JOURNAL OF MEMORY AND LANGUAGE 33, 600–629 (1994)

Forward and Backward Word Translation by Bilinguals

ANNETTE M. B. DE GROOT, LUCIA DANNENBURG, AND JANET G. VAN HELL

University of Amsterdam, The Netherlands

The determinants of performance in word translation by unbalanced bilinguals, fairly fluent in their second language, were studied. Translation was both from the subjects' native (L1) to their second (L2) language and in the reverse direction ("forward" and backward" translation, respectively). The predictor variables were imageability, context availability, definition accuracy, familiarity, word frequency, length (each of these six was determined for the L1 and L2 words separately), and the cognate status of the translation equivalents. Both forward and backward word translation were influenced by meaning variables, familiarity variables, and cognate status. However, meaning played a somewhat more important role in forward than in backward translation, whereas familiarity appeared to have a larger influence in backward translation. A few other differences between forward and backward translation were detected, but, when considering the complete stimulus set, the differences between translation directions were generally small. In some of the subsets of the stimulus materials (particularly noncognates) larger directional differences occurred. Particularly relevant is the finding that meaning affects backward translation, because it suggests a qualification of the "asymmetry model" of word translation as proposed by Kroll and Stewart (1994).

Among the tasks frequently used in the study of bilingualism is word translation, in which subjects simply translate words presented in one of their two languages into the other. Potter, So, von Eckardt, and Feldman (1984) employed this task in conjunction with a second task, picture naming in a second language. Their purpose was to contrast two plausible models of bilingual memory and processing: the "word-association" model and the "concept-mediation" model. The models are depicted in the upper part of Fig. 1.

Both models discriminate between two levels of representation in bilingual memory, one lexical and one conceptual, and are therefore sometimes called "hierarchical" (Snodgrass, 1984). They also both assume the existence of two separate lexicons, one for each of the bilingual's two languages. However, they differ in the connections they assume between the various memory stores. The word-association model postulates direct connections between the representations of translation equivalents in the two lexicons. In this model there are no direct connections between the lexical representations of words in the bilingual's weaker, second, language (L2) and the representations of the associated concepts in conceptual memory. As a consequence, L2 words can only access their concepts indirectly, that is, via the lexical representations of their translations in this bilingual's first language (L1). In contrast, the concept-mediation model denies the existence of direct links between the lexical representations of a translation pair, but instead assumes direct connections between representations in the L2 lexical store on the one hand and the representations of the associated concepts in conceptual memory on the other. L2 words thus enable direct access of conceptual memory. Both models assume that a repre-

We gratefully acknowledge the contribution of Rik Poot, Hanny Comijs, John Hoeks, Paul Oosterveld, and Nico van Ruijven to this study. They assisted in collecting and/or analyzing the data. We thank Robert Dufour, Judith Kroll, and three anonymous reviewers for very helpful textual and theoretical comments. Correspondence and reprint requests may be addressed to Annette M. B. de Groot, Department of Psychology, University of Amsterdam, Roetersstraat 15, 1018 WB Amsterdam, The Netherlands.

0749-596X/94 $6.00

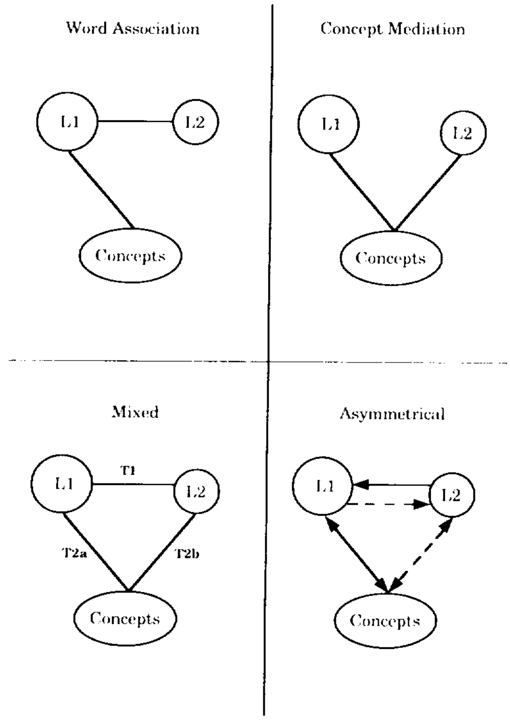

FIG. 1. Some models of bilingual memory. Critical differences between them are the type of connections between the memory systems each of them contains and the strength of these connections.

sentation in conceptual memory is shared by both words of a translation pair, that is, conceptual representation is language-independent. Typically the L1 lexicon will contain more elements than the L2 lexicon. This is depicted in Fig. 1 by the difference in size of the circles representing both lexicons.

The word-association and concept-mediation models of Potter et al. (1984) are reminiscent of two of the bilingual representational structures postulated by Weinreich (1953), namely, the "subordinative" (word association) and "compound" (concept mediation) structures, for both of which Weinreich assumed language-independent conceptual representations, shared by the two languages. In addition to these two, Weinreich considered the existence of a third type of bilingual memory structure, the "coordinate" structure. Unlike in the other two, in a coordinate structure the L1 and L2 words access different, language-specific, conceptual representations. Whether a pair of translations is represented in one language-independent conceptual representation or in two language-specific conceptual representations may depend on a number of factors, for instance, the language history of the bilingual (Ervin & Osgood, 1954) and word type (de Groot, 1993).

Potter et al. (1984) argued that the word-association model predicts shorter latencies for word translation from L1 to L2 than for picture naming in L2, whereas the concept-mediation model predicts equal response times (RTs) for the two tasks (see Potter et al., 1984, for details). Because they obtained equally long RTs in both tasks, they opted for the concept-mediation model. Using the same experimental procedure (that is, comparing translation and picture-naming times) but testing two subject groups that clearly differed in L2 proficiency, Kroll and Curley (1988) observed a data pattern supporting the word-association model for low-proficiency bilinguals but the concept-mediation model for high-proficiency bilinguals (see also Chen & Leung, 1989).

Rather than comparing translation time and picture-naming time, one of us (de Groot, 1992a) used the word-translation task on its own to study bilingualism. Subjects' performance on different classes of words was compared. Translation was always from the subjects' native language (Dutch) to their second language (English). Insight into the structure of bilingual memory and in bilingual processing can be acquired in this basic use of the word-translation task by manipulating semantic variables: if and only if conceptual memory is implicated in the task (concept mediation), performance should respond to these variables. We will substantiate this claim further on.

Among the stimulus characteristics that were varied by de Groot (1992a) were the cognate status of the translation pairs (are the translation equivalents orthographically and/or phonologically similar?); the familiarity, imageability, definition accuracy, and context-availability of the stimulus words; the written word frequency of the stimulus words and of the intended translations; and the length of the stimulus words and of their translations. As in most studies manipulating imageability, this variable was confounded with word concreteness. For convenience, words rated high on imageability will be called *concrete* here; those rated low on imageability will be called *abstract*. Definition accuracy is defined as the ease with which subjects think they could define the word. Subjects rating words on context availability are asked to score words on how easily they think they could come up with a sentence or sentence fragment for each of the words. This variable is known to be highly correlated with word imageability (e.g., Schwanenflugel, 1991) and, as it turned out, with definition accuracy (de Groot, 1992a). The reasons to include each of these stimulus characteristics as a variable are provided by de Groot (1992a) and will not be repeated here.

The following variables turned out to be strong determinants of translation performance: the frequency of the stimulus (L1) and response (L2) words, the cognate status of the translation pairs, and word imageability (although it was argued that context availability might underlie the latter effect; see de Groot, 1992a, Experiment 3). A further interesting finding was that frequency and imageability interacted: the imageability effect was larger for high-frequency words than for low-frequency words. The finding that cognate status and imageability affect bilingual performance converges with the results of other studies, using different bilingual processing tasks (for instance, cross-language semantic priming and repetition priming; cross-language word association; de Groot & Nas, 1991; Jin, 1990; Kolers, 1963; Sánchez-Casas, Davis, & García-Albea, 1992; Taylor, 1976).

A large part of these translation data can be explained in terms of bilingual memory structures of the type illustrated in the lower left part of Fig. 1 (de Groot, 1992a, 1992b, also suggests alternative accounts). This type of "mixed" memory structure integrates the word-association and concept-mediation models in that the lexical representations of the translation equivalents are now both directly connected (link T1, for translation route 1), and indirectly, via a shared representation in conceptual memory (Links T2a + T2b). As pointed out above, opting for the concept-mediation model, Potter et al. (1984) concluded that T1 links do not exist, but studies by Kroll and Stewart (1990, 1994) and de Groot and Nas (1991) point out their presence in memory.

The interpretation of most of the above effects is in terms of the strength of the various links in these memory structures (see de Groot, 1992a, for more detail). The L1 and L2 word-frequency effects can be attributed to the strength of links T2a and T2b. Frequent use of a word in a bilingual's native language will cause the T2a link to become strong. Similarly, frequent use of a word in his or her second language will strengthen the T2b link associated with this word. The stronger both of these links, the better (in terms of speed and errors) translation via conceptual memory will be. There is another reason why the T2a and T2b links, and *also* the T1 links, will be stronger for frequent words than for infrequent words: words often used in monolingual settings are also likely to occur relatively often in translation settings. Each time when a word is translated, the memory connection between the word and its translation that is used, be it T1 or T2a + T2b, will be strengthened. The contiguity of a word and its translation per se may further strengthen T1 in each translation act.

A plausible interpretation of the effect of cognate status is in terms of differences in the strength of the T1 links between cognates and noncognates. Collins and Loftus (1975) suggested that orthographically and phonologically similar words are linked in lexical memory. They only considered the lexical memory of monolinguals, but there is no reason why these lexical connections should only occur between words of the same language. In other words, in addition to being translations, as compared to noncognates cognates have an extra reason to be linked in lexical memory. This could be reflected in relatively strong L1 links (see also de Groot, 1992a).

A way to conceive of the imageability effect is to assume that concrete words are relatively often represented in the structures as illustrated in the lower left part of Fig. 1, whereas abstract words are more often represented in language-specific nodes in conceptual memory than concrete words (cf. Weinreich's, 1953, coordinate representations). Such an architecture would usually provide two translation routes for concrete words (T1 and T2a + T2b), but relatively often just one for abstract words (T1). If additional translation routes benefit performance, concrete words would be translated better than abstract words, as indeed they are. Finally, the interaction between word frequency and imageability can be understood in terms of different translation routes for frequent and infrequent words. Imageability, representing a *semantic* aspect of words, may only be expected to affect performance if conceptual memory is implicated in the translation process. When conceptual memory is not involved, but instead translation comes about via the links between the representations of a pair of translations within lexical memory, no effect of this variable has to be expected. The present interaction thus suggests that infrequent words are translated via the T1 links relatively often, whereas in translating frequent words the route along the T2a and T2b links is more often traversed.

This interpretation of the frequency by imageability interaction is consistent with Kroll and Curley's (1988) finding, mentioned earlier, that low-proficiency bilinguals translate via the word-association (T1) links, whereas more fluent bilinguals translate words through conceptual mediation. Both their finding and the present interaction suggest that L2 practice, through the strengthening of the links between lexical and conceptual memory, causes a shift from word-association translation to conceptually mediated translation (high-proficiency bilinguals have had more L2 practice than low-proficiency bilinguals, and each individual bilingual has had more practice, both in L1 and in L2, with high-frequency words than with low-frequency words). Note that a critical link in this argument, introduced earlier in this article, is that only when conceptual memory is implicated in translation, semantic variables may have an effect on performance. This assumption will play a crucial role in our test of the asymmetry model of bilingual memory of Kroll and Stewart (1990, 1994), which will be introduced before long.

