Language Learning 45:4, December 1995, pp. 683–724

The Development of Bilingual Memory: Evidence from Word Translation by Trilinguals

Annette M. B. de Groot and John C. J. Hoeks
University of Amsterdam

We investigated the relation between foreign-language proficiency and multilingual lexicosemantic organization, using two sets of 48 unbalanced Dutch-English-French trilingual adults as participants. Dutch was the participants' native language. Of their two foreign languages English was the strongest. We tested a developmental hypothesis that assumes a "word-association" lexical structure for the native language and a relatively weak foreign language, here French, but a "concept-mediation" structure for the native language and a stronger foreign language, here English. Support for the hypothesis derived from the participants' performance in two versions of the word-translation task: "translation production" and "translation recognition". Translation was from Dutch to both of the foreign languages. The critical experimental manipulation was word concreteness. We hypothesized that a concept-

Annette M. B. de Groot, Department of Psychology; John C. J. Hoeks, Department of Psychology (now at the Interfaculty Research Unit for Language and Speech, Nijmegen, The Netherlands).

We thank Janet van Hell and the reviewers of this journal for helpful comments.

Correspondence concerning this article should be addressed to Annette M. B. de Groot, University of Amsterdam, Department of Psychology, Roetersstraat 15, 1018 WB Amsterdam, The Netherlands. Internet: pn_groot@macmail.psy.uva.nl.

mediation structure would predict an effect of this manipulation, whereas a word-association organization would not. In accordance with the developmental hypothesis, a clear concreteness effect obtained in Dutch to English translation, but not in Dutch to French translation. Overall, the data suggest that foreign-language proficiency indeed determines multilingual lexicosemantic organization.

Many studies on the organization of lexicosemantic knowledge in bilingual memory begin with the assumption that this organization is the same across different bilinguals and stable within a given bilingual. These studies typically contrast two opposing views on the bilingual lexicosemantic memory system. On the basis of the experimental data, they subsequently adopt one of the contrasting views and reject the second. However, accepting one and the same bilingual memory organization for all bilinguals is hard to reconcile with the results of an increasing number of studies demonstrating that a number of variables influence bilingual performance. If performance in these studies reflects the underlying memory structure (a common view), these studies' combined data strongly suggest that bilingual memory organization differs between bilinguals, that the memory structure of an individual bilingual reorganizes itself constantly over time (practice), and that within a bilingual memory different types of lexicosemantic structures coexist (see de Groot, 1995, for a review). The present study intends to provide additional support for that position, taking a rather new approach (testing trilingual participants). Our investigation focuses on the role of one of the variables that appear to affect bilingual memory organization: namely, the bilingual's (more appropriately here: the multilingual's) level of proficiency in the nonnative language (L2). But this focus exemplifies the more general point: that bilingual (multilingual) memory organization is not fixed within an individual nor uniform across individuals.

Studies that have looked at the relation between L2 proficiency and the organization of lexicosemantic knowledge in bilingual memory are those by Chen (1990), Chen and Ho (1986), Chen

and Leung (1989), Gekoski (1980), Kroll and Curley (1988), Mägiste (1984), Potter, So, von Eckardt, and Feldman (1984), and Tzelgov, Henik, and Leiser (1990). The prototypical design of such a study involves a between-subjects comparison of performance in bilingual processing tasks by groups of bilinguals that differ in L2 proficiency. Such was also the approach in the study by Potter et al. (1984), which was instrumental in drawing the attention of a larger number of researchers to the present issue.

Potter et al. (1984) tested three hypotheses about how bilinguals process words in their L2 and how these words are stored in memory. All three hypotheses distinguish between representations of word forms and representations of word meanings. The representations of word meanings (in conceptual memory) are shared between languages, whereas the representations of word forms (in lexical memory) are language specific. In other words, each member of a pair of translation equivalents (for instance, the English-French pair *flag-drapeau*) has its own representation in lexical memory, whereas the two are represented in a single, shared representation in conceptual memory. The "word-association" hypothesis (Figure 1) postulates direct connections ("lexical associations") between pairs of corresponding word-form representations (e.g., of *flag* and *drapeau*) in lexical memory. Understanding and producing L2 words (which involves conceptual access) proceeds via the associated first language (L1) word-form representations. In contrast, the "concept-mediation" hypothesis (Figure 1) denies the existence of direct connections between L1 and L2 word-form representations, but instead assumes direct connections between L2 word-form representations and the conceptual representations common to the L1 and L2 words. In this view, conceptual access during understanding and producing L2 is not mediated via L1 word-form representations, but proceeds as directly as do production and comprehension in L1. The third hypothesis, sometimes called the "developmental" hypothesis, takes an intermediate position: L2 learners start out developing and using lexical associations, but with further L2 practice, the direct connections between L2 word-form represen-

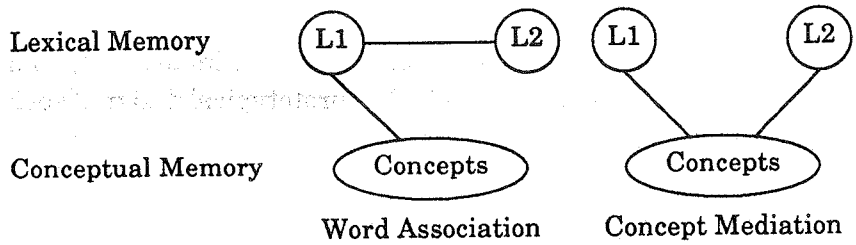

Figure 1. Two lexicosemantic organizations of bilingual memory.

tations and conceptual memory also get developed. The stronger they get, the more often these links will be used in L2 comprehension and production. The word-association links at the lexical representational level then gradually pass into disuse.

Potter et al. (1984) tested these hypotheses by having two groups of bilinguals, with different levels of L2 proficiency, translate words from L1 to L2 and name pictures in L2. They assumed that naming pictures in L2 (and, for that matter, in L1) always requires access of conceptual representations, whereas translating L1 words into L2 would only require conceptual access if the concept-mediation hypothesis were true. Given a word-association structure, translating between L1 and L2 could proceed via the lexical associations. They assumed the (longer) route to the translation response, via conceptual memory, to take more time than the shorter route via the lexical associations. Furthermore, because both require conceptual access, translating from L1 to L2 via conceptual memory and picture naming in L2 were thought to take equally long. In other words, shorter L1 to L2 translation times, compared to the time it would take to name pictures in L2, would provide support for the word-association hypothesis, whereas equal response times (RTs) in the two tasks would support the concept-mediation hypothesis. The developmental hypothesis would receive support if the former data pattern would be obtained for the less-proficient bilinguals, but the latter for the more-proficient bilinguals. In fact, both of Potter et al.'s groups of participants were equally fast in picture naming and translating in L2, and the authors accordingly opted for a concept-mediation

structure in both groups of participants. However, Kroll and Curley (1988) and Chen and Leung (1989), using the same design as Potter et al., indeed obtained the data patterns associated with the concept-mediation and word-association hypotheses in the high-proficiency and low-proficiency groups, respectively.

The merit of Potter et al.'s (1984) study is clear, as is the logic of its design. However, Potter et al.'s methodology, including picture naming as a task, constrains the selection of stimulus materials to concrete (picturable) words. Conclusions concerning bilingual representation and processing derived from these studies may not hold for other word classes as well (see e.g., de Groot, 1993). It is therefore desirable that the question of whether and how the structure of bilingual lexicosemantic memory relates to L2 proficiency be tackled from other experimental angles as well. Indeed, this has been done in a number of studies, employing the "free" and "restricted" word-association techniques (Gekoski, 1980), and bilingual versions of the Stroop interference task (e.g., Chen & Ho, 1986; Mägiste, 1984; Tzelgov et al., 1990). In the common monolingual Stroop task, participants name the ink color of words presented to them. The words themselves refer to colors. Performance depends on the congruence between the color to be named and the color the word refers to: In the case of congruence (e.g., the word *brown* printed in brown ink) performance is better than in the case of incongruence (e.g., the word *yellow* printed in brown ink). The bilingual version of the task includes additional conditions, in which the color words are presented in a language different from the response language (e.g., either of the French words *brun* or *jaune* in brown ink, with English as the response language). Most of these studies indeed demonstrated an effect of L2 proficiency on performance, which the authors interpreted in terms of different bilingual memory structures for the different proficiency levels.

