Language Learning 45:3, September 1995, pp. 467–509

Translation Recognition and Translation Production: Comparing a New and an Old Tool in the Study of Bilingualism

Annette M. B. de Groot and Hannie Comijs
University of Amsterdam

This study explored the "translation-recognition" task. Each trial in this task presents a word pair, one word in L1, the second in L2. The participant has to decide whether or not the words within a pair are translations of one another. Performance is compared with that in translation production, where on each trial the participant has to come up with the translation of the presented word. The results of two experiments, one for 40 adult Dutch learners of English and the second for 80 people from the same population, suggest that translation recognition and translation production generally respond to the same manipulations. An exception to this pattern emerges when cognates and noncognates are focused on separately. With noncognate materials translation production from L1 to L2 shows a larger role of semantic variables than both translation production from L2 to L1 and translation recognition. Whether within the recognition task cognates and noncognates are presented mixed or blocked, and whether the nontranslation pairs consist of perceptually similar or

We thank Janet van Hell, Judith Kroll and two anonymous reviewers of this journal for very helpful comments, and Paul Oosterveld for statistical advice.

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

dissimilar words, or both, do not affect the actual translation-retrieval process.

A frequently used task in the study of bilingualism is word translation. Typically, the participant is visually presented with a word in one of his or her languages and has to produce its translation in the other language out loud. This particular translation task will be referred to as "translation production" in the remainder of this text. It has been used most extensively in the study of bilingual lexicosemantic representation, that is, the organization of form and meaning of the native (L1) and second-language (L2) words in bilingual memory (e.g., Chen, 1990; Chen & Leung, 1989; de Groot, 1992a; de Groot, Dannenburg, & van Hell, 1994; Kroll & Curley, 1988; Kroll & Stewart, 1994; Potter, So, von Eckardt, & Feldman, 1984).

Most models of bilingual lexicosemantic organization (Fig. 1) assume the existence of two layers of units (nodes) in the representational system, one layer representing word forms (lexical memory) and the second representing meaning (conceptual memory). Typical theory believes lexical memory contains two sets of forms, one for each of the bilingual's two languages. In contrast, conceptual memory is often thought to contain language-independent representations; that is, units that each individually represent the meaning of both words in a translation pair.

The various models proposed differ from one another in terms of the connections between the units and what the units represent. For instance, a "word-association" or "subordinative" model (Fig. 1a; e.g., Chen, 1990; Chen & Leung, 1989; Kroll & Curley, 1988; Potter et al., 1984) assumes connections between the unit representing a word's form in L1 and the unit representing its form in L2; no connections exist between L2 word-form representations and the shared, language-independent, meaning unit. In contrast, the "concept-mediation" model (Fig. 1b; e.g., Chen, 1990; Chen & Leung, 1989; Kroll & Curley, 1988; Potter et al., 1984) assumes the latter type of links but not the former. A

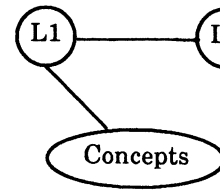

Figure 1a. Word association.

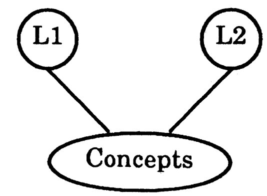

Figure 1b. Concept mediation.

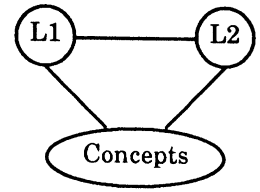

Figure 1c. Mixed.

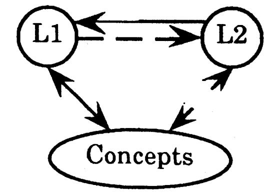

Figure 1d. Asymmetrical.

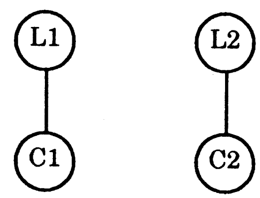

Figure 1e. Coordinate.

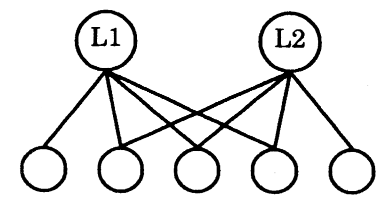

Figure 1f. Distributed.

Figure 1. Models of bilingual memory.

third model (Fig. 1c; e.g., de Groot, 1992a; de Groot & Nas, 1991) accepts both types of links in the lexicosemantic representations. One instance of this type of model (Fig. 1d; Kroll & Stewart, 1994) assumes directional asymmetries in the strength of the connections between the various types of memory units. It proposes *two* links between the word-form representations in lexical memory: a strong link from L2 to L1 and a weaker link in the opposite direction. Furthermore, this type of model assumes stronger links between L1 form representations and the shared conceptual representations than between L2 form representations and the conceptual representations. In yet a further model, the "coordinate" model (Fig. 1e; cf. Weinreich, 1953/1974), the lexicosemantic knowledge of the bilingual is represented completely language-specifically. Both at the lexical and at the conceptual level the L1 and L2 systems are separate. Finally, in the "distributed" model (Fig. 1f; de Groot, 1992b; Taylor & Taylor, 1990) individual units (forms, meanings) and their representations in memory do not map in a one-to-one fashion, but one to many. In this example only the meanings of the L1 and L2 words are represented in more than one memory node, but word form could also be represented in such a distributed fashion.

Distributed *conceptual* representation in bilingual memory can do justice to the fact that the meanings of a word in L1 and its (closest) translation in L2 seldom overlap completely (in other words, that true translation equivalents are rare). They can do so by incomplete overlap of the collection of nodes in conceptual memory that represents the meaning of the translation "equivalents". In the (hypothetical) example (Fig. 1f) the meaning of the L1 word and that of its translation are each represented in four conceptual nodes, but only three are shared between the translation words. Similarly, distributed *lexical* representation can do justice to the fact that in many pairs of languages the forms of the words in a translation pair are often similar and sometimes even identical. Similar L1 and L2 forms would share a set of representational elements at the lexical level, whereas totally dissimilar forms would not. The size of the shared set would be a function of the degree of form similarity.

Word translation in a word-association model could come about via the word-association links within lexical memory. That is, a stimulus word in L1 accesses the L1 word-form representation. From there the link between the L1 and L2 form representations is traced, the L2 form representation accessed, and the corresponding word produced. When translation is from L2 to L1, the same route is tracked, but in the opposite direction, starting with access of the L2 word-form representation. Word translation in a concept-mediation structure could occur via the shared meaning representation in conceptual memory. Similarly, word translation in the distributed model (Fig. 1f) could occur via the representations of the meaning elements that are shared by L1 and L2. In the "mixed" model (Fig. 1c) the task could be performed both ways, via the word-association links and via conceptual memory. In the asymmetrical model, the translation process may use the route along the stronger links. Kroll and Stewart (1994) assumed this to be the route via conceptual memory when L1 words have to be translated in L2, but the direct route between the word-form representations when translation is from L2 to L1. Given that in a coordinate structure neither the direct links between the word-form representations nor the indirect connections via conceptual memory exist, it is as yet unclear how word translation in such a system would come about.