The present study extends the investigation of de Groot (1992a) in three ways: (1) instead of only studying "forward" translation, from the subjects' native language (L1, again Dutch) to their second language (L2, again English), we now also look at word translation from L2 to L1 ("backward" translation); (2) a set of new predictor variables will be included; (3) subjects somewhat more fluent than those participating in de Groot (1992a) will also be tested. The main goal of this study is to identify determinants of both forward and backward translation, but Kroll and Stewart's (1990, 1994) asymmetrical model of word translation will receive special emphasis.

The reason for studying word translation in both directions is that a number of studies, primarily by Kroll and her colleagues (Kroll & Stewart, 1990, 1994), have shown that directionality effects occur when using this task. One of their findings suggesting directionality is that translating words from L1 to L2 took longer than translating words from L2 to L1. This asymmetry effect occurred both for relatively proficient and for less proficient bilingual subjects, although it was larger for the latter group of subjects. Sánchez-Casas et al. (1992) also demonstrated the effect, but only for words with a noncognate translation in the other language (that is, a word orthographically and/or phonologically dissimilar to the to-be-translated word); words with a cognate translation (orthographically and phonologically similar) were translated equally fast in both directions. The data by Sánchez-Casas et al. (1992) thus indicate that the directionality effect depends on word type.

Yet another result of Kroll and Stewart (1990, 1994) suggested directionality in word translation. These authors presented their stimulus words either in categorized lists (all words belonging to the same semantic category, e.g., all garments) or in randomized lists (words selected from several semantic categories and presented in a random order). When the stimulus words were translated from the subjects' L2 (English) to their L1 (Dutch), the categorized and randomized lists showed equally long translation times, but when translating was from L1 to L2, the categorized lists produced longer translation RTs. Because the mixing/blocking manipulation concerned a semantic dimension of the stimulus materials, this finding suggests the processing of meaning in forward translation but not in backward translation.

To account for these directionality effects, Kroll and Stewart (1990, 1994; see also Kroll, 1993) expanded the mixed bilingual memory representations to create an "asymmetrical" memory structure illustrated in the lower right part of Fig. 1. Instead of one direct link between the lexical representations of the L1 and L2 words, they assumed there to be two, one from the L1-node to the L2-node and one in the reverse direction. The link from the L2-node to the L1-node is regarded stronger than the reverse link. Additionally, different strengths were assigned to the link from the L1-node in lexical memory to the L1/L2-node in conceptual memory on the one hand and the link from the L1-node in lexical memory to the L2-node in lexical memory: the former link was assumed to be stronger than the latter. Kroll and Stewart argued that these differences between the strengths of individual links cause forward translation to proceed generally via conceptual memory, whereas backward translation typically exploits the direct links between nodes in lexical memory. They attributed the shorter RTs in backward than in forward translation to the fact that the (direct) translation route from L2 to L1 in lexical memory is shorter than the (indirect) translation route, via conceptual memory, from L1 to L2. The differential effects in forward and backward translation of manipulating a semantic variable is also explained in terms of the use of these different translation routes, although the difference in *length* of the routes is now irrelevant: only when translation comes about via conceptual memory (where word meanings are stored), an effect of manipulating semantic variables may be expected (see also above). Consequently, only forward translation should be sensitive to these manipulations.

It seems to us that especially the latter finding provides strong support for the "asymmetrical" model of word translation (and bilingual representation). The former finding, shorter RTs from L2 to L1 than from L1 to L2, is suggestive, but does not appear to be conclusive, because probably not length per se but the strength of the links constituting a translation route determines response speed. Tracing *n* strong links may thus take an equal amount of time or even less time than tracing *n-m* weaker links. So to be able to attribute unequivocally the RT difference between forward and backward translation to different translation routes, the strengths of all links between the various memory nodes should be known. Without this information, alternative accounts of the RT asymmetry cannot be ruled out. For instance, this asymmetry between translation directions could reflect the difference between "recall" (L1 to L2) and "recognition" (L2 to L1), as some have argued (Snodgrass, 1993). To anticipate, the results of the study reported here, especially those of Experiment 2, indeed suggest that the above directional effect on RT is no reliable indication of asymmetrical word translation as specified by the asymmetry model of Kroll and Stewart (1994).

Unlike the overall RT difference between the two translation directions, the differential effect of the above semantic manipulation in forward and backward translation clearly supports the asymmetry model. Moreover, as pointed out by Kroll and Sholl (1992), the finding converges with another asymmetry, namely that the magnitude of cross-language semantic priming depends on the language of prime and target. When the primes are presented in L1 and the targets in L2 the cross-language effect is larger than when the language of primes and targets is reversed (see e.g. Altarriba, 1990; Keatley, Spinks & de Gelder, 1990; Tzelgov & Eben-Ezra, 1992). As Kroll and Stewart (1990) suggested, the reason may be that L1 primes are more likely to activate conceptual representations than L2 primes (Kroll & Stewart, 1990, p. 6). In sum, it will be especially interesting to see whether or not effects of the present semantic variables will occur in backward translation and, if so, how they compare to the corresponding effects for forward translation.

As mentioned above, the second extension as compared to the study by de Groot (1992a) concerned the use of a new set of predictor variables. Among the predictors included in that study were familiarity, context availability, and definition accuracy (see above). The associated ratings had been done on the set of Dutch words only, that is, the stimulus words in that study, but the response words in the backward condition of the present investigation. For the purpose of the present study, ratings concerning these three variables were performed on the (dominant) English translations of these Dutch words, in other words, on the response words in forward translation but the stimulus words in backward translation. In addition, new imageability ratings were collected, both for the Dutch words and for their translations in English. In de Groot (1992a) imageability (of the Dutch words only) had also served as a predictor variable, but the ratings had been taken from an existing corpus of Dutch imageability ratings. The reason to collect new ratings, also for the Dutch words for which ratings already existed, was that this way the subject samples for all rating studies would be drawn from the same population, first-year psychology students from the University of Amsterdam. This was also the population used in the actual translation experiments. The subjects performing the translation task in Experiments 1 and 2 were tested about 4 months (Experiment 1) and about 11 months (Experiment 2) after starting their university studies. The data of the norming studies were collected all through the subjects' first year at the university.

In sum, the present study aims to reveal determinants of word translation in both directions and to relate the results to bilingual memory representation and processing. Special attention will be devoted to the question whether forward and backward translation exploit different memory pathways.

Norming Studies

Materials

The materials of the norming studies consisted of 458 Dutch nouns and their translations in English. The Dutch set consisted of the words that had been presented as stimulus words in the forward-translation study of de Groot (1992a, Experiment 3). The English set consisted of the dominant translations of the Dutch words. Two of the constraints when first selecting the Dutch words had been that they varied widely both in word frequency and in word imageability. For an assessment of the Dutch words' imageability, an existing corpus of Dutch word-imageability ratings had been consulted (van Loon-Vervoorn, 1985). As already mentioned, for the purpose of the present experiment new imageability ratings for all stimulus words, both in Dutch and in English, were collected from a sample of subjects drawn from the population, also selected from for the actual translation studies (see below for more detail). To assess the words' frequency, again both in Dutch and in English, we consulted the frequency count of the Centre for Lexical Information (CELEX) in Nijmegen, The Netherlands (Burnage, 1990). The English database of this corpus consists of a count of 18.8 million printed words and the Dutch database consists of a count of 42.5 million words. Two further constraints when selecting the Dutch words had been (1) that they either had a unanimous translation in English or a clearly dominant translation and (2) that about half of them had a cognate translation in English, whereas the remaining half had a noncognate translation. Cognate status of the translation pairs was initially judged by the authors and subsequently assessed in one of the norming studies.[1]

Procedure. Nine sets of norming data were (or had been) collected to provide information on nine stimulus characteristics that were to serve as predictor variables in this study (in addition to four predictors that did not require the norming of the materials; see below). One of these sets involved the cognate ratings of the translation pairs. These cognate ratings had been determined in two earlier studies (de Groot, 1992a, Experiment 3; de Groot & Nas, 1991). The subjects providing them had been asked to rate on a 7-point scale how similar they regarded the words within each translation pair (7 = very high similarity; 1 = very low similarity). The subjects were told that a rating should reflect a combined assessment of both spelling and sound similarity of the word pair under consideration. Half of the translation pairs to be rated had consisted of cognates (e.g., *huis–house*) and the remainder of noncognates (e.g., *boom–tree*), as initially judged by the experimenters before starting to collect the rating data. To assess the reliability of the similarity ratings, 134 translation pairs occurred in both studies. Following data collection, a mean cognate rating for each translation pair, collapsed across subjects, was calculated. The ratings for the 134 pairs common to both studies correlated highly ($r = .98$), suggesting that the rating procedure was reliable.

The remaining eight norming studies involved ratings of the (1) context availability, (2) definition accuracy, (3) familiarity, and (4) imageability of all 458 stimulus words in both languages, the data of the Dutch words and their (dominant) English translations being collected in separate studies. The scores on familiarity, defini-

[1] It may seem that a proportion of 50% cognates in the stimulus set is larger than the actual proportion of Dutch–English cognates in the Dutch vocabulary as a whole (but see Snow & Hoefnagel-Höhle, 1978, p. 1124, who mention that approximately 65% of *commonly* used words are Dutch–English cognates). If true, any effects of cognate status to be obtained might be due to a conscious search through the other-language lexicon for a word physically resembling the stimulus word. De Groot (1992a, Footnote 13, p. 1015) reported a study intended to rule out an interpretation of an effect of cognate status in terms of such strategic processing. She argued that this strategic processing should result in a larger effect of cognate status when the proportion of cognates among the materials was large than when small. Contrary to this prediction, the effect of cognate status was equally large when the proportion of cognates to noncognates was 1:3 as when this proportion was 1:1.

tion accuracy, and context availability of the Dutch words had been collected before in three different norming studies (see de Groot, 1992a, Experiment 3). For the purpose of the present study the five remaining sets of norming data were collected, in five separate studies. Every individual subject could participate in only one of the norming studies and none of the subjects participating in a norming study also performed the actual translation task. In all eight norming studies, subjects were randomly assigned to two groups. Each group rated 229 out of the 458 stimulus words. The assignment of words to groups was based on an alphabetical ordering of all 458 words: the first word in alphabetical order was assigned to Group 1, the second to Group 2, the third again to Group 1, and so on. The number of subjects within a group varied between 22 and 32 across the eight studies. The words were presented to the subjects in booklets, eight words per page. The numbers 1 to 7 were printed on the right-hand side of each word. All words were printed underneath one another. Words were randomly assigned to a page, but care was taken that no page contained a preponderance of one particular type of words (e.g., most of them relatively common or most of them concrete). The pages were reshuffled in every new booklet so that every subject rated the words in a unique order. To be able to assess the intergroup reliability of the ratings, 38 words were presented to both groups within each study. All data were collected in test sessions in which minimally 2 and maximally 16 subjects participated. Average time spent on rating the 229 words of a stimulus set varied between 11 and 23 min across the eight studies.

In the two *context-availability norming studies*, one on the Dutch words and one on their (dominant) English translations, the instructions for the subjects were those used by Schwanenflugel and Shoben (1983, as reported in Schwanenflugel, Harnishfeger, & Stowe, 1988), but translated into Dutch and providing the subjects with different example words, better tailored to the Dutch language and culture. The subjects were asked to rate words on a 7-point scale on "how easy it is to come up with a particular context or circumstance in which they might appear" (Schwanenflugel et al. 1988, p. 502; 1 = very hard to think of a context; 7 = very easy to think of a context). The subjects rating English words also received the instructions in Dutch. The reason was that this guaranteed the same degree of understanding of the instructions by the subject groups rating Dutch and English words, and hence, comparability of the ratings across languages. In all further norming studies, the instructions were also presented in Dutch to all groups providing the ratings. Instructions were always presented in written form.