In all of the above studies, L2 proficiency was manipulated between groups of participants. Abunuwara (1992) investigated the relation between foreign-language proficiency and, in his case, *multilingual* memory organization in a new and original way,

namely, by manipulating proficiency within a group of (Arabic-Hebrew-English) trilinguals. An obvious advantage of using a within-subject design is that a possible role of individual differences (e.g., age, family background, cultural differences, intelligence) in the results can definitely be ruled out. All of Abunuwara's participants had Arabic as their first and strongest language. Hebrew was stronger than English in the majority of the participants, but a substantial number had English as the stronger nonnative language. The comparisons in Abunuwara's study that are critical in the present context were those between the trilinguals' performance in conditions tapping their knowledge of L1 and the stronger nonnative language on the one hand, and of L1 and the weaker nonnative language on the other hand. The experimental techniques that he used were trilingual versions of the Stroop interference task and of Potter et al.'s (1984) procedure of comparing picture naming and word translation. Generally, the data were consistent with the studies using the corresponding bilingual paradigms. On the basis of his data, Abunuwara opted for a developmental model of bilingual memory.

As did Abunuwara (1992), we intend to test the relation between proficiency in a foreign language and multilingual lexicosemantic memory using trilinguals rather than bilinguals as participants. However, instead of using the multilingual version of the Stroop task or the technique employed by Potter et al. (1984) and others (comparing picture naming and word translation), we used simple word translation—in two versions (see below). De Groot (1992a), de Groot, Dannenburg, and van Hell (1994), and Kroll and Stewart (1990, 1994) showed that word translation on its own is a useful technique to study bilingual memory. The developmental course we focus on was tested by Potter et al. (1984), Kroll and Curley (1988), and Chen and Leung (1989): a development from a word-association structure to a concept-mediation structure. As argued by de Groot (1992a), de Groot et al. (1994), and Kroll and Stewart (1994), whenever manipulations of semantic variables affect performance in word translation, it can be concluded that conceptual memory is impli-

cated. In other words, an influence of such a manipulation would provide support for a concept-mediation structure. It is hard to conceive of a locus of a semantic effect other than conceptual memory, because word meanings are stored there. In contrast, a null-effect of semantic variables would constitute evidence against the concept-mediation hypothesis. What is more, because in this study we take the developmental hypothesis—which only considers the word-association and concept-mediation structures—as the starting-point, we will regard such a null-effect as support for the word-association structure. However, it should be borne in mind that there may be other memory structures, for instance, a "coordinate" structure (Weinreich, 1953/1974), that may also give rise to null-effects of semantic variables.

The semantic variable to be manipulated here is "word concreteness" or "word imageability". Word concreteness is generally defined as the extent to which the referent of the word can be experienced by the senses, whereas word imageability is defined as the extent to which the referent of the word evokes a mental image. The two variables, however, are highly correlated, and unless attempts have been made to disentangle them (an extremely laborious enterprise), the two can be used interchangeably. We have chosen to label the variable in question "concreteness" in the remainder of this text, because it is much more convenient to refer to "concrete" and "abstract" words than to call these same classes of words "high-imageability" and "low-imageability" words.

We selected our trilingual participants from a homogeneous population: first-year psychology students at the University of Amsterdam with Dutch as their native language and English and French as their only or strongest two foreign languages. Dutch is the language they typically use at home and in by far the majority of all other informal environments they find themselves in. It may thus be expected to be the strongest of their three languages. Our participants had typically started formal training in English and French at school, around the age of 12, and were trained in these languages for about 2 to 4 hours a week. For all participants both

English and French had been final examination subjects at secondary school. Nevertheless, we expect that these two languages would not be equally well established, but that English would be the stronger of the two. The reason is that outside school, in informal situations (listening to music; watching television) they had always been exposed to English much more than to French. Also, their university training typically involves the reading of many English but no French textbooks. Another reason for the expected dominance of English over French is that Dutch and English are more similar to one another (both Germanic languages) than are Dutch and French (the latter being a Romance language). It is plausible that learning a foreign language more similar to the native language taxes the learner less than learning a less similar foreign language. A likely consequence is that the former learning situation leads to a relatively rich stock of foreign-language knowledge.

Our reasons to expect French to be a weaker language in our participants than English introduce two potential confounding factors in this study: (a) English: school training *and* immersion learning; French: primarily school training; and (b) the family relation between Dutch and the nonnative language. We will come back to this issue later (see General Discussion) to argue that it does not jeopardize the more general goal of this study. For the moment it suffices to say that these confoundings are extremely hard to prevent, if at all. Their expected presence in the majority of studies that manipulate the participants' proficiency in a foreign language provides a reason to approach the present research question from different angles and employing different populations of participants, and to look for converging outcomes of the various types of study. The present study fits well in this approach.

The conclusion—on the basis of the above casual observations—that our participants should be more proficient in English than in French, and most proficient in Dutch, will be substantiated in pretests that compare their performance across languages. The pretests involve relatively simple word-processing

tasks: lexical decision, in which participants categorize letter strings as words or nonwords, and semantic access, in which they attempt to retrieve the meaning of a presented word. Our assumption concerning the relative proficiency of our participants in the three languages would be supported if lexical decision or semantic access took less time in Dutch than in English, and less time in English than in French.

The two versions of the main task we used, word translation, are "translation production", in which participants are presented with words in one language and are asked to produce their translation in the target language, and "translation recognition", where pairs of words are shown, the words in each pair in different languages, and the participants have to decide whether or not the words are translations of each other (see also de Groot, 1992a, Experiment 2, and de Groot & Comijs, 1995). In our translation-production study (Experiment 1) the participants translated Dutch words into English (one condition) or into French (the other condition). In our translation-recognition study (Experiment 2) they decided whether Dutch-English word pairs (one condition) and Dutch-French word pairs (the other condition) consisted of translation equivalents.[1]

Translation production from the native and strongest language to a nonnative and weaker language has been referred to as "forward" translation (de Groot et al., 1994). We opt for forward rather than "backward" translation (in the opposite direction) because two recent studies (de Groot et al., 1994; Kroll & Stewart, 1994; but see La Heij, Hooglander, Kerling, & Van der Velden, in press) suggest that this way we would maximize our chances to obtain a developmental effect. More specifically, these studies showed that, as compared to backward translation, reasonably fluent bilinguals effectuate forward translation through conceptual memory relatively often. In contrast, in backward translation these bilinguals employ the word-association connections in lexical memory relatively often. These are the connections that, according to the developmental hypothesis (e.g., Chen & Leung, 1989; Kroll & Curley, 1988), are typically used by (adult) bilinguals

with a low level of proficiency in L2. In other words, backward translation by more-proficient bilinguals may be more similar to the way less-proficient bilinguals translate, in both directions, than is forward translation. Consequently, even if the developmental hypothesis were correct, a differential effect of the concreteness manipulation across the proficiency conditions (here: Dutch-English vs. Dutch-French within the same participants) might not show up in backward translation. The different proficiency conditions might show the same data patterns, and we could, as a consequence, be tempted to reject the developmental hypothesis, but for the wrong reasons and possibly erroneously.