Lexicosemantic organization is not uniform across bilinguals, nor within individual bilinguals. Instead, it depends on a number of factors (see de Groot, 1995, for a review). Different types of models may hold for different groups of bilinguals (e.g., Chen, 1990; Chen & Leung, 1989; Kroll & Curley, 1988) and for different types of words (e.g., de Groot & Nas, 1991; Jin, 1990). Other variables that may determine the representation structure are learning strategy (e.g., Chen, 1990; Chen & Leung, 1989) and the learning environment (e.g., Lambert, Havelka, & Crosby, 1958). One way to discover what type of model holds in a particular case is to look at the effect of manipulating semantic variables on translation performance (de Groot, 1992a; de Groot et al., 1994; Kroll & Stewart, 1994). If such a manipulation exerts an effect, conceptual memory is implicated; if not, conceptual memory is not implicated. The former result thus provides support for a concept-mediation (or a mixed or a distributed) structure; the latter for a word-association or a coordinate structure. Differences in the magnitude of the semantic effects between the two translation directions (larger effects from L1 to L2 than in the reverse direction, or *only* effects from L1 to L2) would support the asymmetrical model. Note that the critical assumption being made here is that effects of semantic variables originate in conceptual memory, which is the place where word meanings are stored.

In all but one of the translation studies enumerated above, the experimental task was translation production. When unbalanced bilinguals are used as participants (as is common) and not only relatively easy words are presented for translation, this task gives rise to many omissions (the stimulus does not evoke a response) and errors. Omissions and errors do not necessarily imply that the elicited knowledge is absent from the memory. It may at least be stored partly, but rooted too weakly to be retrieved by this particular task.

One study (de Groot, 1992a, Experiment 2) introduced a version of the word-translation task expected to be easier than translation production, and thus to result in more, and more often correct, responses: "translation recognition". Each trial in a translation-recognition experiment presents participants with a pair of words, one in L1 and one in L2. The two words are or are not translations of one another, and the participant must decide whether they are. Unless the nontranslation pairs are deliberately constructed to complicate the decision process (e.g., by pairing an L1 word to a semantically related L2 word; Talamas & Kroll, 1994), this task should be easier than translation production. Probably under these circumstances weakly rooted knowledge can be exploited and will often lead to a correct response. Therefore, this task can be used with bilinguals with relatively low levels of L2 fluency, and relatively difficult words (e.g., words with a low frequency of occurrence) can be presented as stimuli, with little risk of obtaining inconveniently few correct responses.

In addition, collecting translation-recognition data involves much less labor and time than collecting data in a translation-production task. In the latter, the retrieved translation is typically output verbally by the participants, and the onset of the response is registered by a voice switch. The task requires either the experimenter's monitoring the workings of the voice switch and typing in the participant's responses or the tape recording of the responses followed by transcribing the tapes. Neither is required for translation-recognition. Furthermore, the analysis of the collected data is much less laborious. For instance, translation-

production data require a cleaning-up procedure in which response times (RTs) associated with the triggering of the voice key to environmental sounds are removed. Furthermore, each response that is not the intended (by the investigator) one needs to be checked in a translation dictionary to see whether under no circumstances it would be a permitted translation of the stimulus word. Separate analyses must be performed on the data for correct and intended responses and for correct but unintended responses. None of these analysis procedures has to be performed on translation-recognition data.

In addition to these practical reasons to explore the translation-recognition task, we think there is a theoretical reason. In translation production, access of conceptual memory (assumed by most models) is followed by access to the lexicon of the output language, selection of the appropriate entry, accessing of its phonology, and response execution. Translation recognition does not involve the latter processing components, but a decision stage instead (discussed at length below). If a particular effect occurs in both tasks, it can therefore be inferred to reside in a processing stage shared by the two tasks, rather than in one of the stages unique to either task. A task comparison thus helps pinpoint the locus of the effects obtained. For instance, an effect occurring both in translation production and in translation recognition cannot be attributed to retrieving the phonology of the response word, because translation recognition does not involve this processing component. Similarly, it cannot be attributed to the process of converting the outcome of the translation-recognition process in an "accept" or "reject" response, because translation production does not involve this process.

However, even apart from the final processing steps, translation production and translation recognition do not necessarily involve exactly the same processing. Indeed, in a way processing in recognition may be "shallower" than in production. Given that only in translation recognition are both terms of a translation pair presented simultaneously, the participants may exploit the direct links between the L1 and L2 word-form representations that exist in some of the models (see Fig. 1) more intensively. If so, conceptual memory would be implicated less in recognition than in production, and, consequently, the effects of semantic variables should be smaller in the former task. Furthermore, to the extent that conceptual memory *is* implicated in translation recognition, the same conceptual information is not necessarily retrieved from that knowledge store in both tasks. Assuming, for instance, different levels of conceptual information, it is plausible that translation recognition addresses more superficial information in conceptual memory than does translation production.

The present study searches for determinants of translation recognition and compares performance in this task with that in translation production. A comparison between these two tasks has been performed before (de Groot, 1992a, Experiments 1 & 2), but on a very small scale: the stimulus set included only 96 words (in production) or 96 translation pairs (in recognition), all with a noncognate relation between the translation words (i.e., a stimulus word and its translation were always perceptually dissimilar). The combination of languages used was Dutch-English. The present study used a much larger selection of Dutch-English stimulus materials. The stimulus materials varied on 13 variables that may determine the efficacy of translation recognition or production. These are the same variables used by de Groot et al (1994). They are: Imageability, Context Availability, Definition Accuracy, Familiarity, Log Word Frequency, and Length of both the Dutch words and of their translations in English (amounting to 12 variables in all), and the cognate status of the translation equivalents (all variables will be explained below). The words' values on Imageability, Context Availability, Definition Accuracy, and Familiarity had been collected before in a set of eight norming studies (de Groot et al., 1994). In each of these studies a group of participants (a different group for each of the four stimulus characteristics and for each of the two languages) rated the words on a 7-point scale with respect to that particular variable. The cognate ratings were derived from an earlier norming study in which yet another group of participants rated

pairs of words, each pair consisting of a Dutch word and its (dominant) English translation. Again the rating was done on a 7-point scale. Further details of the procedures followed in the norming studies can be found in de Groot (1992a) and de Groot et al. (1994). The length measure was derived by simply counting the number of letters in the Dutch words and in their English translations. Finally, the words' written word frequencies were derived from the frequency count of the Centre of Lexical Information (CELEX) in Nijmegen, The Netherlands (Burnage, 1990). This corpus contains a count of 18.8 million printed English words and of 42.5 million printed Dutch words.

Word imageability (e.g., Paivio, 1968) refers to the extent the referent of the word arouses a mental image. This variable is typically confounded with *word concreteness*, the extent to which the word's referent can be experienced by the senses. A word's *context availability* (e.g., Schwanenflugel, Harnishfeger, & Stowe, 1988) is a rating of how easy it would be to come up with a particular context or circumstance in which the word might appear. This variable is highly correlated with word imageability and concreteness (Schwanenflugel et al., 1988), and with *definition accuracy* (de Groot, 1992a; de Groot et al., 1994). This latter variable is a measure of how easy participants think it would be to define the word. *Word familiarity* (e.g., Noble, 1953) indicates the frequency with which the participants have experienced the word. It is considered as a measure of subjective frequency (as opposed to word-frequency measures based on a word count in text corpora, that produces an "objective" measure of word frequency). Finally, *cognate status* is a measure of the orthographic and phonological (combined: perceptual) similarity of the words in a translation pair.