In the *definition-accuracy norming studies*, the corresponding instructions of de Groot (1992a, p. 1010) were used. Subjects were asked to rate words on a 7-point scale on how accurately they thought they could define them. A 1 should be chosen if they thought they could hardly define the word; a 7 if they thought they could define it very accurately.

The procedure in the *familiarity norming studies* was based on that by Noble (1953, pp. 90–91), but here, as in all of the remaining norming studies, ratings were done on a 7-point scale (Noble used a 5-point scale), only the end-points of the scale were labeled verbally in the instructions (Noble explicitly labeled all scale points), and example words different from those provided by Noble were presented. In Noble's study and in our Dutch condition, the subjects were asked to rate words as to the number of times they had experienced them, with a 1 indicating that the subject had never seen, heard, or used the word in his or her life, and a 7 indicating that he or she had seen, heard, or used it nearly every day of his or her life. The instructions for the English stimulus set were the same as for the Dutch set, except that the verbal label for scale-point 7 was not "seen or heard or used nearly every day of your life," but "seen or heard or used relatively often." The reason for this change was that the everyday language of our subjects is Dutch, not English, so that it could be expected that the "nearly every day" instruction would have had the effect that the rating points at the higher end of the scale would hardly ever be chosen for the English words.

Finally, in the two *imageability norming studies* we used the instructions of Paivio, Yuille, and Madigan (1968), but translated into Dutch. Subjects were asked to rate words as to the ease or difficulty with which they arouse mental images. Any word which was thought to arouse a mental image very quickly and easily should be given a high imagery rating; any word thought to arouse a mental image with difficulty or not at all should be given a low imagery rating. As usual, ratings were made on a 7-point scale.

After data collection a mean score for each of the words in each norming study was calculated, collapsing across the subjects within a group. Also, for each of the norming studies the correlation was calculated between the Group 1 and the Group 2 scores for the 38 words common to both groups. In all eight norming studies these two sets of scores correlated highly (all *r*'s between .90 and .98). Therefore, the fact that the scores on a given predictor variable for the complete set of 458 words were collected from two different groups of subjects will be ignored in all further analyses.

Experiment 1

Subjects

Twenty-six first-year psychology students from the University of Amsterdam, The Netherlands, participated as subjects in the backward-translation task, receiving course credit for participation. They were drawn from the same population as the subjects tested in the forward-translation study (De Groot, 1992a) that is to be compared with the present study. The subjects were all unbalanced bilinguals, with Dutch as their native language (L1). They were all relatively fluent in their second language (L2), which was English. They started to learn English at school around the age of 12, and until they went to university (at age 18–19) they had had 3 or 4 h of English classes a week. Their schooling at the university required them to read mainly in English. The subjects, both those translating backward here and those translating forward in de Groot (1992a), were tested about 4 months after starting their university studies. On entering the laboratory they were asked to rate on a 7-point scale their comprehension and production abilities in English (1 = very low; 7 = same as in Dutch, the native language). The mean comprehension and production ratings of the backward-translation subjects were 4.96 (*SD* .87) and 4.38 (*SD* .85), respectively. The corresponding ratings for the forward-translation subjects had been 5.02 (*SD* .87) and 4.55 (*SD* 1.06).

The degree of L1–L2 language balance of our subjects may also be assessed from the performance on simple word-recognition tasks of subject groups comparable to those tested here: when performing lexical decision to reasonably common Dutch words, our subjects typically respond about 20 ms faster than when making these decisions on the corresponding English words (de Groot & Nas, 1991; Swaak, 1992). When pronouncing reasonably frequent Dutch words, the onset of their responses is about 100 ms faster than it is when pronouncing these words in English (see de Groot, 1992a, p. 1008, for a brief report of this pronunciation study).

Apparatus and procedure. The backward-translation experiment was run on an Apple Macintosh Plus computer in a normally lit room. Stimuli were presented in black lowercase letters against a light-grey background on the computer screen. A PASCAL program controlled the stimulus presentation and the recording of the response times (RTs). A microphone that activated a voice-operated switch registered the subjects' responses. Subjects sat facing the screen at a comfortable reading distance. The experimenter sat to the left of the subject, typing the subjects' responses on the computer keyboard and monitoring the workings of the voice switch. Failures of the voice switch to respond to the subject's response and triggering by another sound were noted down by the experimenter.

Prior to the experiment the subjects read an instruction sheet in Dutch. In these instructions they were asked to speak out loud the Dutch translations of the English stimulus words. They were asked to do so as quickly as possible, while making as few errors as possible and to remain silent in case they did not know the stimulus word's translation. The 458 stimulus words were presented in two groups, the same groups of 229 words each as used in the norming studies. The first subject first translated the words of Group 1 and then those of Group 2; the second subject started out with the words of Group 2 and then translated those of Group 1; the third subject again started with the words of Group 1, and so on. The within-group presentation of the words was randomized and different for all subjects. After presentation of the first group of words, the subject paused for about 5 min before presentation of the second group started. The test words of both groups were preceded by 15 practice words, which were the same words for both groups.

The sequence of events on each trial was as follows: prior to the stimulus word, a fixation stimulus (an asterisk) appeared in the middle of the screen for 1 s, slightly above where the word was to appear. Immediately after it disappeared, the stimulus word appeared and remained on the screen until the voice switch registered the onset of the subject's response (or of any other sound). RT was measured from the onset of the stimulus. The experimenter then typed the subject's response (what was being typed did not appear on the screen). Finally, the experimenter touched the RETURN key, 1 s after which the next trial started. The maximum presentation duration for a stimulus was 5 s. Whenever this duration expired, the experimenter typed the word *none* and called the next trial by hitting the RETURN key. The stimuli were presented in blocks of 24 each. After each block the subject was permitted a brief rest before the experimenter initiated the presentation of the next block. The experiment lasted about 90 min.

The forward-translation study (de Groot, 1992a), the data of which are to be compared with the presently collected backward-translation data, was run on a Schneider PC 1640 DD, an IBM compatible computer. The procedure in that study was identical to that adopted here, except of course that the subjects were asked to produce their responses (to the Dutch stimulus words) in English.

Results and Discussion

For each stimulus word a mean backward-translation RT was calculated, collapsed across all of the subjects who had provided a correct response for that word (maximum: $N = 26$). Furthermore, an error score (ER) was calculated for each stimulus word. A response was regarded as an error when it was not listed among the translations of the stimulus word in a popular English-to-Dutch translation dictionary (Martin & Tops, 1984). Apart from error-responses, trials on which other sounds than the translation response had triggered the voice-switch were also excluded when calculating the RTs for the stimulus words. In addition to RT and ER, a third performance measure to be calculated for each stimulus word was an omission score (OS), that is, the number of subjects who did not come up with a response to that word within 5 s following the onset of the stimulus.

Forty-seven subjects had participated in the forward-translation study of de Groot (1992a). To be able to base the comparison of forward and backward translation on equally large sets of data, the data of 21 subjects, randomly selected from the total of 47, were removed from the earlier forward-translation data set. For each stimulus word the mean RT and the ER- and OS-scores based on the data of the remaining 26 subjects were then computed.

For various reasons (for instance, zero or just one correct response out of the 26 potential responses in either forward or in backward translation) the data associated with 18 of the 458 stimulus words were excluded from further analyses. Most of the analyses reported below were based on the data of the two groups of 26 subjects each, 440 data points per subject. Whenever analyses were performed on subsets of these data, this is stated explicitly.

The first analysis was performed on the forward-translation data of the selected 26 subjects and 440 words. In this analysis the correlations between the same 14 variables (10 predictor variables and 4 dependent variables)[2] as those that had been included in the earlier forward-translation study were calculated. This analysis was run to see whether the exclusion of a number of subjects and stimuli from the earlier data set would alter the response pattern. This turned out not to be the case. All correlation coefficients between the dependent variables on the one hand and the original predictor variables on the other were about equally large as before, although they tended to be slightly smaller. None of the old coefficients differed significantly from the corresponding new coefficient. It thus seems legitimate to continue our analyses on this subset of the earlier forward-translation data while maintaining the earlier study as a frame of reference.

Multiple regression analyses. The variance accounted for in a series of three multiple regression analyses on the forward-translation data by de Groot (1992a) had varied between 25 and 48%. In all three of these analyses the 10 predictor variables of that study had been included and each single one had gotten assigned RT, OS, or ER as the dependent variable. To see whether and to what extent the new predictor variables of the present study would raise the accounted variance, these same analyses were now performed with 13 predictors. One of the variables from the earlier study, verbal context availability, was removed from the set of predictor variables.[3] The newly collected Dutch word imageability ratings replaced the imageability ratings used by de Groot (1992a). The reason for this replacement was that this way all sets of norming data would have been gathered from samples drawn from the same subject population. So the new set contained the following 13 predictors: the six variables context availability (CA), definition accuracy (DEF), imageability (IMA), familiarity (FAM), length (LEN), and log-word-frequency (LOGF), each assessed for both the Dutch and the English words separately (adding up to 12), and cognate status (CS). A second set of three multiple regression

[2] In addition to RT, ER, and OS, de Groot (1992a, Experiment 3) included a fourth dependent variable which was based on the *z* score transformations of the remaining three performance measures. This measure will not be included here because, being an aggregate of the other three, it does not provide additional information.

[3] The subjects providing the verbal-context-availability norming data had been asked to rate words on how easy it is to think of a sentence or a sentence fragment in which the word (could) occur(s). The reason we removed this variable from the set of predictor variables was twofold. First, we had reason to doubt the reliability of the verbal-context-availability norming data. When collecting the norming data the same procedure was used as reported for the norming studies presented in the present article: the words to be rated were split up in two lists, to be rated by two different groups of subjects, but 38 words were presented to both groups so that the intergroup reliability of the ratings could be assessed. The intergroup reliability for the verbal-context-availability data was considerably smaller than it was for all remaining sets of norming data in de Groot (1992a) and in the present study. Second, the pattern of correlations between the predictor variables in de Groot (1992a) suggested that subjects in the verbal-context-availability norming study had not complied with the instructions.

analyses was run including this same set of 13 predictors, but each single one now taking backward-translation RT, OS, or ER as the dependent variable. The percentages of accounted variance in these analyses are presented in Table 1. For comparison, the percentages accounted for with only the ten predictors of the earlier study included are also shown.[4] A striking outcome is that adding the new set of predictors increased the percentage of accounted variance dramatically, between 6 and 35% (14.5% on average). Furthermore, it seems prediction is somewhat better for backward translation than for forward translation.

Correlational analyses on the complete data set. For both translation directions the correlations between the 16 variables (three dependent variables and 13 predictors) were calculated. The correlations between the three dependent variables on the one hand and the 13 predictors on the other are presented in Table 2. The correlations between the predictors are shown in Table 3.

As can be seen in Table 2, imageability, context availability, definition accuracy, cognate status, familiarity, and log word frequency, both for the Dutch (D) and for the English (E) words, and both for forward and backward translation, all correlated negatively with the dependent variables. In contrast, the length of the Dutch and English words correlated positively with the dependent variables. In other words, forward as well as backward translation performance (in terms of response speed, number of errors and number of omissions) is relatively good for words that are easy to imagine, to think a context for and to define, that are perceptually similar to their translation, familiar, frequent, and short.