In translation recognition the response is likely to be based on the result of some intersection of forward and backward processing. This means that translation recognition involves a backward component. If this is true, and if the above analysis of backward translation is correct, proficiency effects are less likely to occur in translation recognition than in forward translation production. We nevertheless include translation recognition because we fear Condition Dutch to French in translation production might turn out to be quite difficult for our participants and, consequently, produce many missing data. Presumably, when a retrieval route between a pair of translations is too weak a memory trace to be used successfully in translation production, it will often still be possible to exploit it in translation recognition.

We mentioned that an effect of a semantic variable, here concreteness, would suggest the involvement of conceptual memory in word translation and thus support the concept-mediation hypothesis, whereas the absence of such an effect would support the word-association hypothesis. We know from earlier studies (de Groot, 1992a; de Groot et al. 1994; Kroll & Stewart, 1994) that Dutch participants with a similar knowledge of English (and a similar history of acquiring this knowledge) as our intended participants show semantic effects when translating Dutch words into English in a translation-production task, evidencing concept mediation. We therefore expect this effect to occur again here. The interesting new question in this study is whether or not they

will show a concreteness effect when they translate from Dutch to French, their weaker nonnative language. According to the developmental hypothesis, in the latter condition the effect should not occur (in this case translation still always comes about via word association) or be smaller (in the case of some intermediary stage between pure word association and maximal concept mediation) than when translating is from Dutch to English. These combined effects (a clear concreteness effect when Dutch words are translated into English and a smaller or null-effect of concreteness when the same participants translate Dutch words into French) would support the developmental hypothesis. Finally, as set forth above, such support may be harder to obtain in translation recognition. Nevertheless, any effect, significant or not, should go in the same direction as in translation production, and the data of the translation-recognition condition should thus provide converging evidence.

Experiment 1: Translation Production

Method

Participants. Forty-eight first-year psychology students from the University of Amsterdam participated. Twenty-four of them were selected on the basis of a questionnaire that was completed by all students in a cohort of first-year psychology students (about 700 in all). The selected students all had Dutch as their native language and had had both English and French as final examination subjects at secondary school. In selecting the remaining 24 participants, we followed a less laborious procedure. Forms on a notice board invited first-year psychology students of the University of Amsterdam with Dutch as their native language and both English and French as final examination subjects at secondary school to participate. To get a first, rough estimate of a proficiency difference between English and French we asked each participant, upon arrival in the laboratory, to assess on a 7-point scale his or her comprehension and production levels in

English and French. They were to tick Point 7 on the scale if they considered the skill in question equally well developed as the corresponding skill in Dutch, and Point 1 if they considered it totally lacking. For all participants we calculated a proficiency score for English and French. A score was the mean of the comprehension and production scores of that participant for a given language. As expected, the English proficiency score was higher (M=5.54; SD=0.74) than the French score (M=4.43; SD=1.13). The difference was significant, $F(1, 47)=69.86; p<.0001$).

Pretest. We also assessed the proficiency difference between English and French, and between each of these and our participants' L1, Dutch, more formally by testing all 48 participants in a lexical-decision pretest. This pretest consisted of three parts, differing from one another with respect to the language, Dutch, English, or French. In each part, we presented the participants with letter strings on a screen, one string after the other, and asked them to decide whether or not the string was a word in the language of that test. The participants signaled their responses by pushing one of two buttons on a response box, one for *yes*, and the second for *no*. The order in which we collected the data of the three language parts alternated among the participants. The test materials consisted of three groups of 50 words (all nouns), one of Dutch words, the second of English words, and the third of French words. We used different words in the three language conditions, but across languages the words were matched on log word frequency and concreteness, two variables known to affect lexical-decision performance. In addition to the 50 words, each group of materials contained 40 nonwords. The nonwords within a language condition complied with the orthography and phonology of the language of that condition: They were Dutch-like, English-like, and French-like. Each set of materials also included 25 letter strings to be presented as practice materials prior to the actual test materials. The results of this pretest are summarized in Table 1, collapsed across the two recruitment procedures (see above). Preliminary analyses had shown that there were no differences among the patterns of results for the participants

Table 1
Mean Lexical-Decision Times and Error Rates in the Proficiency Pretest of Experiment 1. Mean Reaction Times are Given by Participants and Items.

	Participants			Items			
	Reaction Time (ms)	*SD*	*N*	Reaction Time (ms)	*SD*	*N*	Error Rate (%)
Dutch							
Words	526	46	48	527	39	50	1.5
Nonwords	654	60	48	655	36	40	4.4
English							
Words	629	54	48	639	82	50	11.0
Nonwords	728	69	48	733	50	40	15.1
French							
Words	658	61	48	670	75	50	12.1
Nonwords	795	78	48	799	63	40	23.8

recruited by the different procedures. In Table 1 (and in Tables 2–4) mean RTs are given by participants and items. As can be seen, lexical decision was fastest and most accurate in Dutch, and slowest and least accurate in French, with performance in Condition English in between. The differences between the three language conditions were all significant on the analysis of both the reaction time data and the error data. This pattern of lexical-decision data thus supports our assumption that French was a weaker language than English, and that Dutch was our participants' strongest language.[2]

Materials. The set of translation-production test materials contained two groups of 40 Dutch words (all nouns, different words in the two groups). Twenty words of each group were concrete, whereas the remaining 20 were abstract. We selected the stimulus words from a Dutch corpus of word-imageability ratings (Van Loon-Vervoorn, 1985). The mean imageability ratings of the concrete words in the two word groups were 6.49 (SD=0.33) and 6.46 (SD=0.34). The mean ratings for the abstract words were 2.96 (SD=0.58), and 3.10 (SD=0.55), in the same order. All ratings had been done on a 7-point scale, where a 1 indicated "the word arouses a mental image with difficulty or not at all", and a 7 indicated "the word arouses a mental image quickly and easily". In both word groups the ratings for the concrete words differed significantly from those for the abstract words (p<.0001 in both cases). Across the two stimulus sets (concrete and abstract) per word group and across the two word groups, we matched the words on log word frequency and "cognate status", two variables known to affect translation performance (e.g., de Groot et al., 1994; Sánchez-Casas, Davis, & García-Albea, 1992). The frequencies were taken from the CELEX frequency count (Burnage, 1990; a count based on a corpus of 42.5 million Dutch words). The variable cognate status refers to the perceptual similarity of translation equivalents. Similar translations (e.g., the Dutch-English pair *appel-apple*) are called "cognates"; dissimilar pairs are called "noncognates". For the present study we simply rated, independently of one another, pairs of Dutch-

English and Dutch-French translations on a 3-point scale as to how similar we thought the words within each pair were. The first words in these pairs were the two groups of test materials for the translation-production study. The second words were their translations in the target language. The ratings involved a combined assessment of both spelling and sound similarity of the pairs under consideration. A second rating by us involved a binary classification of the translation pairs (cognates vs. noncognates). This second procedure resulted in a perfect overlap of the ratings of the two judges; the first resulted in a nearly perfect overlap (r=.97; see de Groot & Nas, 1991, and Kroll & Stewart, 1994, for different procedures to determine the cognate status of translation pairs). Within the two groups of words, concrete and abstract words did not differ significantly with respect to log word frequency and cognate status. Furthermore, the two word groups did not differ significantly from one another with respect to these two variables. All test words, together with their ratings on cognate status, imageability, and log word frequency, are listed, in English, in Table A–1. In addition to the test stimuli, we presented 25 nouns for practice in both translation conditions.