De Groot et al. (1994) showed that the 13 predictor variables cluster on four factors: (a) a Semantic factor loaded on highly by six predictors, namely, word imageability, context availability, and definition accuracy, each of them as assessed for both the Dutch and the English words; (b) a Familiarity factor loaded on most highly by familiarity and written word frequency, again, each of them as determined both for the Dutch and the English words, and moderately loaded on by the context availability and definition accuracy of the English words; (c) a Length factor, on which the length of the Dutch and English words loaded; (d) a Cognate factor, loaded on by the cognate-status variable.

The present study reports two translation-recognition experiments, the data of which will be compared with the corresponding data of the translation-production study of de Groot et al. (1994, Experiment 1). All of these data combined should delineate the circumstances under which translation recognition could be used as a convenient alternative to translation production in the study of bilingualism. The discussion of the data will focus on the effects of the semantic variables, that is, word imageability, context availability and definition accuracy.

Experiment 1

Method

Participants

Forty first-year psychology students from the University of Amsterdam received course credit for participation. Twenty of them took part in Condition Dutch-English, and the remaining 20 in Condition English-Dutch (see below). All participants were unbalanced bilinguals, with Dutch as their native language (L1) and a relatively high level of fluency in their second language, English (L2). They had started to learn English in a school setting about the age of 12, and had received about 3 hours of English training per week until they went to university (mostly at age 18 to 19). Their university textbooks were mainly in English. On entering the laboratory they were asked to rate their comprehension and production abilities in English on a 7-point scale (1=no knowledge of English; 7=knowledge equal to the corresponding skill in Dutch). The mean comprehension and production ratings of the participants in Condition Dutch-English were 5.40 (*SD*=.75) and 5.00 (*SD*=.65), respectively. The corresponding ratings for the participants in Condition English-Dutch were 5.35 (*SD*=.67) and 4.50 (*SD*=.83). Other indications of the participants' level of proficiency in English are the overall response times and error scores in simple word-recognition tasks of other participant samples drawn from the same population. For instance, lexical-decision response time to English words typically ranges between 520 and 620 ms, depending on word characteristics such as frequency of occurrence, as compared to a range of about 480 to 540 ms for Dutch words. The associated error percentages vary between 2 and 5 for English words, and between 0 and 2 for Dutch words.

Materials

The "positive" test materials (requiring a "yes" response) consisted of 440 word pairs, each pair consisting of a Dutch noun and its dominant or only translation in English. The Dutch words of these pairs had served as stimuli in the "forward" translation condition in studies by de Groot (1992a; Experiment 3) and de Groot et al. (1994). In forward translation participants translate L1 words into L2. The English words had served as stimuli in the backward translation condition (from L2 to L1) of de Groot et al. (1994). Both of those studies had employed translation production rather than translation recognition as task. As pointed out before, for all Dutch and English words, ratings on a 7-point scale were available on their imageability, context availability, definition accuracy, and familiarity. The 7-end of the scale indicated easy to imagine, to think a context for, to define, and highly familiar, respectively. Furthermore, frequency and length information was available for all words. Finally, for each translation pair a rating concerning its cognate status was available. About half of the translation pairs consisted of cognates (perceptually similar translations), the remaining half of noncognates (perceptually dissimilar translations). Procedural details about the rating studies can be found in de Groot (1992a) and de Groot et al. (1994).

The Dutch words in the 440 translation pairs were first ordered alphabetically. Subsequently, the translation pairs were split up into two lists on the basis of this alphabetical ordering: The pair with the alphabetically first (Dutch) word was assigned to List 1; that with the alphabetically second word was assigned to List 2; the pair with the word third in order was again assigned to List 1, and so on. The two lists resulting from this procedure thus each consisted of 220 translation pairs. Subsequently, a total of 350 "negative" word pairs (requiring a "no" response), each consisting of a Dutch and an English noun that were not translations of one another, were added, 175 to each list.[1] The words in these pairs were all different from the words in the positive pairs of the same list. Fifteen of the negative word pairs within each list consisted of pseudocognates, that is, a Dutch and an English word that resembled one another perceptually, but that were not each other's translation, nor were they related to one another in any other way. Examples are: *hout* (*wood*)-*house*, *pop* (*doll*)-*pope*, and *mening* (*opinion*)-*meaning* (English translations of the Dutch words in parentheses). All remaining negative pairs consisted of perceptually dissimilar words. But as with the pseudocognates, the words in these pairs were not translations of one another, nor were they related to one another in any other way. The pseudocognates were included to explore the possibility of a bias in the participants to accept a word pair as a translation pair when it consists of words that are perceptually similar, and, conversely, to reject a word pair as a translation pair when it consists of perceptually dissimilar words.

Every participant performed translation recognition to the stimuli in both lists, in two separate sessions, with approximately one week between the two sessions. Half of the participants received the Dutch word of a stimulus pair before the English word; the language order was reversed for the remainder. Prior to each list of experimental stimuli, the participant received 15 word pairs for practice.

Apparatus and Procedure

The experiment was run on an Apple Macintosh Plus computer. A separate response box with two push buttons was attached to the computer. A Pascal program controlled the stimulus presentation and the recording of the response times.

The participants received written instructions (followed by additional oral instructions, if necessary) in which they were told they would be presented with word pairs, each of them consisting of a Dutch and an English word, and that they should decide, both as quickly and as accurately as possible, whether the words within a pair were translations of one another. In case of a translation pair they were to push the right-hand button on the response box with their right forefinger; in case of a nontranslation pair they should push the left-hand button with their left forefinger. A participant took part in either Condition Dutch-English, or in Condition English-Dutch. In Condition Dutch-English the Dutch word always preceded the English word. In Condition English-Dutch the presentation order was reversed. The onset asynchrony between the first and second word of a pair was always 238 ms (14 clock ticks of 17 ms each).

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 was to appear. Immediately after its offset, the first word of a pair was presented. After 238 ms the second word of the pair joined the first on the screen, one line below the first word. Immediately after the participant's response, one of the words *correct*, *slow*, or *wrong* appeared on the screen (in Dutch). The word *slow* was shown when the response took longer than 1,200 ms. Feedback remained on the screen for 2 seconds. Then, 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 of the pair and the moment the response button was pushed was registered as RT. Occasionally, the participant failed to respond at all. The maximum presentation duration in these cases was 5 seconds. The Dutch and English words of a pair were presented successively rather than simultaneously because this way the processing order of the two words within a pair could be controlled. The onset asynchrony between the two was far too short for a translation to be generated by the participant before the second word appeared, but long enough for the first word to have been recognized by then.

The 395 word pairs per session (=per list; see Materials section above) were presented in groups of 40 (the last group contained 35 word pairs only). After each group the average RT and the number of errors for that group were shown on the screen. The word pairs were presented in a random order, a different order for each participant. The order of the four language order (Dutch-English and English-Dutch) by list (List 1 and List 2) conditions rotated systematically among the participants. After monitoring the participant's performance during the practice session, the experimenter left the room. The experiment lasted about 50 minutes per session.