A noteworthy aspect of Table 3 is that all

TABLE 1

PERCENTAGES OF ACCOUNTED VARIANCE IN MULTIPLE REGRESSION ANALYSES WITH 10 OR 13 PREDICTOR VARIABLES AND 3 DEPENDENT VARIABLES IN FORWARD AND BACKWARD TRANSLATION

Dependent variable	10 Predictors	13 Predictors Experiment 1	13 Predictors Experiment 2
		forward	
RT	45	56	57
OS	45	61	55
ER	22	28	23
		backward	
RT	50	63	62
OS	32	67	63
ER	26	32	37

Note. RT, reaction time; OS, omission score; ER, error rate.

semantic variables, that is, variables that require that the subject retrieves the meaning of the word in order to make the rating (IMA-D, IMA-E, DEF-D and DEF-E, and presumably also CA-D and CA-E) all correlated highly with one another. The same holds for the set of variables that reflect how often words are used (FAM-D, FAM-E, LOGF-D, and LOGF-E), although these correlation coefficients tend to be slightly smaller. We will refer to these clusters of variables as the *semantic* and *familiarity* variables, respectively. But note that one of the familiarity variables (FAM-E) correlates moderately (r = .55 and r = .58) with two of the variables of the semantic cluster (CA-E and DEF-E). One interpretation of this finding is that our subjects, when instructed to rate English words (that is, words in their L2) on how easy they can be thought a context with or defined, take into account an assessment of the familiarity of these words in the target language. A further noteworthy aspect of Table 3 is that the semantic variables all correlated negatively with the two length variables, LEN-D and LEN-E. Finally, the two length variables correlated positively with one another, and weak negative correlations occurred between the length variables on the one hand and cognate status on the other.

Factor analyses. The correlations in Ta-

[4] The percentages of accounted variance reported in Table 1 for the analyses of the forward-translation data including 10 predictor variables deviate somewhat from the corresponding percentages reported in de Groot (1992a, p. 1011). The reason is that the analyses are based on different data sets (the present set constituting a subset of the earlier data set; see above).

TABLE 2
CORRELATION MATRIX OF THE 13 PREDICTOR VARIABLES WITH THE DEPENDENT VARIABLES FOR FORWARD AND BACKWARD TRANSLATION IN EXPERIMENTS 1 AND 2

	Forward						Backward					
	RT		OS		ER		RT		OS		ER	
Predictor variable	Expt 1	Expt 2	Expt 1	Expt 2	Expt 1	Expt 2	Expt 1	Expt 2	Expt 1	Expt 2	Expt 1	Expt 2
IMA-D	−.35	−.36	−.29	−.24	−.25	−.21	−.28	−.22	−.14	−.11	−.19	−.17
IMA-E	−.45	−.47	−.42	−.38	−.34	−.27	−.40	−.34	−.28	−.26	−.28	−.28
CA-D	−.43	−.46	−.40	−.36	−.25	−.24	−.37	−.33	−.24	−.22	−.27	−.24
CA-E	−.58	−.60	−.58	−.56	−.39	−.32	−.57	−.55	−.53	−.52	−.38	−.43
DEF-D	−.31	−.32	−.29	−.24	−.22	−.20	−.26	−.22	−.14	−.12	−.19	−.15
DEF-E	−.55	−.58	−.59	−.57	−.37	−.36	−.59	−.58	−.65	−.64	−.40	−.47
CS	−.38	−.36	−.34	−.29	−.28	−.19	−.36	−.36	−.22	−.17	−.31	−.26
FAM-D	−.35	−.35	−.40	−.36	−.23	−.31	−.49	−.46	−.38	−.35	−.34	−.33
FAM-E	−.60	−.59	−.68	−.61	−.32	−.31	−.66	−.69	−.72	−.68	−.45	−.53
LOGF-D	−.41	−.41	−.46	−.42	−.30	−.33	−.52	−.51	−.42	−.37	−.32	−.37
LOGF-E	−.40	−.39	−.45	−.40	−.26	−.25	−.45	−.47	−.46	−.42	−.34	−.38
LEN-D	.24	.21	.21	.18	.22	.10	.13	.08	.03	−.02	.15	.07
LEN-E	.32	.34	.29	.24	.26	.16	.22	.17	.06	.05	.15	.10

Note. RT, reaction time; OS, omission score; ER, error rate; IMA, imageability; CA, context availability; DEF, definition accuracy; CS, cognate status; FAM, familiarity; LOGF, log word frequency; LEN, length; D, Dutch words; E, English words. $p < .05$ if $r > .08$; $p < .01$ if $r > .11$.

ble 3 suggest the existence of at least two underlying factors, a Semantic factor and a Familiarity factor. This was confirmed in a number of factor analyses (Principal Component Analyses) on the data associated with the predictor variables. In some of these analyses the number of factors to extract was fixed by us (at 4), whereas in others it was allowed to be generated spontaneously (the default method); in some of the analyses only the data associated with the 13 predictors of Table 3 were entered, whereas in others the data associated with other variables were included as well (ver-

TABLE 3
CORRELATION MATRIX OF THE 13 PREDICTOR VARIABLES

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13
1. IMA-D	—												
2. IMA-E	.94	—											
3. CA-D	.82	.80	—										
4. CA-E	.71	.80	.76	—									
5. DEF-D	.83	.81	.81	.64	—								
6. DEF-E	.64	.74	.66	.83	.66	—							
7. CS	.16	.20	.13	.16	.11	.15	—						
8. FAM-D	.05	.09	.18	.26	.03	.31	.08	—					
9. FAM-E	.07	.20	.26	.55	.09	.58	.14	.63	—				
10. LOGF-D	−.05	.03	.06	.20	−.01	.30	.09	.64	.64	—			
11. LOGF-E	−.10	−.01	.01	.21	−.04	.29	.11	.55	.71	.78	—		
12. LEN-D	−.33	−.35	−.23	−.23	−.28	−.19	−.26	−.07	−.03	−.13	−.16	—	
13. LEN-E	−.39	−.42	−.32	−.32	−.33	−.24	−.22	−.05	−.09	−.09	−.23	.53	—

Note. IMA, imageability; CA, context availability; DEF, definition accuracy; CS, cognate status; FAM, familiarity; LOGF, log word frequency; LEN, length; D, Dutch words; E, English words. $p < .05$ if $r > .08$; $p < .01$ if $r > .11$.

bal-context-availability; see Footnote 3; length of the stimulus and response words in terms of number of syllables rather than letters), or the data associated with one variable were replaced by those of a conceptually similar variable (e.g., replacing the log word frequency variables by the absolute frequencies). The factor patterns emerging from these analyses were generally quite similar. When the number of factors was extracted by default always three or four factors emerged. One of them could clearly be identified as a Semantic factor and the second as a Familiarity factor. In case of three factors, the length variables loaded most highly on the third factor, but also Cognate Status (CS) loaded substantially on this factor. In case of four factors the pattern was the same, except that only the length variables now loaded substantially on the third factor. Cognate Status was the only variable to load on the fourth factor. When the number of factors to extract was fixed at four, the factor pattern was as when four factors were spontaneously extracted. Table 4 presents the rotated factor matrix (Varimax Rotation Solution) for the analysis with the number of factors to extract set at four and the data associated with the 13 predictors of Table 3 included.

TABLE 4
ROTATED FACTOR MATRIX

Variable	I	II	III	IV
IMA-D	.91	−.11	−.23	.05
IMA-E	.92	.00	−.23	.08
CA-D	.90	.06	−.10	.01
CA-E	.86	.31	−.05	.08
DEF-D	.88	−.07	−.18	−.02
DEF-E	.81	.39	.01	.07
CS	.09	.07	−.17	.98
FAM-D	.10	.79	.04	.02
FAM-E	.26	.87	.11	.10
LOGF-D	−.02	.88	−.11	−.01
LOGF-E	−.07	.88	−.22	.01
LEN-D	−.17	−.05	.84	−.14
LEN-E	−.26	−.08	.82	−.05

Note. IMA, imageability; CA, context availability; DEF, definition accuracy; CS, cognate status; FAM, familiarity; LOGF, log word frequency; LEN, length; D, Dutch words; E, English words.

All six semantic variables load substantially on Factor I, the Semantic factor. In addition, FAM-E and LEN-E load moderately on this factor. The four familiarity variables load most highly on the second factor, which can thus be identified as a Familiarity factor. As can be seen Context Availability and Definition Accuracy of the English words (CA-E and DEF-E) also load on this variable, but only moderately so. This supports the suggestion (see above) that an assessment of the familiarity of English words is taken into account by our subjects when asked to rate these words on how easily they can be defined or thought a context with. The two length variables load particularly highly on the third factor (the Length factor). Finally, cognate status is the only variable loading on the fourth factor (the Cognate factor). The four factors together account for 82.2% of the common variance.

Forward vs backward translation. In view of the asymmetry model, particularly noteworthy are the correlations between the dependent variables and the *semantic* predictors in backward translation (Table 2). As already pointed out above, only when conceptual memory is implicated in the translation process an effect of semantic variables may be expected. Consequently, the fact that the dependent variables and the semantic predictors in backward translation correlate significantly provides a strong indication that also when words are translated from L2 to L1 conceptual representations are accessed. It thus seems that a strong version of the asymmetry model, that would claim that conceptual memory is never involved in backward translation, can be rejected on the basis of these data. However, careful inspection of the correlations in Table 2 suggests that a weaker version, stating that in backward translation conceptual memory is less often implicated than in forward translation, may hold: except for the correlations between

Definition Accuracy of the English words (DEF-E) and the dependent variables, whenever there is a difference between a correlation of a semantic predictor and a dependent variable for forward translation and the corresponding one for backward translation, the correlation tends to be lower for backward translation. To test for significant differences between a given correlation coefficient for forward translation and the corresponding coefficient for backward translation, *z* scores were calculated on the basis of the Fischer *zs* of all such pairs of correlation coefficients involving a semantic predictor on the one hand (IMA-D; IMA-E; CA-D; CA-E; DEF-D; DEF-E) and a dependent variable on the other. Indeed, the coefficients within four of these pairs (out of a total of 18), all with Omission Score as the dependent variable (but see Experiment 2), differed significantly from one another (two-tailed tests), and all in the direction predicted by the (weak version of the) asymmetry model. They are listed in the left part of Table 5.

Z scores were also calculated for the Fischer *zs* of the pairs of coefficients (one coefficient for forward translation and one for backward translation) involving a nonsemantic predictor on the one hand and a dependent variable on the other. The differences that were significant, five out of the 21, are again listed in the left part of Table 5. From the direction of these differences one could conclude that word familiarity (the FAM and LOGF variables) affects backward translation somewhat more than forward translation and that the opposite holds for the length variables. Finally, we may tentatively conclude that the relatively large coefficients involving the (semantic) variable DEF-E in backward translation, may relate to the fact that this variable is correlated with the familiarity variables (see above).