Apparatus and procedure. The experiment was run on an Apple Macintosh Plus Computer. A PASCAL-program controlled the stimulus presentation and the recording of the RTs. The participants' responses were recorded by a microphone that activated a voice-operated switch. The translation-production task started after the participant had finished the lexical-decision pretest. The participant sat facing the screen and the experimenter sat to the left of the participant, monitoring the workings of the voice switch. We instructed the participants to speak out loud the translation of the stimulus word as quickly as possible while making as few errors as possible, and to remain silent if they did not know the translation. The experimenter noted failures of the voice switch to respond to the participant's response and triggering by another sound. The sequence of events on each trial was as follows: Prior to a stimulus word, a fixation stimulus (an asterisk) appeared on the screen, slightly above where the stimu-

lus word was to appear. After it disappeared, the stimulus word immediately appeared and remained on the screen until the voice switch registered the onset of the participant's response (or of any other sound). The duration between the onset of the stimulus word and the moment a sound triggered the voice switch was registered as the RT. The experimenter then typed the participant's response, if a response was given (what was being typed did not appear on the screen). Finally, the experimenter touched the RETURN key, 1 second after which the next trial started. The maximum presentation duration of a stimulus was 8 seconds. Whenever this duration expired without a response having been given by the participant, the experimenter typed the word *none* and called the next trial by hitting the RETURN key. The stimuli were presented in blocks (one block of 25 practice stimuli followed by two blocks of 20 experimental stimuli). After each block, the participant got a brief rest before the experimenter initiated the presentation of the next block. We presented the stimuli in a random order, different for every participant. The order of the two translation sessions, Dutch to English and Dutch to French, alternated among the participants.

Results and Discussion

For each of the 48 participants we calculated four mean RTs, one for each of the four conditions formed by the two levels of the variable concreteness (concrete vs. abstract) and the two levels of the variable translation condition (Dutch to English and Dutch to French). In calculating these means, we excluded RTs associated with translation errors and voice-switch registration errors (the latter occurred only on a small percentage of trials), and RTs shorter than 400 ms. We excluded RTs shorter than 400 ms because from previous word-translation studies we knew that generating a translation response within this short amount of time was virtually impossible; RTs shorter than 400 ms had to be due to premature responding or faulty registration. We regarded a response as an error when it was not listed among the transla-

tions of the stimulus word in a popular translation dictionary (of course, different dictionaries were consulted for the different translation conditions). We performed a 2×2 (concreteness × translation condition) ANOVA on these mean RTs. In this analysis we treated both variables as within-subject variables. We also ran the corresponding 2×2 ANOVA on the mean RTs for all stimulus words, collapsed across (the correct responses of) all 48 participants, treating both variables as between-items variables. We also performed this same pair of analyses on the numbers of translation errors per participant and per item. Finally, we executed this pair of analyses on the numbers of omissions per participant and per item, that is, on the number of times no translation was given before the maximum stimulus duration (8 seconds) expired. A summary of the results of these analyses is presented in Table 2.

On the RT analyses both main effects were significant: concreteness, $F_1(1, 47)=13.68$, $p<.001$, and $F_2(1, 76)=5.24$, $p<.05$; translation condition, $F_1(1, 47)=113.83$, $p<.0001$, and $F_2(1, 76)=44.10$, $p<.0001$. Concrete words were translated faster than abstract words, and translating from Dutch to English was faster than from Dutch to French. The difference in RT between translating from Dutch to English on the one hand and from Dutch to French on the other once more suggests that our participants were more proficient in English than in French.

The interaction between concreteness and translation condition was also significant, $F_1(1, 47)=59.79$, $p<.0001$, and $F_2(1, 76)=5.32$, $p<.05$. Pairwise comparisons on the analyses by participants showed that the concreteness effect was significant when translating from Dutch to English ($p<.0001$) and from Dutch to French ($p<.01$). But the effects in Condition Dutch to English and in Condition Dutch to French went in opposite directions. Pairwise comparisons on the analyses by items showed a significant concreteness effect in Condition Dutch to English only ($p<.01$). As shown in Table 2, on this analysis concreteness had absolutely no effect when Dutch words were translated into French ($p=.99$; see note 2). In all, these data suggest that the concreteness effect

Table 2

Mean Response Times, Error Rates, and Omission Scores for All Translation by Concreteness Conditions in Translation Production (Experiment 1). Mean Reaction Times are Given by Participants and Items.

	Participants			Items				
	Reaction Time (ms)	*SD*	*N*	Reaction Time (ms)	*SD*	*N*	Error Rate (%)	Omission Score (%)
Dutch to English								
Concrete	971	192	48	994	427	20	1.1	1.7
Abstract	1,519	349	48	1,542	399	20	4.6	9.8
Effect	548			548			3.5	8.1
Dutch to French								
Concrete	2,017	584	48	2,060	525	20	10.5	15.6
Abstract	1,816	448	48	2,058	719	20	7.7	25.3
Effect	–201			–2			–2.8	9.7

indeed depends on the level of proficiency of a participant in his or her foreign language(s). However, the fact that, on the subject analysis, a concreteness effect was obtained in Condition Dutch to French, and that this effect was due to abstract words being processed faster than concrete words, is incompatible with both the word-association and the concept-mediation views of word translation. The former predicts a null-effect of concreteness and the latter predicts the opposite from what is obtained: It predicts better performance for concrete words (see General Discussion for more detail). However, the analyses of the omission data, to be discussed shortly, throw a different light on this unexpected concreteness effect when Dutch words were translated into French.

On the analyses of the translation errors, of the main effects only that of translation condition was significant on both analyses, $F_1(1, 47)=43.79, p<.0001$, and $F_2(1, 76)=16.42, p<.0001$: Fewer errors were made in Condition Dutch to English than in Condition Dutch to French. The difference between the conditions again shows our participants' greater competence in English than in French. The main effect of concreteness was not significant ($F<1$ on both analyses). Finally, the interaction between the two variables was significant, $F_1(1, 47)=12.82$, $p<.001$, and $F_2(1, 76)=4.10$, $p<.05$. Pairwise comparisons of the means of the participant analysis showed that the concreteness effect was statistically reliable ($p<.05$ or better) in both translation conditions. Again, the effect in Condition Dutch to French went in a direction opposite from that in Condition Dutch to English. However, on the item analysis the concreteness effect was significant in neither of the two translation conditions.

So far it seems that in Condition Dutch to French translating abstract words is somewhat easier than translating concrete words. But, as anticipated above, the analyses of the omission data throw a different light on the observed unexpected concreteness effects in this condition. These analyses produced a significant main effect of translation condition, $F_1(1, 47)=106.48, p<.0001$, and $F_2(1, 76)=17.10, p<.0001$. Fewer omissions occurred in Condition Dutch to English (5.7% overall) than in Condition Dutch to

French (20.5% overall), again showing the differential proficiency in English and French of our participants. The main effect of concreteness was also significant, $F_1(1, 47)=62.27$, $p<.0001$, and $F_2(1, 76)=6.24$, $p<.05$. Finally, the interaction between concreteness and translation condition was not significant, $F_1(1, 47)=1.04$, $p>.10$, and $F_2(1, 76)<1$.

Particularly noteworthy is the high number of omissions when Dutch abstract words were translated into French. This finding warns us against accepting the relatively short RT (but only on the analysis by participants) and small number of errors in that condition, both as compared to the corresponding concrete materials, as an indication that for some unexpected reason these abstract words were relatively easy to translate. The large number of omissions in this condition suggests that they were not. Assuming that the words hardest to translate typically result in an omission, the RT and error data in the Dutch to French condition with abstract words concern relatively many (as compared to the corresponding condition with concrete words) easy words. The assumed larger proportion of difficult words included in the RT and error data of the concrete words might then have resulted in the slower and more often incorrect responses to these words. But whether or not our interpretation of these aspects of the data is correct, our participants clearly showed different response patterns in the two translation conditions. This fact per se suggests different lexicosemantic memory structures for the Dutch-English and Dutch-French knowledge stores of our participants.

Overall, the translation-production data support the idea that the level of proficiency in a foreign language determines whether or not a concreteness effect occurs when words are translated from the native to that foreign language. Here, a robust effect occurred only when the participants translated from Dutch to English, the stronger of their nonnative languages. The data in this condition are consistent with the concept-mediation hypothesis, whereas those in Condition Dutch to French fit the word-association hypothesis better.