Results and Discussion

For each participant in both language-order conditions (Dutch-English and English-Dutch) we calculated mean RTs for the four conditions formed by the two levels of the variable List (List 1 and List 2) and the two levels of the variable Word-Pair Type (translation and nontranslation). When calculating these means, we excluded RTs associated with error responses. We also excluded RTs shorter than 100 ms and longer than 1,400 ms (less than 1% in all). Furthermore, we calculated for each participant the number of errors for each of these four conditions. We entered these mean RTs and error scores into two 2×2×2 (List×Word-Pair Type×Language Order) analyses of variance, one analysis on the RT data, and one on the error data. In these analyses we treated List and Word-Pair Type as within-subject variables and Language Order as a between-subjects variable. The data of these analyses are summarized in Table 1.

Table 1

Mean Response Times (in ms) and Error Rates (in Percentages) for All Language Order by List (1 vs. 2) by Word-Pair Type Conditions in Experiment 1

	Dutch-English				English-Dutch			
	List 1		List 2		List 1		List 2	
Word Pair Type	Reaction Time	Error Rate	Reaction Time	Error Rate	Reaction Time	Error Rate	Reaction Time	Error Rate
Translation	567	5.3	563	5.5	561	7.4	554	7.7
Nontranslation	628	5.9	629	5.1	630	6.7	624	4.7

The analysis on the RT data showed a main effect of Word-Pair Type, $F(1, 38)=177.43, p<.0001$, MSE=1003, with translation pairs being responded to faster (561 ms) than nontranslation pairs (628 ms). None of the remaining main effects nor any of the interactions was significant (all Fs<1). On the analysis of the error data the main effect of Word-Pair Type was again significant, $F(1, 38)=21.73$, $p<.0001$, MSE=35.23, with slightly more errors being made on translation pairs (6.5%) than on nontranslation pairs (5.6%). This main effect was qualified by significant interactions between Language Order and Word-Pair Type, $F(1, 38)=5.62, p<.05$, MSE=28.50, and between List and Word-Pair Type, $F(1, 38)=4.90, p<.05$, MSE=18.99. As shown in Table 1, the difference in error rates between word-pair types was larger when English words preceded Dutch words than with the reverse language order, and (collapsed across the language-order conditions) it was larger in List 2 than in List 1. None of the remaining main effects and interactions was significant ($p>.10$ in all cases).

Correlational Analyses

Subsequently, we calculated a mean RT for each of the language-order conditions separately, over correct responses for each of the 440 translation word pairs, collapsing across participants. Reaction times shorter than 100 ms and longer than 1,400 ms (less than 1% in all) were excluded from the calculation of these means. These mean RTs were then correlated with the 13 predictor variables included in the production study of de Groot et al. (1994). This was done for both language-order conditions. The resulting correlation matrices are shown in Table 2. Each of them is presented next to the correlation matrix for the corresponding condition in the production study of de Groot et al. (1994, Experiment 1): The matrix for language-order condition Dutch-English is associated with that for forward translation production (from Dutch to English; L1 to L2); the matrix for Condition English-Dutch is associated with that for backward translation production (English to Dutch; L2 to L1).[2]

Table 2

Correlations Between the 13 Predictor Variables and Reaction Times in Translation Production (de Groot et al., 1994, Experiment 1) and Translation Recognition (Experiment 1)

Predictor	Dutch-English: Translation Production	Dutch-English: Translation Recognition	English-Dutch: Translation Production	English-Dutch: Translation Recognition
Imageability-Dutch	–.35	–.16**	–.28	–.22
Imageability-English	–.45	–.29**	–.40	–.35
Context Availability-Dutch	–.43	–.27**	–.37	–.32
Context Availability-English	–.58	–.47*	–.57	–.54
Definition Accuracy-Dutch	–.31	–.14**	–.26	–.18
Definition Accuracy-English	–.55	–.51	–.59	–.57
Cognate Status	–.38	–.41	–.36	–.39
Familiarity-Dutch	–.35	–.38	–.49	–.43
Familiarity-English	–.60	–.64	–.66	–.66
Log Word Frequency-Dutch	–.41	–.43	–.52	–.45
Log Word Frequency-English	–.40	–.42	–.45	–.40
Length-Dutch	.24	.02**	.13	.08
Length-English	.32	.17*	.22	.17

**$p<.01$ and *$p<.05$ involve the comparisons between Translation Production Dutch-English and Translation Recognition Dutch-English

$p<.05$ if $r>.08$. $p<.01$ if $r>.11$.

As can be seen, the direction of the correlation coefficients in translation recognition was generally the same as that of the corresponding coefficients in translation production. Imageability, Context Availability, Definition Accuracy, Cognate Status, Familiarity, and Log Word Frequency, of both the Dutch and the English words, all correlated negatively with RT, whereas the length of the English words correlated positively with RT. Unlike in production, the length of the Dutch words in the translation pairs did not correlate with RT. In other words, translation-recognition times are relatively short when the translation pairs refer to words that are easy to imagine, to think up a context for, and to define; the words in these pairs are familiar and frequent; and the translations are perceptually similar. Furthermore, translation-recognition RT is shorter when the English term in the translation pair is short than when it is longer.

The correlations between RT on the one hand and the semantic predictor variables (Imageability-Dutch; Imageability-English; Context Availability-Dutch; Context Availability-English; Definition Accuracy-Dutch; Definition Accuracy-English) on the other suggest that conceptual memory is also implicated in translation recognition, as it appeared to be in translation production. However, the sizes of the relevant coefficients in Condition Dutch-English generally appear to be smaller in recognition than in production. This trend suggests that the role of conceptual memory is smaller in the former task. This difference between tasks does not appear to exist in Condition English-Dutch.

Multiple Regression Analyses

The variance in forward-translation RT accounted for by the 13 predictor variables in the production study of de Groot et al. (1994) had been 56% (Experiment 1) and 57% (Experiment 2); the corresponding percentages for backward-translation RT were 63% and 62%. We performed two multiple regression analyses to calculate the corresponding percentages of accounted variance in translation recognition in the present study. In one of them the RT in Condition Dutch-English served as the dependent variable; in the second the RT in Condition English-Dutch was the dependent variable. We included the same 13 predictor variables. The percentages of accounted variance were about as large as in translation production: 58% for Condition Dutch-English, and 61% for Condition English-Dutch. Thus, as was the case with translation production, these 13 predictor variables account for a large percentage of variance. No further data of the regression analyses will be provided here, because the high intercorrelations of a number of the predictor variables (e.g., Imageability-Dutch and Imageability-English correlate .94; see de Groot et al., 1994, Table 3, for the complete correlation matrix) make it extremely difficult to interpret the solved regression equation. This phenomenon is known as the "multicollinearity problem" (Kerlinger & Pedhazur, 1973, p. 396).