The analyses above suggest that semantic variables play a slightly more important role in forward than in backward translation, supporting a weak version of the asymmetry model. Recall that a directional difference in translation RT—longer RTs in forward translation than in backward translation—has been regarded as a second source of support for the asymmetry model (Kroll & Stewart, 1990; 1994). So to provide a second test of the asymmetry model,

TABLE 5
SIGNIFICANT DIFFERENCES BETWEEN CORRELATION COEFFICIENTS FOR FORWARD AND BACKWARD TRANSLATION

Correlation between	Experiment 1			Correlation between	Experiment 2		
	Forward	Backward	Sign.		Forward	Backward	Sign.
IMA-D and OS	−.29	−.14	$p < .05$	IMA-D and RT	−.36	−.22	$p < .05$
IMA-E and OS	−.42	−.28	$p < .05$	IMA-D and OS	−.24	−.11	$p < .05$
CA-D and OS	−.40	−.24	$p < .01$	IMA-E and RT	−.47	−.34	$p < .05$
DEF-D and OS	−.29	−.14	$p < .05$	IMA-E and OS	−.38	−.26	$p < .05$
				CA-D and RT	−.46	−.33	$p < .05$
				CA-D and OS	−.36	−.22	$p < .05$
				DEF-E and ER	−.36	−.47	$p < .05$
FAM-D and RT	−.35	−.49	$p < .05$	FAM-E and RT	−.59	−.69	$p < .05$
FAM-E and ER	−.32	−.45	$p < .05$	FAM-E and ER	−.31	−.53	$p < .001$
LOGF-D and RT	−.41	−.52	$p < .05$	LOGF-E and ER	−.25	−.38	$p < .05$
LEN-D and OS	.21	.03	$p < .01$	LEN-D and OS	.18	−.02	$p < .05$
LEN-E and OS	.29	.06	$p < .001$	LEN-E and RT	.34	.17	$p < .01$
				LEN-E and OS	.24	.05	$p < .01$

Note. RT, reaction time; OS, omission score; ER, error rate; IMA, imageability; CA, context availability; DEF, definition accuracy; FAM, familiarity; LOGF, log word frequency; LEN, length; D, Dutch words; E, English words.

the overall RTs for forward and backward translation were calculated. In addition, the overall omission and error scores were calculated for both translation directions. The left column of Table 6 summarizes the results of these analyses. The values reported there for each translation direction are the overall means of the 440 individual stimulus means, each of them collapsed across maximally 26 subjects. Paired t tests (all dfs 439) were performed to see whether performance in forward translation differed from that in backward translation. The small RT-difference (23 ms) between translation directions was not significant (t = 1.44, p > .10). In contrast, the differences in omission and error scores (8.6 and 2.9%) were statistically reliable (t = 13.12, p < .001 and t = 4.52, p < .001, respectively). These effects were in the opposite direction: more omissions but fewer errors occurred in forward translation than in backward translation.

Unanimous responses. Recall that one of the constraints when selecting the Dutch words had been that they had a unanimous or a clearly dominant translation in English. In the norming studies only these unanimous or dominant translations, not the correct but weaker (mostly extremely weak) alternatives of the dominant translations, had been rated on the relevant characteristics. Yet, when calculating the overall translation RT of all correct responses to a stimulus word, these "unintended" correct responses were also included. As a consequence, the reported correlational analyses involving the predictor variables on the one hand (based on ratings of the unanimous and dominant responses only) and RT as the dependent variable on the other hand are not totally pure (none of the remaining analyses suffers from this imperfection). We may expect the effect of this inaccuracy to be negligible, because the proportion of unintended but correct translations was small (generally only one or two of all correct translations to the stimulus words that did not produce a unanimous response). Yet, in order to assess its effect more formally, the above analyses were performed two more times: once containing only the data associated with the Dutch stimuli to which all subjects gave the intended English translation (forward unanimous) and once containing only the data for the English stimuli to which only the intended Dutch translation response was given (backward unanimous). These analyses concerned 279 and 282 pairs of translation equivalents, respectively.

The main conclusions drawn above regarding the complete stimulus set also all hold for the forward and backward unanimous sets. For both subsets of the data, performance in forward and backward

TABLE 6

MEAN REACTION TIMES (IN MS), OMISSION SCORES (IN PERCENTAGES), AND ERROR RATES (IN PERCENTAGES) FOR FORWARD AND BACKWARD TRANSLATION FOR THE COMPLETE SET AND FOR SUBSETS OF THE DATA

	Complete set		Noncognates		Cognates		Abstract words		Concrete words	
	Expt 1	Expt 2	Expt 1	Expt 2	Expt 1	Expt 2	Expt 1	Expt 2	Expt 1	Expt 2
RT forward	1315	1197	1463	1315	1099	1008	1526	1406	1130	1011
RT backward	1292	1230	1390	1328	1111	1044	1436	1351	1127	1076
Effect	23	−33	73	−13	−12	−36	90	55	3	−65
OS forward	14.4	9.1	18.8	11.7	6.0	3.3	22.7	14.3	7.5	3.8
OS backward	5.8	4.2	7.6	5.3	2.0	1.6	8.5	6.2	1.9	1.5
Effect	8.6	4.9	11.2	6.4	4.0	1.7	14.2	8.1	5.6	2.3
ER forward	5.9	4.2	8.2	5.3	2.1	1.9	9.6	6.4	2.9	1.6
ER backward	8.8	6.6	12.9	9.0	3.4	2.5	12.5	9.4	5.3	3.4
Effect	−2.9	−2.4	−4.7	−3.7	−1.3	−0.6	−2.9	−3.0	−2.4	−1.8

translation was again negatively correlated with imageability, context availability, definition accuracy, cognate status, familiarity and log word frequency, but positively so with length. More specifically, the dependent variables again correlated significantly with all semantic predictor variables both in forward and in backward translation, suggesting an involvement of conceptual memory when translating in both directions.

The analyses comparing forward vs backward translation RT, OS, and ER within the forward and backward unanimous sets also showed the same results as the analyses encompassing the complete data set. For neither of the two subsets did the forward and backward translation RTs differ significantly from one another, whereas in both subsets the number of omissions was significantly larger and the number of errors significantly smaller in forward than in backward translation.

Differential Analyses

Comparisons between translation directions. Considering the tenet of the asymmetry model, the most salient outcome of the above analyses is that, statistically, responding was not faster in backward translation than in forward translation. Yet, the direction of the 23 ms effect that *was* obtained was as predicted by the asymmetry model. It is plausible that this small effect is not caused by stimuli randomly distributed over all types of words in the experimental materials, but that it is due to one or more stimulus categories of a circumscribed type. To determine which stimulus characteristics might be critical, we could split up our stimulus set according to a number of arbitrarily chosen stimulus dimensions, for instance, all predictors included is this study, and see when translation asymmetries in RTs occur. However, such an exploratory approach would inevitably produce an unwieldy set of data. Also, we would be bound to hit one or more Type-1 errors along the way, that, not acknowledged as such, might mislead us when interpreting the data. So instead we will only look at two variables we have reason to think could play a role in getting the asymmetry or not getting it: cognate status and imageability.

The reason for including cognate status in these analyses is that Sánchez-Casas et al. (1992) obtained a directionality effect on translation times for noncognates but not for cognates. To see whether this noncognate-specific directionality effect also occurred in the present study, we rank-ordered our stimuli according to their cognate-status and created two stimulus groups: one containing the 150 stimuli with the highest ratings (mean rating: 5.51) and a second including the 150 words with the lowest ratings (mean rating: 1.20). The same analyses as performed on the complete stimulus set were then done on these groups of cognates and noncognates separately. A directionality effect on translation times was indeed obtained for noncognates (t = 2.43, $p < .05$), but not for cognates (t = 0.53, $p > .10$). The corresponding means are reported in the middle columns of Table 6. The directionality effect on omission scores was significant for both cognates (t = 5.17, $p < .001$) and noncognates (t = 9.33, $p < .001$). The directionality effect on error scores was marginally significant for cognates (t = 1.83, p = .07) and significant for noncognates (t = 3.96, $p < .001$). For cognates as well as noncognates the direction of these effects was as for the complete stimulus set, that is, more omissions but less errors occurred in forward translation than in backward translation.

The stimulus materials of Kroll and Stewart (1990, 1994), the study that led them to develop the asymmetry model, were taken from eight semantic categories consisting of concrete exemplars. It is possible that Kroll and Stewart's directionality effects relate to their choice of exclusively concrete materials. To test this assumption, imageability (highly confounded with word concreteness; e.g., Paivio et al., 1968) was chosen here as the second variable to per-

TABLE 7
SIGNIFICANT DIFFERENCES BETWEEN CORRELATION COEFFICIENTS FOR FORWARD AND BACKWARD TRANSLATION

Correlation between	Experiment 1 Noncognates			Experiment 2 Noncognates		
	Forward	Backward	Sign.	Forward	Backward	Sign.
IMA-D and OS	−.44	−.15	$p < .01$	−.38	−.10	$p < .05$
IMA-E and OS	−.54	−.31	$p < .05$	−.48	−.26	$p < .05$
CA-D and RT	−.60	−.42	$p < .05$	−.62	−.42	$p < .05$
CA-D and OS	−.58	−.33	$p < .05$	−.56	−.29	$p < .01$
DEF-D and OS	−.42	−.19	$p < .05$	−.37	−.15	$p < .05$
	Cognates No significant differences			Cognates No significant differences		

Note. RT, reaction time; OS, omission score; IMA, imageability; CA, context availability; DEF, definition accuracy; D, Dutch words; E, English words.

form differential analyses on. We first rank-ordered all 440 stimulus words according to their word-imageability scores. Rather than choosing the imageability scores of the words in one of the two experimental languages as the basis for this rank-ordering, we first calculated the combined imageability score of each word on the basis of the z score transformations of the imageability of both its Dutch and its English form, and then entered these combined scores in the ranking procedure. We regarded this method suitable because the imageability of the Dutch and English words correlate highly ($r = .94$; see Table 3). Subsequently, two stimulus groups were formed, one consisting of the 150 words with the highest ratings (concrete words) and a second containing the 150 words with the lowest ratings (abstract words). Paired t tests were then done on the forward vs the backward RT, omissions and errors for the concrete and abstract word groups separately. The means for all conditions are presented in the right part of Table 6. A directionality effect on RT emerged, but surprisingly it was on the abstract words ($t = 2.73$, $p < .001$), and not on the concrete words ($t = .13$, $p > .10$). The by now familiar directionality effects on omissions and errors also occurred again ($t = 7.31$, $p < .001$, and $t = 10.20$, $p < .001$, for omissions on concrete and abstract words; $t = 2.51$, $p < .05$, and $t = 2.23$, $p < .05$, for errors on concrete and abstract words).

Finally, analogous to the analyses of the complete stimulus set, in addition to these differences in RT, OS, and ER between forward and backward translation, for the above subsets of noncognates and cognates the correlations between the dependent and the predictor variables were computed, both for forward and for backward translation of each subset. Subsequently, to test for significant differences between the correlation coefficients for forward translation and the corresponding coefficients for backward translation, z scores were calculated for the Fischer zs of all such pairs of correlation coefficients involving a *semantic* predictor variable on the one hand and a dependent variable on the other.[5] Table 7 lists all (and only) the pairs of coefficients for which a significant difference occurred.

The results converge with the above analyses on forward and backward translation RT. Recall that noncognates had

[5] These analyses were not performed on the subsets of concrete (high-imageability) and abstract (low-imageability) words because imageability itself is one of the semantic variables the effect of which we focus on in these differential analyses and it is a variable that is highly correlated with the other semantic variables of interest (definition accuracy and context availability). It may thus be expected that within the subsets of concrete and abstract words the relevant variables would hardly show any variance.

shown the directionality effect predicted by the asymmetry model: longer translation RT from L1 to L2 than vice versa (Table 6). In agreement with that result, in five cases (out of the 18) a semantic predictor variable correlated significantly higher with a dependent variable in forward translation than in backward translation (in one case a marginally significant effect in the same direction was obtained). An effect in the reverse direction never occurred. Unlike noncognates, cognates were translated equally fast in both directions (Table 6). Analogous to that outcome, in the present analyses in all cases a semantic predictor variable and a dependent variable correlated equally highly statistically in forward and backward translation. These correlation coefficients for cognates, for both forward and backward translation, were about equally large as those for noncognates in the backward condition (most of them centering around $r = .25$), suggesting that they are relatively often translated via the direct lexical route and equally often so in both directions.