Experiment 2: Translation Recognition

Method

Participants. Forty-eight first-year psychology students from the University of Amsterdam participated. None of them had participated in Experiment 1. They all had Dutch as their native language and had had both English and French as final examination subjects at secondary school. (They had indicated this to be the case on forms on a notice board; cf. Experiment 1.) Again, to get a first estimate of a proficiency difference between English and French, we asked each participant upon arrival in the laboratory to assess on a 7-point scale his or her comprehension and production levels in English and French. (See Experiment 1 for details.) As expected, the English proficiency score was higher (M=5.35; SD=0.73) than the French score (M=4.03; SD=1.11). The difference was significant, $F(1, 47)=72.12$, $p<.0001$).

Pretest. We also again assessed more formally a proficiency difference between English and French, and between each of these two languages and our participants' native language Dutch. This was done with a lexical-decision pretest for half of the participants. The procedure and stimulus materials in this pretest were as used in Experiment 1. The remaining 24 participants were pretested in a "semantic-access" task. Like the lexical-decision pretest, it consisted of three parts, the parts differing in language, Dutch, English, or French. Every participant performed in all three parts, and the order of the parts rotated among the participants. In each part we presented the participants with a series of words, one word per trial, and required them to push a response button the moment they felt they had retrieved the meaning of the word presented on that trial. We stressed the button should be pushed as quickly as possible after meaning retrieval, that pushing the button before the word's meaning had actually been retrieved should be avoided, and that the button should not be pushed if the presented word was not known. The reason to include the semantic-access task as a new pretest was

Table 3

Mean Lexical-Decision and Semantic-Access Times and Corresponding Error and Omission Rates in the Proficiency Pretest of Experiment 2. Mean Reaction Times are given by Participants and Items.

	Participants			Items				
	Reaction Time (ms)	*SD*	*N*	Reaction Time (ms)	*SD*	*N*	Error Rate (%)	Omission Rate (%)
Lexical Decision								
Dutch								
Words	545	50	24	546	43	50	1.8	
Nonwords	665	69	24	667	40	40	4.3	
English								
Words	645	47	24	659	95	50	10.4	
Nonwords	731	66	24	733	52	40	12.5	
French								
Words	643	46	24	654	82	50	11.0	
Nonwords	784	60	24	785	62	40	24.1	

Semantic Access							
Dutch							
Easy	535	90	24	536	46	50	0.2
Difficult	671	140	24	649	125	15	36.9
English							
Easy	786	176	24	804	191	50	11.4
Difficult	1,064	229	24	1,095	248	15	69.2
French							
Easy	846	177	24	906	274	50	15.6
Difficult	1,051	217	24	1,193	295	15	69.7

simply to obtain yet a further indication of a difference between proficiency in English and in French in our population of participants. In a task like this, where essentially there are no wrong responses, the participant could push the response button at odd moments. That participants do not behave erratically, however, is suggested by a number of other studies using similar tasks (e.g., Paivio, 1966; Snodgrass & Vanderwart, 1980). The test materials consisted of three groups of 65 words each (all nouns), each group consisting of Dutch, English, or French words exclusively. Fifty of the words in each language condition were the same 50 words as presented in the corresponding lexical-decision conditions. In addition, each language condition included 15 words, in the language of that condition, selected to be unknown by many of the participants. We included these "difficult" words to provide a means to check whether the participants complied with the instructions. If indeed they did, the responses to these words should take relatively long and these words should show a large number of omissions. Finally, each language condition included 18 practice trials. The results of the lexical-decision and semantic-access pretests are summarized in Table 3. As can be seen, in one respect the lexical-decision data differ from those of the participants in Experiment 1 (see Table 1): English and French words were processed equally fast. But another aspect of the data suggests that it would be incorrect to conclude from this that the present participants were equally proficient in English and French: The main effect of language, collapsing across words and nonwords, showed a significant difference between Conditions English and French ($p<.01$), both in terms of RT and in terms of percentage of errors. This finding suggests a proficiency difference between English and French. Collapsing across the easy and difficult words, the semantic-access data clearly suggest that the participants in that condition were less proficient in French than in English. All differences between the three language conditions were statistically reliable on both the RT and the omission analyses. But the difference between French and English words materialized only for easy words. That participants were not

slower on French difficult words than on English difficult words suggests that participants in this task set a deadline (presumably implicitly) for responding. Overall, the pretest-data of Experiment 2 do suggest that our participants were more fluent in English than in French (and most fluent in Dutch).

Materials. The test materials contained two groups of 80 word pairs each. In one of the groups the word pairs all consisted of a Dutch and an English noun; in the second they all consisted of a Dutch and a French noun. Within 40 of the word pairs of each group the second word was the translation of the first. These stimuli thus required a *yes* response. The first words of these "positive" pairs were the test materials of the translation-production experiment (Experiment 1). Within the remaining 40 of the word pairs of each group, the "negative" pairs, the second word was not a translation of the first word. These stimuli thus required a *no* response. A consequence of using the materials for translation production as the basis for constructing the present groups of positive word pairs was that half of the stimuli in the positive material sets consisted of concrete words, whereas the other half contained abstract words only. Therefore, we selected negative pairs that also varied on this dimension. A further constraint when selecting the materials for the negative trials was that a number of them consisted of "pseudocognates". A pseudocognate pair consists of two words, one in one language and the second in the other, that resemble one another perceptually but that have totally different meanings (e.g., the English-French pairs *crime-crise*, *anger-ange*, and *guest-geste*). The stimulus sets for translation production had included both words with a cognate translation in the response language and words with a noncognate translation. Consequently, the present positive set also partly consisted of cognates and partly of noncognates. To prevent the participants' deliberately or inadvertently using perceptual similarity between the words in the word pairs as a cue to the response (being inclined to push the *yes* button in case of similarity and the *no* button otherwise), each of the three groups of negative stimuli contained a number of such pseudocognate pairs, about as many

as the number of cognates in the corresponding positive set. In addition to the test pairs, each group of stimuli contained 25 word pairs for practice.

Apparatus and procedure. The apparatus was the same as that used in Experiment 1, except that instead of a voice-operated switch a response box with two push buttons was attached to the computer. We informed the participants that they would be presented with pairs of words, one word in one language and the second in another language, and that they were to decide, both as quickly and as accurately as possible, whether the words within a pair were translations of one another. They were to push the right-hand button with their right forefinger in case of a translation pair, and the left-hand button with their left forefinger if the word pair did not consist of translations. The sequence of events during each trial was as follows. First, a fixation stimulus (an asterisk) was presented for 1 second, slightly above where the first word of a pair would appear. Immediately after its offset, the first word of a pair, the Dutch word, was presented. After 333 ms the second word of the pair joined the first on the screen, one line below the first word. The reason for not presenting the words of a pair simultaneously but with a brief delay between them was that this way the order in which they would be processed could be controlled and would be the same as the processing order in translation production. The chosen onset asynchrony of 333 ms is far too short for a translation to be generated by the participant before the actual appearance of the second word of a pair, but long enough for the participant to have recognized the first word by the time the second is presented. Immediately after the participant's response, one of the words *correct*, *slow*, or *wrong* appeared on the screen (in Dutch). The word *slow* appeared when the response took longer than 1600 ms. We provided this feedback, which implicitly pointed at the relevance of both speed and accuracy of responding, because it helped keep the participants alert. The feedback remained on the screen for 2 seconds. After this, both the word pair and the feedback disappeared from the screen simultaneously. The fixation stimulus reappeared 1 second later. The

duration between the onset of the second word in the word pair and the moment the participant pushed a button was registered as RT. The stimuli (80 test stimuli—40 requiring a *yes* response and 40 requiring a *no* response—and 25 for practice) were presented again in blocks (of 20 or 25 stimuli each). After each block the average RT and the number of errors for that block appeared on the screen. The order of the translation conditions rotated among the participants. Within a translation condition, we presented the stimuli in a random order, a different order for all participants. After monitoring the participant's performance during the practice session, the experimenter withdrew until the next translation session started.

