-length effect in number processing is task-dependent. *Perception & Psychophysics, 50,* 449–458.

Ito, Y., & Hatta, T. H. (2003). Semantic processing of Arabic, Kanji, and Kana numbers: Evidence from interference in physical and numerical size judgments. *Memory & Cognition, 31,* 360–368.

Izura, C., & Ellis, A. W. (2004). Age of acquisition effects in translation judgement tasks. *Journal of Memory and Language, 50,* 165–181.

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

Koechlin, E., Naccache, L., Block, E., & Dehaene, S. (1999). Primed numbers: Exploring the modularity of numerical representations with masked and unmasked semantic priming. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 1882–1905.

Kroll, J. F. (1993). Assessing conceptual representations for words in a second language. In R. Schreuder & B. Weltens (Eds.), *The bilingual lexicon: Vol. 6. Studies in bilingualism* (pp. 53–82). Amsterdam: John Benjamins.

Kroll, J. F., & de Groot, A. M. B. (1997). Lexical and conceptual memory in the bilingual: Mapping form to meaning in two languages. In A. M. B. de Groot & J. F. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives* (pp. 169–199). Mahwah, NJ: Erlbaum.

Kroll, J. F., Michael, E., Tokowicz, N., & Dufour, R. (2002). The development of lexical fluency in a second language. *Second Language Research, 18,* 137–171.

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

La Heij, W., Hooglander, A., Kerling, R., & Vandervelden, E. (1996). Nonverbal context effects in forward and backward word translation: Evidence for concept mediation. *Journal of Memory and Language, 35,* 648–665.

Logan, G. D., & Klapp, S. T. (1991). Automatizing alphabet arithmetic: I. Is extended practice necessary to produce automaticity? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 179–195.

Lorch, R. F., & Myers, J. L. (1990). Regression analyses of repeated measures data in cognitive research. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 149–157.

McCloskey, M. (1992). Cognitive mechanisms in numerical processing: Evidence from acquired dyscalculia. *Cognition, 44,* 107–157.

Meuter, R. F. I., & Allport, A. (1999). Bilingual language switching in naming: Asymmetrical costs of language selection. *Journal of Memory and Language, 40,* 25–40.

Moyer, R. S., & Landauer, T. K. (1967, September 30). Time required for judgments of numerical inequality. *Nature, 215,* 1519–1520.

Naccache, L., & Dehaene, S. (2001). The priming method: Imaging unconscious repetition priming reveals an abstract representation of number in the parietal lobes. *Cerebral Cortex, 11,* 966–974.

Nas, G. (1983). Visual word recognition in bilinguals: Evidence for a cooperation between visual and sound based codes during access to a common lexical store. *Journal of Verbal Learning and Verbal Behavior, 22,* 526–534.

Pesenti, M., Thioux, M., Seron, X., & De Volder, A. (2000). Neuroanatomical substrates of Arabic number processing, numerical comparison, and simple addition: A PET study. *Journal of Cognitive Neuroscience, 12,* 461–479.

Pinel, P., Dehaene, S., Riviere, D., & Lebihan, D. (2001). Modulation of parietal activation by semantic distance in a number comparison task. *Neuroimage, 14,* 1013–1026.

Reynvoet, B., & Brysbaert, M. (1999). Single-digit and two-digit Arabic numerals address the same semantic number line. *Cognition, 72,* 191–201.

Reynvoet, B., Brysbaert, M., & Fias, W. (2002). Semantic priming in number naming. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 55*(A), 1127–1139.

Sánchez-Casas, R. M., Davis, C. W., & Garcia-Albea, J. E. (1992). Bilingual lexical processing: Exploring the cognate/non-cognate distinction. *European Journal of Cognitive Psychology, 4,* 293–310.

Sholl, A., Sankaranarayanan, A., & Kroll, J. F. (1995). Transfer between picture naming and translation: A test of asymmetries in bilingual memory. *Psychological Science, 6,* 45–49.

Smith, M. C. (1997). How do bilinguals access lexical information? In A. M. B. de Groot & J. F. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives* (pp. 145–167). Mahwah, NJ: Erlbaum.

Talamas, A., Kroll, J. F., & Dufour R. (1999). From form to meaning: Stages in the acquisition of second-language vocabulary. *Bilingualism: Language and Cognition, 2,* 45–58.

Thomas, M. S. C., & Allport, A. (2000). Language switching costs in bilingual visual word recognition. *Journal of Memory and Language, 43,* 44–66.

Tokowicz, N., & Kroll, J. F. (2003). *Cross-language ambiguity and concreteness effects: The role of meaning and form ambiguity in cross-language processing.* Manuscript submitted for publication.

Tzelgov, J., Yehene, V., Kotler, L., & Alon, A. (2000). Automatic comparisons of artificial digits never compared: Learning linear ordering relations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 103–120.

Van Hell, J. G., & de Groot, A. M. B. (1998a). Conceptual representation in bilingual memory: Effects of concreteness and cognate status in word association. *Bilingualism: Language and Cognition, 1,* 193–211.

Van Hell, J. G., & de Groot, A. M. B. (1998b). Disentangling context availability and concreteness in lexical decision and word translation.

Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 51(A), 41–63.

Van Wijnendaele, I., & Brysbaert, M. (2002). Visual word recognition in bilinguals: Phonological priming from the second to the first language. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 619–627.

von Pein, M., & Altarriba, J. (2003). *The development of linguistic knowledge in naïve learners of American Sign Language.* Manuscript submitted for publication.

Von Studnitz, R. E., & Green, D. W. (1997). Lexical decision and language switching. *International Journal of Bilingualism, 1,* 3–24.

Winer, B. J., Brown, D. R., & Michels, K. M. (1991). *Statistical principles in experimental design* (3rd ed.). New York: McGraw-Hill.

Received September 6, 2001
Revision received October 15, 2003
Accepted March 22, 2004 ■