Table 3
Mean Response Times (in ms) and Error Rates (in Percentages) for Pseudocognates and Their Controls in Experiment 1

		Pseudocognates		Controls	
Language Order	List	Reaction Times	Error Rates	Reaction Times	Error Rates
Dutch-English	1	780	30.0	618	4.7
Dutch-English	2	738	22.7	622	1.3
English-Dutch	1	757	26.7	607	5.7
English-Dutch	2	781	25.3	647	1.0

Comparisons Within Language Orders, Between Tasks

To substantiate the above observation that in Condition Dutch-English semantic variables play a smaller role in recognition than in production, we calculated z-scores on the basis of the Fischer-zs of all six pairs of coefficients (one for production and the corresponding one for recognition) involving a semantic predictor on the one hand and RT on the other (see Hays, 1963, p. 532, for this type of analysis). When translation production Dutch to English (forward) and translation-recognition Dutch-English were thus compared, indeed five out of the six comparisons turned out to be significant. They are marked with one ($p<.05$) or two ($p<.01$) asterisks in Table 2. The exception is the comparison involving Definition Accuracy-English. The reason may be that this predictor variable correlates relatively highly ($r=.58$ in de Groot et al., 1994) with Familiarity-English, as if participants instructed to rate English words on how easy it would be to define them in fact perform some familiarity assessment of those words (cf. the factor analysis discussed earlier). A comparison of English to Dutch (backward) production and English-Dutch recognition, did not show any significant differences.

These same comparisons of translation production and translation recognition, but now involving the correlations between RT on the one hand and the *nonsemantic* predictors on the other, showed that two of them were significant, both concerning the length variables (Table 2). They suggest a larger role of length in forward translation production than in the corresponding recognition condition.

Comparisons Between Language Orders, Within Translation Recognition

An additional set of analyses, comparing the corresponding coefficients of the two language-order conditions in translation recognition (for instance, comparing the coefficients for the correlation between Imageability-Dutch and RT: –.16 and –.22 in Table 2), revealed no significant differences, neither within the pairs of coefficients involving semantic predictors, nor within those involving nonsemantic predictors. The finding that the semantic predictor variables do not show a directional effect in translation recognition contrasts with the finding obtained in earlier production studies (Kroll & Stewart, 1994; de Groot et al., 1994), where the involvement of conceptual memory was larger in forward than in backward translation. We will come back to this finding below.

Concentrating on the semantic variables, the correlational data reported above support the conclusion that conceptual memory is implicated in translation recognition. The second of the above comparisons (between language orders, within translation recognition) showed that the order in which the translations are presented, Dutch word or English word first, has no effect on the magnitude of the semantic effects in translation recognition. The first comparison (within language orders, between tasks) had shown that semantic variables play a larger role in forward translation production than in forward translation recognition, and that this task difference does not occur in backward translation, where English words precede the Dutch words. The conclusions of the above two comparisons lead to a third conclusion; namely, that semantic variables play a smaller role (in other words, that conceptual memory is implicated less) in translation recognition (in both language-order conditions) and in backward translation production than in forward translation production. The involvement of conceptual memory in translation recognition and in backward translation production appears to be about equally large.

Comparing Pseudocognate and Control Nontranslation Pairs

Recall that 15 of the nontranslation pairs within both List 1 and List 2, and in both Condition Dutch-English and English-Dutch consisted of pseudocognates, words that are perceptually similar but with completely different meanings. A final set of analyses on the data of Experiment 1 focused on these pseudocognates. They were matched on Imageability, Log Word Frequency and length with 15 of the common (i.e., perceptually dissimilar) nontranslation pairs of the same list by language-order condition. The matching always involved the second word of the nontranslation pairs; in other words, it concerned English words in the two lists of Condition Dutch-English, and Dutch words in the two lists of Condition English-Dutch. The reason the matching was based on the second words was that these were the words from which RT was measured. The pseudocognates and their controls within the same list never differed on Imageability, Log Word Frequency and length (p>.70 in all cases). Subsequently, unpaired t-tests were performed on the average translation-recognition RTs (including correct responses only) and the error scores of the pseudocognates on the one hand and their controls on the other. In all four list by language-order conditions the RT for pseudocognates was significantly longer than that for their controls, and in all cases significantly more errors were made to the former type of nontranslation pairs. The results of these analyses are summarized in Table 3.

These findings strongly suggest that the participants were biased toward accepting the stimulus as a translation pair when the words that constitute the stimulus were perceptually similar. Conversely, in the case of dissimilar words they were biased toward a nontranslation response. Such a bias inhibits correct responses to pseudocognates and facilitates correct responses to dissimilar nontranslation pairs. But, potentially more serious, it also affects the processing of true translation pairs: It facilitates correct responses to cognate translations but inhibits correct responses to noncognate translations. This type of bias does not operate in translation production, because in that task only one term of a translation pair is presented per trial. That this term has a cognate or a noncognate translation in the other language will only be clear after the translation has actually been retrieved from memory.

The results of Experiment 1 thus suggest that this bias is operative in translation recognition. They do not point out, however, when exactly during processing it exerts its effect. It may operate early in processing and have the effect that the actual translation-recognition process is bypassed on a certain number of trials (the error scores for noncognates would provide an indication of what percentage of trials). On these trials, where the decision is simply based on perceptual similarity or dissimilarity of the two words in a stimulus, neither the lexical nor the conceptual system needs to be accessed. Processing is thus even

shallower than when the decision is based on tracing the link between the two word-form nodes of a translation pair. But the consequence will be that, as in the latter situation, conceptual memory is not implicated. This should again become reflected in the size of the effects of the semantic variables, small effects suggesting shallow (here only perceptual) processing.

But it may also be the case that the bias only affects the final processing stage, that is, response selection and execution. This is the stage in which the outcome of the actual translation-recognition process is converted into the appropriate response, and this response is executed. Similar bias effects affecting response selection and execution in other "binary-classification" or two-choice tasks (as is translation recognition) have been demonstrated before, for instance in lexical decision to words preceded by semantically related or unrelated prime words (see Neely, 1991, for a review). Experiment 2 was designed to resolve this issue. We approached it by manipulating the felicity conditions of the bias, by blocking the presentation of cognates, noncognates, pseudocognates, and perceptually dissimilar nontranslation pairs. For instance, in one condition all translation pairs could consist of cognates whereas all nontranslation pairs are perceptually dissimilar. In a second condition, all translation pairs could be noncognates and all nontranslation pairs perceptually dissimilar. The first of these conditions provides optimal circumstances for the bias to operate and be effective. In the second condition, however, the perceptual relation between the words—which now does not distinguish between translation and nontranslation pairs—can in no way assist the response process. To the extent that the bias is under the participants' strategic control, they will try to frustrate it. If the bias operates in an early stage, it may be expected (see above) that processing is more shallow in the conditions where it is optimally operative than in the conditions in which it is (partly) disabled. A consequence of this would be that semantic variables exert a smaller effect in the conditions with an unimpeded bias than in those with a frustrated bias. In contrast, if the bias only exerts its effect during response selection and execution, it may be expected that processing depth is independent of whether or not the bias is frustrated. Semantic effects would then be equally large in conditions with optimal versus unfavorable conditions for the bias.

Experiment 2

Method

Participants

Eighty participants from the same population as used in Experiment 1 took part. Based upon their order of arrival in the laboratory, each was assigned to one of four list conditions (Noncognates-Dissimilar, Cognates-Dissimilar, Cognates-Similar, and Cognates-Dissimilar/Similar; see Materials section), 20 participants per condition. We asked the participants on entering the laboratory to rate their comprehension and production abilities in English on a 7-point scale (1=no knowledge of English; 7=knowledge equal to the corresponding skill in Dutch). The mean comprehension ratings of the participants in the four conditions were 5.40 (*SD*=.60), 5.30 (*SD*=.86), 5.35 (*SD*=.67), and 5.40 (*SD*=.68), in the order of the conditions given above. The mean production ratings of the participants in the four conditions were 5.10 (*SD*=.79), 4.60 (*SD*=.82), 4.85 (*SD*=.59), and 4.65 (*SD*=.99), in the same order.