Results and Discussion

We performed a number of preliminary analyses to see whether the participants had complied with the instructions to respond both quickly and accurately on all trials, positive and negative. The data suggested they had: Error percentages were rather low, both on positive (4.4% overall) and on negative trials (3.0% overall). As in de Groot and Comijs (1995), responding was faster on positive trials (547 ms and 551 ms for the analyses by participants and items, respectively) than on negative trials (668 ms and 670 ms for the analyses by participants and items, respectively). A further set of preliminary analyses focused on the difference between negative pairs consisting of pseudocognates and those consisting of words that were not perceptually similar ("negative controls"). In both translation conditions, pseudocognates were processed more slowly and less accurately than negative controls (Condition Dutch-English: 722 ms and 7.1% errors for pseudocognates vs. 634 ms and 1.6% errors for negative controls; Condition Dutch-French: 722 ms and 3.3% errors for pseudocognates vs. 680 ms and 2.4% errors for negative controls; see de Groot & Comijs, 1995, for similar findings). All further analyses deal only with the "critical" data, that is, those associated with the positive trials.

For each of the 48 participants, we subsequently calculated four mean RTs on the critical data, namely, one RT for each of the four conditions formed by the two levels of the variable concreteness (concrete vs. abstract) and the two levels of the variable translation condition (Dutch-English and Dutch-French). In calculating these means, we excluded RTs associated with errors, and those shorter than 300 ms and longer than 1500 ms. We excluded these outliers (few occurred) because they presumably reflected premature responding and attention lapses, respectively, and not the duration of translation retrieval. We set the minimum lower than in Experiment 1, where it was 400 ms, because translation recognition takes less time than translation production. With a minimum of 400 ms we might have inadvertently excluded a number of appropriate responses.

Furthermore, for all critical items we calculated a mean RT, collapsed across all 48 participants (after preliminary analyses had shown that there were no differences between the patterns of results for the participants in the lexical-decision pretest and those in the semantic-access pretest). Finally, we calculated participant and item means for all four concreteness-by-translation conditions on the errors made on critical trials. We then ran on these means the same 2×2 (concreteness×translation condition) ANOVAs as performed on the translation-production data. The results of these analyses are presented in Table 4.

The analysis by participants on the RT data produced two significant main effects and a significant interaction, but on the corresponding item analysis only the main effect of translation condition was reliable: translation condition, $F_1(1, 47)=16.78$, $p<.001$, and $F_2(1, 76)=10.12$, $p<.01$; concreteness, $F_1(1, 47)=11.85$, $p<.01$, and $F_2(1, 76)=1.21$, $p>.10$; translation condition by concreteness, $F_1(1, 47)=13.80$, $p<.001$, and $F_2(1, 76)=1.42$, $p>.10$. The interaction between the variables on the analysis by participants showed that an effect of concreteness occurred only in Condition Dutch-English ($p<.0001$).

The two analyses on the error data revealed a significant effect of translation condition, $F_1(1, 47)=56.29$, $p<.0001$, and $F_2(1,$

Table 4
Mean Response Times and Error Rates for all Translation by Concreteness Conditions in Translation Recognition (Experiment 2). Mean Reaction Times are Given by Participants and Items.

	Participants			Items			
	Reaction Time (ms)	*SD*	*N*	Reaction Time (ms)	*SD*	*N*	Error Rate (%)
Dutch–English							
Concrete	504	67	48	506	74	20	1.6
Abstract	544	76	48	544	51	20	2.3
Effect	40			38			0.7
Dutch–French							
Concrete	569	97	48	577	82	20	7.7
Abstract	571	115	48	576	79	20	6.1
Effect	2			–1			–1.6

76)=5.83, $p<.05$. The main effect of concreteness approached significance on neither of the two error analyses ($F<1$ in both cases). The interaction between the two variables was only reliable on the participant analysis, $F_1(1, 47)=4.89$, $p<.05$, and $F_2<1$. Paired comparisons of the means of the analysis by participants showed that in Condition Dutch-French slightly more errors were made to concrete words than to abstract words ($p<.05$).

The differences in terms of RT and errors between the translation-recognition conditions Dutch-English and Dutch-French once more suggest that our participants were more competent in English than in French. More importantly, the RT data showed a concreteness effect only when translation recognition involved a Dutch word on the one hand and an English word on the other. But this time the effect only occurred on the analyses by participants. Yet, the present data corroborate the conclusions drawn from the translation-production data.

General Discussion

The Dutch-English translation conditions in this study clearly support the idea that in this condition conceptual memory is implicated: The concreteness effect was large and robust when translation production served as the experimental task. As anticipated, the effect was much smaller in translation recognition. Furthermore, in the latter task the effect was only reliable on the analysis by participants. Yet, the overall data pattern for translation recognition was consistent with that obtained in translation production and thus substantiates the conclusions to be drawn from the production data. The Dutch-French conditions showed a completely different data pattern. In translation production, a concreteness effect again occurred, but it was considerably smaller than the effect in the Dutch-English condition. What is more, it went in the opposite direction. Furthermore, it only materialized in the analysis by participants. On the item analysis a null-effect of concreteness was obtained. The translation-

recognition task showed this same null-effect of concreteness. All in all, the data strongly suggest that only when our participants translated from Dutch to English did they directly access and exploit representations in conceptual memory. In terms of the hypotheses regarding bilingual word storage and processing, the data of the Dutch-English conditions support the concept-mediation hypothesis, whereas the data of the Dutch-French conditions fit the word-association hypothesis much better.

This study provides a number of indications that our participants were more fluent in English than in French: the participants' own proficiency assessments; the data of the lexical-decision (Experiments 1 and 2) and semantic-access (Experiment 2) pretests; the translation-production (Experiment 1) and translation-recognition (Experiment 2) data. The results, combined with the established proficiency difference between our participants' two nonnative languages, thus support the developmental hypothesis of bilingual memory introduced earlier. In other words, the data are consistent with the idea that the bilingual lexicosemantic system develops from a word-association structure to a concept-mediation structure as proficiency in the nonnative language increases. A further, more general, conclusion that we can draw—because we manipulated proficiency within trilingual participants rather than between groups of bilingual participants—is that different types of lexicosemantic memory structures can coexist within one and the same multilingual individual. Our trilingual participants appear to possess word-association structures for their Dutch-French language pair, but concept-mediation structures for their Dutch-English language pair.[3]

This study regarded an effect of concreteness, a meaning variable, as indicative of concept mediation, whereas the absence of such an effect was seen as support for the word-association hypothesis. But how exactly could a concreteness effect come about? De Groot (1992a) discussed two options. One of them derives from a qualification of the common assumption of both the concept-mediation and word-association hypotheses, that trans-

lation equivalents share a representation in conceptual memory. This qualification involves the premises that only the translations of concrete words are represented in a common node in conceptual memory and that, in contrast, translations of abstract words have language-specific conceptual representations. An additional assumption is that all translation pairs known by a participant, concrete and abstract, have direct connections between their form representations in lexical memory (i.e., word-association links). Of course, word translation can only proceed via a common conceptual representation if such a representation exists. Thus, for concrete words two routes to the translation response exist (word association and concept mediation), whereas for abstract words there is just one route (word association). If additional translation routes benefit performance, concrete words would be translated with greater ease than abstract words. (See de Groot, 1993; de Groot & Nas, 1991, for further discussions of word-type dependent bilingual conceptual representation.)