Summarizing so far, it seems we can reject a strong version of the asymmetry model that would state that forward translation always involves conceptual mediation whereas backward translation never does. However, a weak version stating that at least for one category of words (noncognates) conceptual mediation occurs more often in forward than in backward translation is vindicated by the present set of data. The translation of cognates, in both directions, appears to come about via the direct lexical connections between word representations relatively often.

Comparisons within translation directions. In the introduction to this article it was suggested that for cognates the direct connections between the lexical representations of a pair of translation equivalents are stronger than they are for noncognates. At that point the notion of a directionality of these links had not been introduced yet. Taking the directionality into account now, the suggestion would be that for cognates not only is the lexical connection from an L2 word representation to an L1 word representation strong, as the asymmetry model claims, but that the lexical connection in the reverse direction is also strong. In fact, the data reported above (equal RTs for cognates in forward and backward translation and no significant differences between translation directions in the magnitude of the correlations between semantic predictor variables on the one hand and dependent variables on the other) suggest that the lexical links in the two directions are equally strong. The consequence will be that when cognates are translated, this nonconceptual route will be used relatively often in both directions. In contrast, as concluded above, the translation of noncognates more often proceeds via conceptual memory in forward translation than in backward translation. From this general picture the following predictions can be derived: (1) in forward translation the effects of semantic variables should be smaller for cognates than for noncognates (see also de Groot, 1992a); (2) the difference between the effects of semantic variables when translating cognates and noncognates should be smaller in backward translation than in forward translation. It is even plausible that for backward translation no such difference between cognates and noncognates exists at all; (3) the (equal or about equal) effects of semantic variables for cognates and noncognates in backward translation should be about equally large as these effects are for cognates in forward translation. To test these predictions, *within* the two translation directions the magnitude of the coefficients for the correlations between the various semantic predictors and the dependent variables in the cognate set of materials was compared with the magnitude of the corresponding coefficients in the set of noncognates. These analyses were based on the same collections of correlation coefficients as employed above, but the relevant comparisons differed. Whereas there the critical comparisons were *between* translation directions but *within* the sets of cognates or noncognates, here the critical comparisons were *between* cognates and noncognates but *within* translation directions.

The support for the first of the above predictions (forward translation) was substantial: out of the total of 18 comparisons (6 semantic predictors by 3 dependent variables), 11 were significant (at the 5% level or better) and six of the remaining seven were marginally significant. The direction of the effect was always as predicted, that is, the correlation coefficient for the noncognate materials was always larger than the corresponding coefficient for the cognate materials. As pointed out, this effect suggests a larger influence of semantic variables when the words to be translated are noncognates than when they are cognates. The results of these analyses are presented in Table 8 (significant effects only).

The second prediction was also supported convincingly: out of the 18 comparisons (again 6 semantic predictors by 3 dependent variables) made between cognates and noncognates within the backward-translation data set, only two were statistically reliable (both at the 5% level), and one was marginally significant. The significant contrasts are listed in Table 8. The differences that *did* materialize all once more showed a larger influence of semantic variables when noncognates are translated than when cognates are translated. The third prediction was also supported and was already discussed above, when presenting the comparisons *between* translation directions.

To resume, the data so far support a weak version of the asymmetry model in that forward translation of noncognates more often appears to proceed via conceptual memory than backward translation of noncognates. But conceptual memory seems also implicated, although to a lesser extent, in backward translation of noncognates. Contrary to the translation of noncognates, translating cognates appears to be symmetrical in that the involvement of conceptual memory is about the same when translating these words in both directions. Moreover, this involvement appears the same as when translating noncognates backward.

The next experiment once more investigates forward and backward translation of the same stimulus set, but with subjects expected to be slightly more experienced in

TABLE 8
SIGNIFICANT DIFFERENCES BETWEEN CORRELATION COEFFICIENTS FOR NONCOGNATES AND COGNATES IN EXPERIMENT 1

Correlation between	Forward			Correlation between	Backward		
	Noncognate	Cognate	Sign.		Noncognate	Cognate	Sign.
IMA-D and RT	−.48	−.22	$p < .05$	CA-E and OS	−.58	−.41	$p < .05$
IMA-D and OS	−.44	−.22	$p < .05$	DEF-E and OS	−.65	−.47	$p < .05$
IMA-E and RT	−.55	−.25	$p < .01$				
IMA-E and OS	−.54	−.27	$p < .01$				
IMA-E and ER	−.32	−.10	$p < .05$				
CA-D and RT	−.60	−.38	$p < .05$				
CA-D and OS	−.58	−.30	$p < .01$				
CA-D and ER	−.33	−.06	$p < .05$				
CA-E and OS	−.67	−.48	$p < .05$				
CA-E and ER	−.40	−.18	$p < .05$				
DEF-E and ER	−.39	−.11	$p < .05$				

Note. RT, reaction time; OS, omission score; ER, error rate; IMA, imageability; CA, context availability; DEF, definition accuracy; D, Dutch words; E, English words.

their L2, again English. If the subjects will indeed turn out to be more proficient in L2, it will be interesting to see whether and in what way this increased proficiency affects the data pattern.

Experiment 2

Subjects

Fifty-two new subjects were tested, 26 performing forward translation and 26 performing backward translation. Again, all subjects were first-year psychology students from the University of Amsterdam, with Dutch as their native language and English as their second language, and received course credit for participation. As mentioned before (see Introduction), these subjects were tested about 11 months after entering the university, on average 7 months later than the subjects in Experiment 1. Because their schooling at the university required them to read mainly in English, the subjects in Experiment 2 may be expected to be somewhat more fluent in English than those in Experiment 1. Whether this is indeed the case will be tested in an overall analysis of translation performance (in terms of RT, OS and ER) of the subjects in Experiments 1 and 2 (see below). The subjects' own assessment of their comprehension and production abilities in L2 (on a 7-point scale; 7 = same as in Dutch; see Experiment 1) indeed suggests a difference in L2 proficiency of the subjects between Experiments 1 and 2: the overall scores for Experiments 1 and 2 were 4.64 and 5.05, respectively. A 2 (Experiment: 1 vs 2) by 2 (Translation direction: forward vs backward) by 2 (Ability: comprehension vs production) ANOVA on the subjects' ratings showed a main effect of Experiment, $F(1,99) = 7.82, p < .01$ ($df = 99$ rather than 100 because for one of the subjects no ratings were available). The main effect of Ability was also significant, $F(1,99) = 56.63$, $p < .001$ (5.18 for comprehension and 4.52 for production). The main effect of Translation direction was not significant, nor was any of the interactions. The mean scores of the subjects in the forward condition of Experiment 2 were 5.46 (*SD* .76) for comprehension and 4.73 (*SD* .96) for production (7 = same as in Dutch); the corresponding scores of the subjects in the backward condition were 5.28 (*SD* .68) and 4.72 (*SD* .74).

Materials. The materials were identical to those of Experiment 1 except that the 18 words of which the data were excluded from Experiment 1 (see Experiment 1 Results) were not included in the stimulus set. The materials thus consisted of 440 Dutch words and their (dominant) English translations.

Apparatus and procedure. The experiment was run on an Apple Macintosh Plus computer with a voice-operated switch attached. Stimulus presentation and response registration were as in Experiment 1. As before, the stimuli were divided in two groups. Subjects alternately translated the stimuli of Group 1 followed by those of Group 2 or those of Group 2 followed by those of Group 1. Furthermore, subjects were alternately tested in the forward-translation condition or in the backward-translation condition. Thus, of every four further subjects one participated in each of the four stimulus-group-order by translation-direction conditions.

Results and Discussion

The subjects' English comprehension and production ratings suggest a higher L2 fluency level in Experiment 2 than in Experiment 1. To see whether these subjective assessments are substantiated by objective performance measures, overall mean RTs (collapsed across the RTs of all correct responses to the 440 stimulus words) were calculated for every subject in both experiments. A 2 (Experiment: 1 vs 2) by 2 (Translation direction: forward vs backward) by 26 (subjects) ANOVA on these mean RTs was then calculated. Indeed, the main effect of Experiment was significant, $F(1,100) = 5.67$, $p < .05$, and the effect was in the expected direction: translation times were longer in Experiment 1 than in Experiment 2. The main effect of Translation direction and the interaction between the two variables were not significant, $F < 1$ in both cases. Furthermore, this same ANOVA was run on the sum of the omission and error scores per subject. Again the main effect of Experiment was significant, $F(1,100) = 10.02$, $p < .01$, with more errors and omissions in Experiment 1 than in Experiment 2. The effect of Translation direction was marginally significant, $F(1,100) = 3.44$, $p = .07$, with slightly more errors and omissions in forward translation than in backward translation. The interaction did not approach significance ($p > .10$). Both analyses thus converge on the conclusion that the subject groups in Experiments 1 and 2 indeed differ in L2 proficiency.

The majority of all further analyses on the data of Experiment 2 parallel analyses performed in Experiment 1. The results are presented in the corresponding tables of Experiment 1 and the order in which they are discussed parallels that of the corresponding results of Experiment 1. The Results conclude with an overall subject analysis of Experiments 1 and 2. Generally, Experiment 2 replicated the results of Experiment 1. There is one clear exception, to be discussed shortly.

Table 1 shows that prediction was again quite good. The percentage of accounted variance on the multiple regression analyses varied between 23% and 63% on the various analyses. Again prediction was somewhat better in backward translation than in forward translation.

The correlations between the three dependent variables on the one hand and the 13 predictors on the other (Table 2) are on the whole very similar to the corresponding correlations in Experiment 1. Both the direction of the effects and the relative size of the coefficients are as in Experiment 1, and in most cases even the absolute size of the coefficients for Experiments 1 and 2 are about the same.

Table 5 again lists the *significant* differences between a correlation coefficient for forward translation and the corresponding coefficient for backward translation. The data again suggest that semantic variables affect forward translation somewhat more than backward translation and that the opposite holds for the familiarity variables: in six cases (out of the 18), a semantic predictor and a dependent variable correlated significantly higher in forward than in backward translation (in two cases a marginally significant difference in the same direction was obtained); in three cases (out of the 12) a familiarity predictor correlated higher with a dependent variable in backward than in forward translation. Yet, in one case a semantic predictor correlated *higher* with a dependent variable in *backward* translation. Note that here a "semantic" predictor is involved (DEF-E) that correlates moderately (but, relative to the remaining semantic predictors, highly; see Table 3) with one of the familiarity variables (FAM-E). As discussed before, this suggests that when instructed to rate *English* words on how easy they can be defined, our subjects take into account an assessment of the familiarity of these words. If true—and indeed the factor analyses support this idea (see Table 4)—it is no longer surprising that in this case the direction of the effect parallels the direction obtained with the familiarity predictors. Finally, as before, word length affected forward translation more than backward translation.

Table 6 shows the one result on which Experiment 2 notably differs from Experiment 1. It reports the differences between forward and backward translation in terms of RT, omission scores and error scores. The new omission and error data replicate Experiment 1: more omissions but fewer errors occurred in forward than in backward translation. Except for the effect on the error data in the cognate set, all these effects were statistically reliable on paired t tests ($p < .05$ or better). Note however that the omission and error scores are both lower in Experiment 2 than in Experiment 1, reflecting the higher L2 proficiency of the subjects in Experiment 2 (see also the earlier ANOVA over subject scores).