Materials

The experimental materials consisted of four lists of 290 word pairs each, each word pair consisting of a Dutch and an English noun. A sample of the materials is presented in Table 4.

Within each list there were 150 positive word pairs, consisting of translation equivalents, and 140 negative pairs, containing two words that were not translations of each other (see note 1). All positive materials were taken from the stimulus materials used in

Table 4
A Sample of the Stimulus Materials of Experiment 2

	Word Pair Type	
	Translation	Nontranslation
Noncognates-Dissimilar	spijt-regret	aandacht (attention)-blanket
	krant-newspaper	paraplu (umbrella)-meeting
	lawaai-noise	speld (pin)-reason
	aap-monkey	mening (opinion)-lie
Cognates-Dissimilar	les-lesson	aandacht (attention)-blanket
	resultaat-result	paraplu (umbrella)-meeting
	domein-domain	speld (pin)-reason
	schouder-shoulder	mening (opinion)-lie
Cognates-Similar	les-lesson	dier (animal)-diet
	resultaat-result	broek (trousers)-break
	domein-domain	lucht (air)-lunch
	schouder-shoulder	dreiging (threat)-dressing
Cognates-Dissimilar/Similar	les-lesson	dier (animal)-diet
	resultaat-result	broek (trousers)-break
	domein-domain	rook (smoke)-speech
	schouder-shoulder	schoteltje (saucer)-moustache

Note. For nontranslation pairs the English translation of the Dutch words is presented in parentheses.

our earlier translation studies. The positive word pairs in List 1 were the 150 translation pairs with the lowest cognate ratings in those studies (noncognates). The average cognate rating for those pairs was 1.20 (*SD*=.08; Range: 1.00–1.32 on a scale of 1 to 7). The positive word pairs in Lists 2 through 4 were the 150 translation pairs with the highest cognate ratings (cognates). The average cognate rating for those pairs was 5.51 (*SD*=.70; Range: 4.16–7.00 on a scale of 1 to 7). The negative word pairs in List 1 (Noncognates-Dissimilar) and List 2 (Cognates-Dissimilar) all consisted of a Dutch word and an English word that were perceptually dissimilar (the same negative word pairs were used in Lists 1 and 2). The negative word pairs of List 3 (Cognates-Similar) consisted of a Dutch and an English word that were perceptually similar to one another (called pseudocognates in Experiment 1). Finally, the negative stimuli in List 4 (Cognates-Dissimilar/Similar) consisted of both perceptually dissimilar and similar pairs, 50% of each type. All of the 50% similar pairs (70 in all) also occurred as similar pairs in List 3. Similarly, all of the 50% dissimilar pairs (again 70 in all) also occurred as dissimilar pairs in Lists 1 and 2. For all four lists none of the words in a negative pair also occurred among the positive materials of the same list. Prior to the experimental stimuli, we presented the participants with 20 stimuli for practice.[3]

Apparatus and Procedure

The apparatus was the same as used in Experiment 1. Experiment 1 had shown that language-order is immaterial in translation recognition. So, contrary to Experiment 1, the Dutch word now always preceded the English word. In all other respects the procedure of Experiment 2 was identical to that of Experiment 1. The order of the four list conditions (Noncognates-Dissimilar, Cognates-Dissimilar, Cognates-Similar, Cognates-Dissimilar/Similar) rotated systematically among the participants. So, out of every four participants one participated in each of the four list conditions. The experiment lasted about 40 minutes.

Results and Discussion

For each participant in all four list conditions we calculated two mean RTs, one for each of the two levels of the variable Word-Pair Type (translation and nontranslation). When calculating these means, we excluded RTs associated with error responses and those shorter than 100 ms and longer than 1400 ms (the latter two groups combined again made up less than 1% of the complete data set). We also calculated two error scores per participant, one for the translation pairs and the second for the nontranslation pairs. A score was the percentage of errors that the participant made on that particular type of word pairs. We entered these mean RTs and error scores into two 4×2 (List×Word-Pair Type) analyses of variance, one analysis on the RT data and one on the error data. List was treated as a between-subjects variable and Word-Pair Type as a within-subject variable. The data of these analyses are summarized in Table 5.

The two main effects and the interaction between them were all significant on both analyses; List: $F(3, 76)=6.97$, $p<.001$, MSE=11417 on the RT analysis, and $F(3, 76)=10.48$, $p<.0001$, MSE=19.98 on the error analysis; Word-Pair Type: $F(1, 76)=470.93$, $p<.0001$, MSE=671 on the RT analysis, and $F(1, 76)=18.68$, $p<.0001$, MSE=8.65 on the error analysis; List by Word-Pair Type: $F(3, 76)=9.04$, $p<.0001$, MSE=671 on the RT analysis, and $F(3, 76)=4.73$, $p<.01$, MSE=8.65 on the error analysis. Newman-Keuls tests on the means for the four list conditions showed that responding was significantly faster ($p<.01$) and more accurate ($p<.01$) in Condition Cognates-Dissimilar than in the remaining three conditions. The latter three did not differ significantly from one another. The obvious source of this effect is that Condition Cognates-Dissimilar is the only condition where the bias always works in the direction of the required response: Similar words always require a translation response; dissimilar words a nontranslation response. In other words, the required response is always congruent with the natural bias of the participants to accept the stimulus as a translation pair when the words look alike, and to reject it when

Table 5

Mean Response Times (in ms) and Error Rates (in Percentages) for all List by Word-Pair Type Conditions in Experiment 2

	List							
	Noncognates Dissimilar		Cognates Dissimilar		Cognates Similar		Cognates Dissimilar/Similar	
Word Pair Type	Reaction Time	Error Rate	Reaction Time	Error Rate	Reaction Time	Error Rate	Reaction Time	Error Rate
Translation	555	6.7	460	2.9	535	6.4	529	6.1
Nontranslation	619	6.7	537	3.7	657	10.8	622	9.0

they do not look alike. The consequence will be that all responses, both to translation and to nontranslation pairs, will be relatively fast.

The main effect of Word-Pair Type showed that participants responded to translation pairs faster and more accurately than to nontranslation pairs. This finding basically replicates the corresponding effect in Condition Dutch-English of Experiment 1, except that there the effect only occurred on RT. The interaction between List and Word-Pair Type indicated that the difference between translation and nontranslation pairs in terms of both RT and errors was particularly large in Condition Cognates-Similar. Table 5 shows that this effect has to be due primarily to the extremely slow responses to the nontranslation pairs in this condition; that is, the pairs that required a response that conflicted with the participants' bias to accept perceptually similar words as translations.

In the above analyses we collapsed the data of the perceptually similar and dissimilar nontranslation pairs of Condition Cognates-Dissimilar/Similar. Two further analyses, one on RTs and one on percentages of errors, looked at possible differences between these two types of nontranslation pairs in this condition. The RT for dissimilar pairs (576 ms) turned out to be significantly shorter ($p<.0001$) than that for similar pairs (675 ms). Analogously, the error percentage was lower for dissimilar nontranslation pairs (2.9%) than for similar nontranslation pairs (15.0%). This finding replicates within a single list condition what this experiment also shows between the conditions Cognates-Dissimilar and Cognates-Similar. The effect also occurred in Experiment 1 (see Table 3), where the present similar and dissimilar pairs were referred to as pseudocognates and controls, respectively.