This interpretation of the concreteness effect also enforces a modification of the present developmental hypothesis of multilingual memory, in two ways. First, we no longer assume that the word-association connections cease to exist with increasing proficiency in the foreign language after (for concrete words) a connection between the translation words has been established through conceptual memory. Second, instead of accepting a development from word-association translation to concept-mediation translation (and the associated memory structures), which is independent of word type, we postulate a word-type dependent development. There would be a development from a word-association representation to a (modified) concept-mediation representation only for concrete words. For abstract words, the word-association hypothesis would hold for both early and late stages of development. Consistent with this view that only the translations of concrete words share a representation in conceptual memory is the finding by Jin (1990) that cross-language semantic priming (e.g., from the French word *père* to the English word *mother*) occurs for concrete words but not for abstract ones. The received

interpretation of this cross-language effect is in terms of conceptual representations shared by translation words and connected to the conceptual representations of semantically related words, which again are shared by translation words (see de Groot & Nas, 1991; Kroll & Sholl, 1992, for an overview of studies on cross-language semantic priming). Concrete translation words might share their conceptual representations, whereas abstract translation words might not, because the meanings of concrete words are more similar across languages than those of abstract words (Taylor, 1976).

A second way the concreteness effect could come about rests on the view that individual word meanings, rather than being represented in a single node, are distributed over various nodes in conceptual memory, each node representing, for instance, one meaning aspect of the word (see e.g., de Groot, 1992b; Masson, 1991; Taylor & Taylor, 1990). Depending on the degree of meaning overlap between a word and its translation, the translation pair could share larger or smaller numbers of these nodes (the more meaning overlap, the more nodes shared). If indeed the meanings of a concrete word and its translation are more similar than those of an abstract word and its translation, a concrete translation pair would share more of the nodes in conceptual memory than an abstract translation pair. The more nodes shared, the easier retrieval via conceptual memory will be (see de Groot, 1992b, for more detail). This conception of the concreteness effect does not demand the developmental hypothesis be modified, as did the above view. In a relatively late developmental stage, words of all types, both concrete and abstract, may be translated via conceptual memory. The process is simply more efficient in the case of concrete words, because of the relatively large number of nodes in conceptual memory that are shared between the words in concrete translation pairs.

On the basis of our data (a word-association pattern for the Dutch-French language pair and a concept-mediation pattern for the Dutch-English language pair in the same individuals) we argued above that different types of lexicosemantic structures

may coexist within the same multilingual head. If the first of the above two interpretations of the concreteness effect held, this per se would provide additional support for the same point: that different types of structures coexist within the same multilingual head. But now it would also explicitly apply to *bi*linguals: Concrete words would be represented in (modified) concept-mediation structures, whereas abstract words would be represented in word-association structures. Bilingual word processing tasks (e.g., the present word-translation task) have not only shown an effect of word concreteness, but also of word frequency and the cognate status of translations. To explain these effects, analogous to the above interpretation of the present concreteness effect, differences in bilingual representational structures have to exist between high-frequency and low-frequency words, and between cognates and noncognates (de Groot, 1995). These word-type effects and the present trilingual proficiency effects thus all converge on the conclusion that every individual multilingual's memory contains lexicosemantic structures of various types.

In addition to proficiency in the foreign language(s) and word type, de Groot (1995) discussed the potential role of a number of other variables in multilingual lexicosemantic representation. Three of them are: (a) the environment in which the foreign language(s) is (are) learned (e.g., Chen, 1990; Ervin & Osgood, 1954; Gekoski, 1980; Lambert, Havelka, & Crosby, 1958; and Weinreich, 1953/1974, for discussions of the role of this variable in multilingual memory representation); (b) whether or not the languages under consideration belong to the same family of languages; (c) whether or not the foreign language(s) is (are) temporarily out of use. As mentioned above, the English and French learning environments of our participants may not have been matched completely: Our participants received both English and French training at school, but were also immersed in English in many everyday situations outside school. The French experience of our participants is likely to have been much more restricted than their immersion in English.

Regarding the second of the above three variables, we noted

that Dutch and English belong to the same family of languages (Germanic) whereas French belongs to a different family (Romance). The studies by Kroll and Curley (1988) and Chen and Leung (1989) suggest that this variable may not be all that relevant. Both these studies tested the developmental hypothesis and obtained support for it; yet Kroll and Curley compared proficiency groups bilingual in two alphabetic languages, whereas Chen and Leung tested groups bilingual in one alphabetic and one logographic language. Because both studies obtained support for the developmental hypothesis, it appears that the script relation between a bilingual's two languages does not play a critical role in multilingual lexicosemantic organization. If indeed the script relation is immaterial, it is likely that the family relation is also not critical.

The foreign languages we studied also varied on the third of the above variables. Until the day of participating in this study, our participants were confronted with English daily (e.g., when reading their university textbooks), but typically they had not used French for months. De Groot (1995) asks whether bilingual lexicosemantic organization not only develops (from a word-association to a concept-mediation structure) with increasing proficiency in L2, but also reverts in the opposite direction when existing structures are not being used for some time. De Groot adds that one should not construe these processes of learning and forgetting as the actual replacement of one type of memory structure by another, but as the strengthening and weakening of connections between memory units. Links that are not used for some time become dormant, to be revived again under the appropriate circumstances. Applied to the present study, this line of thought suggests that direct connections between French word-form representations and conceptual memory actually existed in our participants' memories but were out of use at the time of testing. As a consequence, the participants had to fall back on the use of the direct links between the French and Dutch word-form representations (producing a data pattern that supported the word-association hypothesis). In fact, this conception provides a

means to explain why, after no less than six years of schooling in French, our participants demonstrated no hint of a concept-mediation pattern for the language combination Dutch-French. Conversely, studies that have demonstrated a concept-mediation data pattern for participants after a relatively short period of L2 training, may have tested participants who still used their L2 at the moment of testing (or had recently used it).

To conclude, the present study clearly supports the idea that different lexicosemantic structures may coexist in one and the same multilingual memory system and that the representational system is not stable but develops over time and use. We focused on the role of foreign-language proficiency in lexicosemantic organization. That this variable may have been confounded with a number of other variables (a state of affairs that could not, and typically cannot, be prevented) does not detract from this general conclusion. The present data thus clearly oppose the theory of one and the same, stable, multilingual memory structure for all bilinguals/multilinguals.

Revised version accepted 20 July 1995

Notes

[1]In fact, each participant ultimately took part in three translation conditions. In addition to the Dutch-English and Dutch-French conditions, there was also an English-French condition. We included this condition only for exploratory purposes, with the only clear (and obvious) prediction (substantiated by the data) that, involving two nonnative languages, it would be the hardest condition. We chose not to report the associated data here, because not doing so allows a much more to-the-point presentation of the data, related directly to the theoretical starting-point of this study, the developmental hypothesis. The data of the English-French condition can be summarized as follows: It basically showed a null-effect of concreteness and RTs considerably slower than those in the two remaining translation conditions. Also, relatively many errors and omissions occurred in that condition.

[2]As can be seen in Table 1 (and more so in, e.g., Table 2), the mean RTs for the analyses by participants and items generally differ from one another, particularly in conditions with relatively many errors. The trials on which an error was made were not included in the RT analyses and these empty cells weigh differently in calculating the mean RTs (per participant, per

condition, collapsed across the items in that condition; and per item, collapsed across all participants in a condition) that constitute the input for the two types of analyses. Particularly in conditions with large numbers of missing data, as in translation production (see Table 2), the mean RTs of the analyses by participants and items can differ substantially.