The deviating results concern RT. In Experiment 1 forward and backward translation had taken equally long statistically on the analysis of the complete stimulus set. The analysis on the corresponding data of Experiment 2 showed a significant difference between forward and backward translation RT ($t = 2.49$, $p < .05$; note that this effect was not significant on the ANOVA of the data by subjects). What is more, the direction of the effect is contrary to that predicted by the asymmetry model: RT was *shorter* in forward than in backward translation. The difference between forward and backward translation RTs was also significant, again with forward translation being *faster* than backward translation, in the analyses on two of the stimulus subsets, namely, on the cognates-subset ($t = 2.05$, $p < .05$) and on the subset of concrete words ($t = 4.28$, $p < .0001$). A further noteworthy finding is that, overall, responding was faster in Experiment 2 than in Experiment 1. This finding mirrors over items the outcome of the ANOVA on the subjects' mean RTs reported earlier.

Overviewing Table 6, it seems legitimate to conclude that, not surprisingly, increased fluency in L2 causes an increase in the speed with which words can be translated and a decrease in the numbers of errors and omissions that occur in the process. But a more interesting finding is also suggested by the data. It appears that with enhanced fluency more speed is gained in forward translation than in backward translation. The result of this differential effect is twofold: (1) In the material subsets where, with lower fluency, backward translation was faster than forward translation (noncognates and abstract words) the directional difference disappears; (2) in the subsets where, with lower fluency, backward and forward translation took equally long (cognates and concrete words) a directional difference emerges, with the shorter RT for forward translation.

This finding that forward translation gains more with increases in L2 fluency than backward translation and even overtakes backward translation in speed is not an isolated one. In our laboratory Swaak (1992) obtained the same effect, this time both over subjects and items. She had her subjects translate words that may be expected to be quite familiar to them in both Dutch and English, and hence easy to translate. All subjects could be expected to be quite fluent on these materials. Indeed, translation RTs were very short in that study. But more interestingly, a directional asymmetry occurred. Translation times were significantly *faster* for forward translation (887 ms) than for backward translation (947 ms).

Excluding unintended but correct responses. Instead of the analyses of only the unanimous responses in Experiment 1 (see there), a different but conceptually similar analysis was performed on the data of Experiment 2. Again the purpose was to remove the unintended but correct translations (for which no norming data were available) from the analyses. Rather than removing from the analyses *all* responses to Dutch and English stimulus words to which correct but unintended responses had been produced, now only these unintended responses themselves were omitted, retaining the intended responses to these "ambiguous" stimulus words. Subsequently, the above RT analyses were run once more on the remaining data (the omission and error data remain the same; obviously, an unintended but correct response should neither be categorized as an error nor as an omission). The results replicated those of the complete data set. The multiple regression analyses showed that our set of 13 predictors accounted for 53% of the variance in forward-translation RT and for 58% of the variance in backward-translation RT (cf. Table 1). The correlations between forward RT and backward RT on the one hand and the 13 predictors on the other hand were all very similar to, and never significantly different from, the corresponding correlations for the entire data set. Finally, on the new analyses the overall forward-translation RT was again shorter than the overall backward-translation RT (1170 and 1216 ms, respectively). As in the analysis on the complete data set of Experiment 2, this difference between translation directions was significant ($t = 3.38$, $p < .001$).

Correlational analyses on subsets of the materials. In the next set of analyses the correlations between the predictor variables and the dependent variables were calculated for the cognate and noncognate sets of materials separately, and the differences between a given correlation coefficient for forward translation and the corresponding coefficient for backward translation were computed. The right part of Table 7 reports all the *significant* differences involving a *semantic* predictor. The results closely replicate the corresponding findings of Experiment 1. Out of the 18 comparisons, five were significant in the noncognate set, the same five as in Experiment 1. As in Experiment 1, in all cases the coefficient was larger for forward translation. In two cases a marginally significant effect in the same direction was obtained. In the analysis on the cognate set, no significant differences occurred.

What is particularly noteworthy about this last set of analyses is that the pattern of results replicates that of Experiment 1 despite the difference between the two experiments in terms of directional effects on RT. In Experiment 1 both the directionality effect on RT (for some subsets of stimuli) and the sizes of the correlation coefficients for the two translation directions had supported the asymmetry model. Recall that the model predicts both larger semantic effects and longer translation times in forward than in backward translation. Here, only the former of the two effects showed up. In contrast, the latter effect reversed. This supports our view (see Introduction) that the directionality effect on RT obtained by Kroll and Stewart (1994) does not constitute a reliable indication of asymmetrical word translation.

The final set of analyses on the data of Experiment 2 concerned the comparison of cognates and noncognates *within* translation directions. Recall that we predicted the following three results (see Experiment 1 for further detail): (1) in forward translation the effect of semantic variables should be smaller for cognates than for noncognates; (2) the difference between the effects of semantic variables when translating cognates on the one hand and noncognates on the other should be smaller in backward translation than in forward translation, or it might even be absent in backward translation; (3) the effects of semantic variables for both cognates and noncognates in backward translation should be about equally large as these effects are for cognates in forward translation. All three findings were again obtained. Table 9 shows the first two of them.

Of the 18 correlations between a semantic predictor and a dependent variable, 6 were significantly higher for noncognates than for cognates in forward translation. In three cases a marginally significant difference occurred, again in the predicted direction. None of the remaining differences was significant. In contrast to the forward data, the backward-translation data showed no significant differences between corresponding correlation coefficients for noncognates and cognates. The latter finding suggests that in backward translation noncognates and cognates are translated the same way. Finally (not shown in Table 9), the third result also materialized, particularly when predictors involving semantic ratings of Dutch words (uncontaminated by a familiarity assessment) were involved.

TABLE 9
SIGNIFICANT DIFFERENCES BETWEEN CORRELATION COEFFICIENTS FOR NONCOGNATES AND COGNATES IN EXPERIMENT 2

Correlation between	Forward			Backward		
	Noncognate	Cognate	Sign.	Noncognate	Cognate	Sign.
IMA-D and RT	−.48	−.25	$p < .05$	No significant differences		
IMA-E and RT	−.56	−.27	$p < .01$			
IMA-E and OS	−.48	−.20	$p < .01$			
CA-D and RT	−.62	−.39	$p < .01$			
CA-D and OS	−.56	−.26	$p < .01$			
CA-E and OS	−.61	−.41	$p < .05$			

Note. RT, reaction time; OS, omission score; IMA, imageability; CA, context availability; D, Dutch words; E, English words.

Overall subject analysis of experiments 1 and 2. Although the present data refute a strong version of the asymmetry model in that semantic variables clearly affect backward translation, the data are consistent with a weak version of the model that would accept a larger role of semantic variables in forward than in backward translation. However, due to the correlational nature of this study all of the relevant analyses reported thus far were performed over items rather than subjects. To be able to generalize over subjects the finding that meaning affects forward translation more than backward translation, a final analysis was performed, now over subjects. For each subject in the forward and backward translation conditions of Experiments 1 and 2 two mean RTs were calculated, one for the 150 words with the highest imageability ratings and one for the 150 words with the lowest imageability ratings (ratings were based on z score transformations of the imageability of both the Dutch and the English form of the words; see before). These means were entered in a 2 (Experiment: 1 vs 2) by 2 (Translation direction: forward vs backward) by 2 (Imageability: high vs low) by 26 (subjects) ANOVA. Consistent with the weak asymmetry model, the interaction between Imageability and Translation direction was significant, $F(1,100) = 15.70$, $p < .0001$, the imageability effect being larger in forward translation (325 ms) than in backward translation (238 ms). But also the smaller effect in backward translation was significant ($p < .001$), which once more provides evidence against the strong asymmetry model. In addition to this interaction, two of the main effects were significant: responding was (87 ms) faster in Experiment 2 than in Experiment 1, $F(1,100) = 6.54$, $p < .05$, and high-imageability (concrete) words were responded to faster (by 281 ms) than low-imageability (abstract) words, $F(1,100) = 648.61$, $p < .0001$. The effect of Experiment was already reported before. It reflects the larger L2 fluency of the subjects in Experiment 2. The main effect of Translation direction was not significant ($F < 1$), as it had not been in Experiment 1. Except for the interaction between Imageability and Translation direction, none of the interactions was significant. The absence of a significant interaction between Experiment and Translation direction—despite the presence of such an interaction on the analysis over items—replicates its absence on the 2 (Experiment) by 2 (Translation direction) ANOVA on the average RT per subject (collapsed across *all* 440 stimulus words) reported at the outset of the Results Section of Experiment 2. A reason for its absence may be that in the present study, including far more stimulus words than subjects, effects on item analyses are more likely to be detected than effects on subject analyses. But recall that the relevant finding, that with high fluency levels forward translation becomes faster than backward translation, has been obtained before (Swaak, 1992; see above). As noted before, in that study the effect was reliable both over items and subjects.

Conclusion

All data reported in this study converge on the conclusion that both forward and backward translation are affected by semantic and familiarity variables and by the cognate relation between translation equivalents and their length. Together these (groups of) variables account for a large percentage of the variance in translation performance. In view of the asymmetry model of word translation (Kroll & Stewart, 1990; 1994) particularly relevant is the fact that meaning also plays a role in backward word translation. This finding evidences the involvement of conceptual memory in backward translation and thus refutes a strong version of the asymmetry model that would claim that conceptual memory is never implicated in backward translation. However, a weaker version *is* supported by the data.

Kroll and Stewart (1990, 1994) regard two of their findings as indicative of their asymmetry model: faster backward than forward translation and a role of meaning in forward but not in backward translation. The first of these symptoms of asymmetrical translation is not reliably present in this study. In Experiment 1, when the set of materials was analysed as a whole the effect did not turn up. However, when subsets of the materials were focused on, the predicted asymmetry emerged for noncognates and for abstract words (Table 6). In contrast, in Experiment 2, when analyzing the set of materials as a whole an asymmetry in translation time occurred, but backward translation turned out to be slower, not faster, than forward translation. This same effect also turned up when analyzing subsets of materials (cognates and concrete words).

In a weaker form than stated above, the second of the above two indications of asymmetrical translation *was* generally present in the data: in both experiments semantic variables played a weaker role in backward than in forward translation. This result was obtained for both the complete data sets of Experiments 1 and 2 (Table 5) as well as for a subset of the data sets (noncognates; Table 7). In all then, as the data by Kroll and Stewart (1994), the present data also point at a translation asymmetry, but the difference between translation directions appears to be quantitative rather than qualitative: conceptual memory is implicated in backward translation, but less often than in forward translation. From this follows a second conclusion, namely, that lexical memory is implicated more in backward translation than in forward translation.

Apart from the weaker role of semantic variables in backward translation, three further findings also suggest a directional difference in word translation. One is that the number of translation omissions was consistently larger in forward than in backward translation (Table 6). The second is that consistently more errors were made in backward than in forward translation (Table 6). Finally, the data weakly indicate that familiarity variables determine backward translation somewhat more than forward translation whereas the opposite holds for the length variables (Table 5). The first and second findings may relate to differences in the L1 and L2 vocabularies of our, unbalanced, bilinguals. Their L2 vocabulary is presumably smaller than their L1 vocabulary (see also Fig. 1), and it is likely that the subjects often possess faulty knowledge about L2-words that *do* have a representation in the L2-lexicon. If the to-be-produced word is unfamiliar or totally unknown, no response will be given. This will be relatively often the case if the intended response is an L2-word. Hence the larger number of omissions in forward translation. In contrast, if the subject makes a wrong guess about the meaning of the presented stimulus, but the translation of the wrongly guessed meaning is readily available, an error response will be produced. This situation will more often be the case in backward than in forward translation; hence the larger number of errors in backward translation.