In sum, the analyses reported in this section all point at a bias toward an "accept" response in the case of perceptual similarity of the words presented on a trial, and towards a reject response when the words are perceptually dissimilar.

Correlational Analyses

To determine the correlations between 12 of the earlier 13 predictor variables on the one hand (see below) and RT in the various conditions on the other hand, for all four list conditions we calculated a mean RT for each of the 150 translation word pairs, collapsing over participants (correct responses only). Once more, RTs shorter than 100 ms and longer than 1,400 ms were excluded from the data sets (less than 1% of the data in all). The resulting mean RTs were then correlated with the 12 predictor variables. The correlation matrices for the three conditions with cognates as translation pairs (Conditions Cognates-Dissimilar, Cognates-Similar, and Cognates-Dissimilar/Similar) are presented in the last three columns of Table 6. The matrix for the noncognate condition (Condition Noncognates-Dissimilar) is presented in the last column of Table 7. The excluded variable, as compared to Experiment 1, is cognate status. The reason is that within the various list conditions cognate status now hardly varied, because the sets of materials were composed such that they only contained clear cognates or only clear noncognates as translation pairs.

For comparison with the data of the present conditions, in which *all* translation pairs consisted of cognates (blocked), Table 6 also presents the matrix regarding the data of the same 150 cognate-translation pairs, but presented mixed with noncognates, that is, the data associated with these 150 cognate translations from Experiment 1 (Condition Dutch-English). A comparison of the cognate data collected under the mixed circumstances of Experiment 1 with the present blocked data will reveal a possible effect of this presentation manipulation (mixed vs. blocked). Furthermore, for comparison with translation production, the data associated with these same 150 cognate translations in the Dutch to English translation-production condition of de Groot et al. (1994; Experiment 1) are also presented in Table 6. Note that, as in Experiment 1 but unlike in Experiment 2, in that study cognates and noncognates were presented mixed. The Dutch to English production data, rather than the English to Dutch data,

Table 6

Correlations Between the 12 Predictor Variables and RT in Translation Production Dutch to English (de Groot et al., 1994, Experiment 1), Translation Recognition Mixed (Experiment 1), and Translation Recognition Blocked (Experiment 2); Cognate-Data

			Blocked		
			Translation Recognition		
Predictor	Translation Production	Translation Recognition Mixed	Cognates Dissimilar	Cognates Similar	Cognates Dissimilar/ Similar
Imageability-Dutch	–.23	–.21	–.32	–.36	–.34
Imageability-English	–.25	–.23	–.34	–.38	–.34
Context Availability-Dutch	–.38	–.31	–.29	–.40	–.36
Context Availability-English	–.49	–.41	–.38	–.51	–.42
Definition Accuracy-Dutch	–.25	–.17	–.28	–.30	–.30
Definition Accuracy-English	–.44	–.43	–.46	–.51	–.50
Familiarity-Dutch	–.42	–.31	–.29	–.38	–.41
Familiarity-English	–.68	–.57	–.40	–.63**	–.59*
Log Word Frequency-Dutch	–.44	–.33	–.25	–.38	–.38
Log Word Frequency-English	–.46	–.30	–.25	–.38	–.41
Length-Dutch	–.09	–.18	.04	–.05	–.02
Length-English	–.04	–.15	.08	–.01	.06

$p<.05$ if $r>.14$. $p<.01$ if $r>.19$. **$p<.01$ and *$p<.05$ involve the comparisons between Blocked Translation Recognition Cognates-Dissimilar on the one hand and Blocked Translation Recognition Cognates-Similar and Blocked Translation Recognition Cognates-Dissimilar/Similar on the other hand

Table 7
Correlations Between the 12 Predictor Variables and RT in Translation Production Dutch to English (de Groot et al., 1994, Experiment 1), Translation Recognition Mixed (Experiment 1), and Translation Recognition Blocked (Experiment 2); Noncognate-Data

	Translation		
		Recognition	
Predictor	Production	Mixed	Blocked
Imageability-Dutch	–.48	–.16**	–.30
Imageability-English	–.55	–.31*	–.42
Context Availability-Dutch	–.60	–.34**	–.47
Context Availability-English	–.64	–.48*	–.57
Definition Accuracy-Dutch	–.42	–.18*	–.31
Definition Accuracy-English	–.60	–.49	–.58
Familiarity-Dutch	–.42	–.41	–.46
Familiarity-English	–.59	–.65	–.66
Log Word Frequency-Dutch	–.46	–.45	–.52
Log Word Frequency-English	–.42	–.51	–.53
Length-Dutch	.21	–.09	.08
Length-English	.38	.25	.30

$p<.05$ if $r>.14$; $p<.01$ if $r>.19$.
$^{**}p<.01$, $^{*}p<.05$ involve the comparisons between Translation Production and Translation Recognition mixed.

are provided here because the experimental circumstances under which the former of these two data sets was collected are most similar to those of the present Experiment 2 (the Dutch words always preceding the English words). Analogously, in addition to the noncognate data of Experiment 2, Table 7 presents the data of these same 150 noncognates collected in the Dutch-English Condition of Experiment 1, mixed with cognates, and in the Dutch to English translation-production condition of de Groot et al. (1994).

Comparing the Blocked Cognate Conditions

An examination of the correlation coefficients concerning the three blocked-cognate conditions (of Experiment 2) in Table 6

suggests that the differences between the conditions are generally small. Indeed, a statistical comparison of the coefficients within each triplet of coefficients (e.g., the triplet –.32, –.36, and –.34 for the correlation between RT and Imageability-Dutch) revealed that only within one of the triplets significant differences between pairs of coefficients occurred. This triplet concerned the correlations between RT and Familiarity-English. The correlation coefficient for Condition Recognition Cognates-Dissimilar (–.40) was statistically smaller than that for Conditions Recognition Cognates-Similar and Recognition Cognates-Dissimilar/Similar (–.63 and –.59, respectively). In the present context it is particularly interesting that the coefficients for the correlations between RT on the one hand and all of the *semantic* variables on the other never differed between the three blocked-cognate conditions. This finding suggests that conceptual memory is implicated to the same extent in these three conditions. In other words, semantic effects on the processing of cognates in translation recognition are independent of the nature of the nontranslation pairs, similar, dissimilar, or both.

Blocked vs. Mixed Presentation of Cognates and Noncognates

A further comparison of interest is that between the blocked and mixed conditions *within* the translation-recognition task. A brief inspection of the blocked and mixed conditions in Table 6 *suggests* that when performing translation recognition on cognates, semantic variables play a somewhat larger role when these cognates are presented blocked than when presented mixed. Similarly, when performing this task on noncognates, semantic variables *appear* to play a slightly larger role when these noncognates are presented blocked than when mixed with cognates (see Table 7). However, a statistical comparison of the relevant pairs of correlation coefficients did not reveal any significant difference between a coefficient for a blocked condition and the coefficient for the corresponding mixed condition. So the suggestion that there is more semantic processing in the blocked than in the mixed condition is not substantiated statistically. Finally, when comparing the blocked and mixed translation-recognition conditions on the size of the coefficients that concern the *nonsemantic* variables, again no significant differences occurred. In other words, the role of the nonsemantic variables in translation recognition to both cognates and noncognates does not differ statistically in the blocked and mixed conditions.