[3]The absence of a concreteness effect when translation was from Dutch to French may seem inconsistent with the finding that concrete words are easier to learn in a new language than abstract words, also under learning conditions in which the use of keyword imagery mediation is not encouraged (see Ellis & Beaton, 1993, for a review). However, the present translation study (and similar studies) on the one hand and L2 vocabulary learning studies on the other hand deal with different questions, and may each require their own explanation of a concreteness effect (or its absence). The present translation data have been explained in terms of already existing, developing bilingual memory representations. The learning studies should explain why it is easier to form a bilingual representation of a concrete word than of an abstract word. A reason could be that concrete words are more meaningful than abstract words (de Groot, 1989; Kieras, 1978), and that the more meaningful the material to be learned, the easier it is to learn. Another is in terms of elaboration of to-be-learned words, concrete and abstract words differing from one another in the extent to which they enable elaboration. If concrete words allow more extensive elaboration than abstract words, then, following learning, for concrete words there may be more retrieval cues than for abstract words, and consequently, memory for concrete words will be better.

References

Abunuwara, E. (1992). The structure of the trilingual lexicon. *European Journal of Cognitive Psychology, 4*, 311–322.

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

Chen, H.-C. (1990). Lexical processing in a nonnative language: Effects of language proficiency and learning strategy. *Memory & Cognition, 18*, 279–288.

Chen, H.-C., & Ho, C. (1986). Development of Stroop interference in Chinese-English bilinguals. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 12*, 397–401.

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

de Groot, A. M. B. (1989). Representational aspects of word imageability and word frequency as assessed through word association. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 15*, 824–845.

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

de Groot, A. M. B. (1992b). Bilingual lexical representation: A closer look at conceptual representations. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 389–412). Amsterdam: North-Holland.

de Groot, A. M. B. (1993). Word-type effects in bilingual processing tasks: Support for a mixed representational system. In R. Schreuder & B. Weltens (Eds.), *The bilingual lexicon* (pp. 27–51). Amsterdam: Benjamins.

de Groot, A. M. B. (1995). Determinants of bilingual lexicosemantic organisation. *Computer Assisted Language Learning, 8*, 151–180.

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

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

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

Ervin, S. M., & Osgood, C. E. (1954). Second language learning and bilingualism. *Journal of Abnormal Social Psychology, 49*(Suppl.), 139–146.

Ellis, N. C., & Beaton, A. (1993). Psycholinguistic determinants of foreign language vocabulary learning. *Language Learning, 43*, 559–617.

Gekoski, W. L. (1980). Language acquisition context and language organization in bilinguals. *Journal of Psycholinguistic Research, 9*, 429–449.

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

Kieras, D. (1978). Beyond pictures and words: Alternative information-processing models for imagery effects in verbal memory. *Psychological Bulletin, 85*, 532–554.

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: Vol. 2. Clinical and educational implications* (pp. 389–395). London: John Wiley & Sons.

Kroll, J. F., & Sholl, A. (1992). Lexical and conceptual memory in fluent and nonfluent bilinguals. In R. J. Harris (Ed.), *Cognitive processing in bilinguals* (pp. 191–204). Amsterdam: North-Holland.

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

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

La Heij, W., Hooglander, A., Kerling, R., & Van der Velden, E. (in press). Nonverbal context effects in forward and backward translation: Evidence for concept mediation. *Journal of Memory & Language*.

Lambert, W. E., Havelka, J., & Crosby, C. (1958). The influence of language-acquisition contexts on bilingualism. *Journal of Abnormal & Social Psychology, 56*, 239–244.

Mägiste, E. (1984). Stroop tasks and dichotic translation: The development of interference patterns in bilinguals. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 10*, 304–315.

Masson, M. E. J. (1991). A distributed memory model of context effects in word identification. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 233–263). Hillsdale, NJ: Lawrence Erlbaum.

Paivio, A. (1966). Latency of verbal associations and imagery to noun stimuli as a function of abstractness and generality. *Canadian Journal of Psychology, 20*, 378–387.

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 & 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.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning & Memory, 6*, 174–215.

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.

Taylor, I., & Taylor, M. M. (1990). *Psycholinguistics: Learning and using language*. Englewood Cliffs, NJ: Prentice Hall.

Tzelgov, J., Henik, A., & Leiser, D. (1990). Controlling Stroop interference: Evidence from a bilingual task. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 16*, 760–771.

Van Loon-Vervoorn, W. A. (1985). *Voorstelbaarheidswaarden van Nederlandse woorden* [Imageability ratings of Dutch words]. Lisse, The Netherlands: Swets & Zeitlinger.

Weinreich, U. (1974). *Languages in contact: Findings and problems*. The Hague, The Netherlands: Mouton. (Original work published 1953)

Table A–1
Test Words, and Their Ratings on Cognate Status, Imageability, and Log Word Frequency

	Cognate Status	Imageability Rating	Log Word Frequency
Condition Dutch-English			
Concrete Words			
moon	1	6.87	1.81
mother	1	6.58	2.77
head	1	6.37	2.74
farmer	0	6.50	2.00
tree	0	6.90	2.14
king	1	6.27	2.00
daughter	1	5.93	2.08
night	1	5.87	2.42
wall	0	6.72	2.17
skin	0	6.10	1.94
kitchen	1	6.63	1.98
leaf	0	6.33	2.06
sea	1	6.87	2.15
horse	0	6.93	2.20
knee	1	6.57	1.93
money	0	6.33	2.45
girl	0	6.63	2.55
gentleman	0	5.98	2.33
church	0	6.63	2.31
dog	0	6.73	2.22
M	0.45	6.49	2.21
SD	0.51	0.33	0.27
Abstract Words			
change	0	3.50	2.17
certainty	0	2.73	1.86
soul	1	2.63	1.88
quality	2	3.13	1.88
fact	1	2.50	2.55
example	0	2.10	2.36
cause	0	2.10	2.04
science	0	3.50	2.10
distance	0	3.93	2.10

Table A–1 (continued)
Test Words, and Their Ratings on Cognate Status, Imageability, and Log Word Frequency

	Cognate Status	Imageability Rating	Log Word Frequency
Condition Dutch-English			
Abstract Words (continued)			
insight	1	3.10	2.00
whole	1	2.63	1.96
meaning	0	2.80	2.23
space	0	3.17	2.18
impression	0	2.20	2.19
direction	0	3.80	2.30
attention	0	3.27	2.30
state	2	3.67	2.46
resistance	0	2.43	1.78
proposal	0	3.57	1.90
success	2	2.40	1.97
M	0.50	2.96	2.11
SD	0.76	0.58	0.21
Condition Dutch-French			
Concrete Words			
breast	0	6.42	2.05
smoke	0	6.27	1.82
house	0	6.87	2.80
father	0	6.35	2.76
face	0	6.43	2.70
road	0	6.50	2.63
book	0	6.73	2.59
room	1	6.50	2.56
table	1	6.70	2.39
colour	1	6.00	2.19
uncle	0	5.90	2.04
tooth	1	6.90	1.95
train	2	6.87	1.91
wood	1	6.60	1.93
army	0	6.00	1.85
paper	0	6.90	2.07

Table A–1 (concluded)
Test Words, and Their Ratings on Cognate Status, Imageability, and Log Word Frequency

	Cognate Status	Imageability Rating	Log Word Frequency
Condition Dutch-French			
Concrete Words (continued)			
office	0	5.80	1.82
aunt	2	6.20	2.02
police	2	6.53	1.98
river	0	6.67	1.84
M	0.55	6.46	2.20
SD	0.76	0.34	0.35
Abstract Words			
chance	1	3.66	2.30
influence	1	2.23	2.28
task	1	3.43	2.17
information	2	3.40	2.04
power	0	2.73	2.28
result	2	3.57	2.12
case	0	2.17	2.73
criticism	2	2.63	1.86
action	2	3.83	1.80
thing	0	3.47	2.57
turn	0	2.87	1.83
advantage	0	2.70	1.84
demand	0	3.47	1.91
past	0	3.13	1.99
age	0	3.33	2.01
decision	0	3.70	2.02
language	0	3.97	2.19
reason	1	2.63	2.35
difference	0	2.53	2.24
attempt	0	2.57	2.04
M	0.60	3.10	2.13
SD	0.82	0.55	0.25