A clue to the answer as to why familiarity might play a somewhat larger role in backward translation may be found in the conclusion just drawn that lexical memory is implicated more in backward than in forward translation. Assuming that lexical memory is at least one of the loci of familiarity effects, a larger role of familiarity in backward translation would then follow. Indeed it was suggested before (see Introduction) that familiarity/frequency may not only become reflected in the strength of the indirect connection (via conceptual memory) between the L1 and L2 lexical representations, but also—and particularly relevant here—in the strength of the direct connection (within lexical memory) between these representations.

One of the reasons why word length may affect word translation (see de Groot, 1992a, for more detail) is that preparing a retrieved response for production takes longer the longer the response. The directional effect of length (Table 5), can be understood if particularly the preparation of *English* response words is affected by length (maybe because of uncertainty about the pronunciation of these words). In the Subjects section of Experiment 1, data were reported suggesting that our subjects experience particular problems when producing English words: whereas lexical decision to common English words is only about 20 ms slower than lexical decision to these words in Dutch, English words in a pronunciation task take about 100 ms longer to initiate than do Dutch words. It is plausible that problematic response-preparation is the cause of this delay for English words, and that the problem is boosted with long words. This would cause the length effect of English words (LEN-E) in forward translation (where English is the response language). The length effect of Dutch words (LEN-D) in forward translation may then be understood from the correlation between LEN-E and LEN-D (r = .53; see Table 3). Indeed, when LEN-D is partialed out of the correlation between LEN-E and RT in forward translation, LEN-E still correlates significantly with RT (r = .23 and r = .28 in Experiments 1 and 2, respectively); in contrast, when LEN-E is partialed out of the correlation between LEN-D and RT in forward translation, the correlation between LEN-D and RT reduces to near zero (r = .09 and r = .04, in the same order).

It remains to be considered what could have caused the differences in the results of Kroll and Stewart's study (1990, 1994) and of this study. In the case of conflicting results between studies, three possible causes are commonly considered: differences between experimental procedures, subject populations, or materials. It is unlikely that the first of these is presently responsible. Several minor differences between Kroll and Stewart's procedure and ours can be pointed out, but they all seem to be immaterial. A difference between subject populations may have caused a difference between the results of the two studies, but the effects of subject differences are likely to have been very small. Kroll and Stewart drew the subjects from their critical experiment from our population of first-year psychology students at the University of Amsterdam and tested them in our laboratory. Testing was about seven months after the subjects had entered the university (cf. here: Experiment 1: four months; Experiment 2: 11 months). Therefore, Kroll and Stewart's subjects may be expected to have been slightly more proficient in L2 than our subjects in Experiment 1, but somewhat less so than our subjects in Experiment 2. It is possible that at least part of the differences in the directional effects of RT between the studies are due to this difference in time of testing and, hence, in population, because the present study suggested that this effect indeed responds to minor differences in L2 fluency of the subjects. But the most interesting difference between Kroll and Stewart's study and the present one, the effect of meaning in backward word translation here, but not in Kroll and Stewart's study, cannot be explained that way. The reason is that this effect occurred in both of the present experiments, and should therefore have occurred in Kroll and Stewart's study as well (with their subjects' L2 proficiency halfway between that of the subjects in our two experiments).

What remains as a likely cause of the differential effect of meaning in backward translation is a difference in the experimental materials. What may be critical is that our materials were selected to differ on many word characteristics, whereas Kroll and Stewart's stimuli all belonged to eight semantic categories of concrete examples. Blocking concrete words may somehow affect the likelihood of using the conceptual route to the translation response, and differentially so for different translation directions. However, it is as yet unclear why this should be so. A reason could be that similar meanings may cause interference. Especially in the categorized condition of Kroll and Stewart (where all words to be translated were taken from one and the same semantic category, e.g., they all referred to garments) the meaning similarity of the stimulus words may cause a tendency in the subjects to want to give the same response word to a number of different stimulus words. To prevent the errors that would thus be produced, they might adopt a strategy of ignoring meaning as much as possible. The strategy appears successful in backward translation (recall that there was no effect of list presentation—categorized or randomized—in backward translation), but not in forward translation, where an inhibitory effect of categorization indeed occurs.

That meaning affected backward translation in this study, employing this subject population, is not really surprising. At the time of testing, our subjects, though still unbalanced bilinguals, had had quite some experience in L2 comprehension. As a consequence, it may be expected that reasonably strong direct links from L2 lexical representations to conceptual representations will have developed, albeit these links are bound to be still weaker than the links from L1 lexical representations to conceptual representations. Similarly, the "production" links in the reverse direction, from conceptual memory to L2 lexical representations, may also be quite strong in our subjects. In fact, an interpretation of the finding that the subjects in Experiment 2 are faster in translating some subsets of the stimuli from Dutch to English than in the reverse direction (Table 6) would be that for these subjects and these materials the production links between conceptual memory and the L2 lexicon are even stronger than the comprehension links. The general speed-up of forward (and backward) translation in Experiment 2 compared to Experiment 1, despite the fact that the difference in L2 experience between the subjects in the two experiments was only 7 months on average, suggests that link strength responds to relatively minor changes in L2 experience.

Note that in this conception we have just replaced the one, weak, connection between the L2 lexicon and conceptual memory as encompassed in the asymmetry model (see Figure 1) by two links: one from conceptual memory into the L2 lexicon and one from the L2 lexicon into conceptual memory. Similarly, the connection between the L1 lexicon and conceptual memory as proposed by the asymmetry model could be split up in two, one for comprehension and one for production. In our subjects, the latter two types of links may be expected to be stronger than the corresponding links between conceptual memory and the L2 lexicon, but the differences between them are likely to diminish further with increased L2 experience, and they may cease to exist when the point of language balance is reached. It is plausible that at that point the rather small effects of translation direction as observed in this study also cease to exist. Indeed, with such balanced bilinguals it will no longer make sense to talk about "forward" and "backward" translation.

REFERENCES

ALTARRIBA, J. (1990). *Constraints on interlingual facilitation effects in priming in Spanish–English bilinguals.* Unpublished dissertation, Vanderbilt University.

BURNAGE, G. (1990). *CELEX: A guide for users.* Nijmegen, The Netherlands: SSN.

CHEN, H. -C., & LEUNG, Y. -S. (1989). Patterns of lexical processing in a nonnative language. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **15**, 316–325.

COLLINS, A. M., & LOFTUS, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review,* **82**, 407–428.

ERVIN, S. M., & OSGOOD, C. E. (1954). Second language learning and bilingualism. *Journal of Abnormal Social Psychology, Supplement,* 139–146.

GROOT, A. M. B. DE (1992a). Determinants of Word Translation. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **18**, 1001–1018.

GROOT, A. M. B. DE (1992b). Bilingual lexical representation: A closer look at conceptual representations. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning.* Amsterdam: Elsevier Science Publishers.

GROOT, A. M. B. DE (1993). Word-type effects in bilingual processing tasks: Support for a mixed representational system. In R. Schreuder & B. Weltens (Eds.), *The bilingual lexicon.* Amsterdam/Philadelphia: John Benjamins.

GROOT, A. M. B. DE, & NAS, G. L. J. (1991). Lexical representation of cognates and noncognates in compound bilinguals. *Journal of Memory and Language,* **30**, 90–123.

JIN, Y.-S. (1990). Effects of concreteness on cross-language priming in lexical decisions. *Perceptual and Motor Skills,* **70**, 1139–1154.

KEATLEY, C. W., SPINKS, J. A., & DE GELDER, B. (1994). *Asymmetrical* cross-language priming effects. *Memory and Cognition,* **22**, 70–84.

KOLERS, P. A. (1963). Interlingual word associations. *Journal of Verbal Learning and Verbal Behavior,* **2**, 291–300.

KROLL, J. F. (1993). Accessing conceptual representations for words in a second language. In R. Schreuder & B. Weltens (Eds.), *The bilingual lexicon.* Amsterdam/Philadelphia: John Benjamins.

KROLL, J. F., & CURLEY, J. (1988). Lexical memory in novice bilinguals: The role of concepts in retrieving second language words. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory: Current research and issues, volume 2.* London: Wiley.

KROLL, J. F., & SHOLL, A. (1992). Lexical and conceptual memory in fluent and nonfluent bilinguals. In R. J. Harris (Ed.), *Cognitive processing in bilinguals.* Amsterdam: North-Holland.

KROLL, J. F., & STEWART, E. (1990). *Concept mediation in bilingual translation.* Paper presented at the 31st Annual Meeting of the Psychonomic Society, New Orleans, November 1990.

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.

LOON-VERVOORN, W. A. VAN (1985). *Voorstelbaarheidswaarden van Nederlandse Woorden (Imageability ratings of Dutch words).* Lisse, The Netherlands: Swets & Zeitlinger.

MARTIN, W., & TOPS, G. A. J. (1984). *Van Dale Groot woordenboek Engels–Nederlands* (Van Dale large English–Dutch dictionary). Utrecht, The Netherlands/Antwerpen, Belgium: Van Dale Lexicografie.

NOBLE, C. E. (1953). The meaning–familiarity relationship. *Psychological Review,* **60**, 89–98.

PAIVIO, A., YUILLE, J. C., & MADIGAN, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monograph Supplement,* **76**, 1–25.

POTTER, M. C., SO, K.-F., VON ECKARDT, B., & FELDMAN, L. B. (1984). Lexical and conceptual representation in beginning and proficient bilinguals. *Journal of Verbal Learning and Verbal Behavior,* **23**, 23–38.

SÁNCHEZ-CASAS, R. M., DAVIS, C. W., & GARCÍA-ALBEA, J. E. (1992). Bilingual lexical processing: Exploring the cognate/non-cognate distinction. *European Journal of Cognitive Psychology,* **4**, 293–310.

SCHWANENFLUGEL, P. J. (1991). Why are abstract concepts hard to understand? In P. J. Schwanenflugel (Ed.), *The psychology of word meanings.* Hillsdale, NJ: Erlbaum.

SCHWANENFLUGEL, P. J., HARNISHFEGER, K. K., & STOWE, R. W. (1988). Context availability and lexical decisions for abstract and concrete words. *Journal of Memory and Language,* **27**, 499–520.

SCHWANENFLUGEL, P. J., & SHOBEN, E. J. (1983).

Differential context effects in the comprehension of abstract and concrete verbal materials. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **9**, 82–102.

SNODGRASS, J. G. (1984). Concepts and their surface representations. *Journal of Verbal Learning and Verbal Behavior*, **23**, 3–22.

SNODGRASS, J. G. (1993). Translating vs. picture naming: Similarities and differences. In R. Schreuder & B. Weltens (Eds.), *The bilingual lexicon*. Amsterdam/ Philadelphia: John Benjamins.

SNOW, C. E., & HOEFNAGEL-HÖHLE, M. (1978). The critical period for language acquisition: Evidence from second language learning. *Child Development*, **49**, 1114–1128.

SWAAK, J. (1992). Unpublished Masters Thesis (Stageverslag), University of Amsterdam, The Netherlands.

TAYLOR, I. (1976). Similarity between French and English words—A factor to be considered in bilingual language behavior? *Journal of Psycholinguistic Research*, **5**, 85–94.

TZELGOV, J., & EBEN-EZRA, S. (1992). Components of the between-language semantic priming effect. *European Journal of Cognitive Psychology*, **4**, 253–272.

WEINREICH, U. (1953). *Languages in contact: Findings and problems*. New York: Linguistic Circle of New York. (Reprinted in 1974 by Mouton, The Hague.)

(Received July 26, 1993)
(Revision received December 10, 1993)