Comparing Translation Production and Translation Recognition

The appropriate between-task comparisons encompass the recognition data from Experiment 1, where, as in the production study, cognate and noncognate translations were presented mixed. Table 6 shows that for cognates the correlations between RT and the semantic predictors are about equally high in production and recognition (although they tend to be slightly higher in production). This observation was confirmed in a statistical analysis that compared the sizes of the corresponding correlation coefficients (e.g., –.23 and –.21 for the correlation between RT and Imageability-Dutch). These pairwise comparisons revealed that none of the differences between the two coefficients within a pair (one for production and one for recognition) was significant. In contrast, Table 7 indicates that with noncognate materials the semantic variables exerted a larger effect in production than in recognition. Indeed, out of the six critical comparisons (involving a semantic predictor variable; e.g., –.48 vs. –.16 in Table 7), five were significant (at the .05 level or better). The exception was the comparison involving Definition Accuracy-English. Here the effect was in the same direction, but it failed to reach significance. A comparison of production and recognition on the correlations between RT and the nonsemantic variables, indicated that the influence of these nonsemantic variables was, statistically, always the same in the two tasks. This was true for both cognates (Table 6) and noncognates (Table 7).

Cognates vs. Noncognates

A final comparison is that between cognates on the one hand and noncognates on the other under comparable experimental circumstances. As shown by de Groot et al. (1994), in forward (but not in backward) translation production, semantic variables play a larger role when noncognates are translated than when translating cognates (cf. the production data in Tables 6 and 7). To see whether this difference between cognates and noncognates also occurs in translation recognition, we compared statistically the sizes of the corresponding correlations between RT and the *semantic* predictors of the Translation Recognition Mixed conditions of Tables 6 (cognates) and 7 (noncognates) (e.g., –.21 and –.16 for the correlation between RT and Imageability-Dutch). Although the coefficients tended to be slightly larger for noncognates than for cognates, none of the six comparisons revealed a significant difference between the two coefficients within a pair. In other words, when in translation recognition cognates and noncognates were presented mixed, conceptual memory appeared to be implicated to the same extent when processing cognates and noncognates.

We performed the same analyses on the blocked data sets, comparing any of the three blocked cognate sets of Table 6 with the blocked noncognate set of Table 7. The result was basically the same as in the mixed condition. Although also here the coefficients tended to be slightly larger for noncognates than for cognates, with one exception the relevant comparisons (18 in all, 3×6) revealed no significant differences between the coefficients for cognates on the one hand and noncognates on the other. The exception was the comparison of the coefficients for the predictor Context Availability-English in the Cognates-Dissimilar cognate condition (–.38) and the noncognate condition (–.57). This difference was significant at the .05 level. In general then it can be concluded that, unlike in translation production, in translation recognition semantic variables appear to play an about equally large role when cognates or noncognates are processed.

These findings qualify the results of Experiment 1 in one important respect. Recall that Experiment 1 had shown that the role of semantic variables is larger in forward translation production than in translation recognition (with Dutch words in first position and English words in second position). The present analyses show that this conclusion only holds for noncognates. When processing cognates, the effects of semantic variables are equally large in forward translation production and translation recognition. Statistically, these findings occur both when production is compared with mixed recognition and when it is compared with blocked recognition.

Experiment 1 had suggested that participants performing translation recognition are biased by the perceptual similarity or dissimilarity of the words in the presented word pairs. It was proposed that in the case of similarity they are biased to accept the word pair as a translation pair, whereas in the case of dissimilar words they are biased toward a reject response. Two results of Experiment 2 support that suggestion: (a) RT was relatively short when all translation pairs were cognates—and thus perceptually similar—and all nontranslation pairs were perceptually dissimilar (Condition Cognates-Dissimilar). Under those circumstances the bias can be optimally exploited because its direction always converges with the required response; (b) Nontranslation pairs consisting of perceptually similar words were processed more slowly and less accurately than nontranslation pairs consisting of dissimilar words (see also Experiment 1).

In addition to providing additional support for the operation of the bias, the data of Experiment 2 suggest the locus of its effects. If the translation-recognition process itself would be affected by the bias, semantic variables should show relatively small effects in conditions where the bias can be optimally exploited, here, in Condition Cognates-Dissimilar. Under those circumstances the actual translation-recognition process would be bypassed relatively often. In other words, the response would not be based on a retrieval process within a memory structure of the types depicted in Figure 1. In support of this possibility, the coefficients

for the correlations between RT and the semantic variables indeed tended to be somewhat smaller in Condition Cognates-Dissimilar than in the remaining two blocked cognate conditions. However, the differences between the three conditions were generally extremely small and far from significant. Statistically then, there were no differential effects of the semantic variables in the three blocked-cognate conditions. This finding rules out the actual translation-recognition process as the locus of the above bias effects, and suggests that they occur in the processing stage in which the response is selected and executed: In the case of similar translation pairs this stage is shortened; in the case of dissimilar translation pairs it is lengthened (and vice versa in the case of nontranslation pairs).

A final point to note here is that blocking versus mixing cognates and noncognates does not seem to constitute an extremely critical manipulation in translation recognition. Statistically, all variables (semantic and nonsemantic) exerted the same effect under blocked and mixed conditions. This held for cognates as well as noncognates. Yet, the data signaled a slightly larger role of semantic variables with blocked presentation.

Conclusions

The combined data of Experiments 1 and 2 provide a wealth of information on translation-recognition performance. Furthermore, combined with data from earlier production studies, they reveal the similarities and differences between performance in translation production and translation recognition. In so doing, they indicate under what circumstances the two tasks can be used interchangeably.

At a very general level of task comparison, all except one of the word characteristics that earlier work had shown to affect performance in production turned out to affect recognition as well. The exception was the length of the Dutch words of the translation pairs, which appeared to influence production but not recognition (Experiment 1, Table 2).[4] The direction of the observed effects was always the same as in translation production. In other words, translation recognition came about relatively fast on stimuli consisting of words that are (thought to be) easy to imagine, to think a context for, and to define, that are considered familiar and occur frequently in language use, and on stimuli consisting of words that resemble one another perceptually. The percentage of variance accounted for by the complete set of predictor variables was about 60%, as large as it had been in production. We may thus conclude that we now know to a large extent the variables that determine performance in both translation production and translation recognition.

However, when considering the *sizes* of the correlation coefficients, one difference between the two tasks surfaces. From the differences between tasks in the size of the coefficients it can be concluded that the role of semantic variables is larger in forward translation production of *noncognates* than in forward translation recognition (the Dutch word of a translation pair presented before the English word) of these stimuli. All other comparisons showed an equal role of semantic variables in the two tasks (a smaller role than for noncognates in forward production): The backward versions of the two tasks showed no differential effects of task, and the forward versions of the two tasks did not show differential effects on the translation of cognates. In the introduction of this study, effects of semantic variables were presented as indications of the involvement of conceptual memory in the process of word translation. The above results can thus be summarized and rephrased as follows: Conceptual memory appears to be implicated to the same extent in translation recognition (in both directions; see Experiment 1) and in backward translation production. However, it is implicated more in forward translation production, but only when the task involves the translation of noncognates. Forward translation of cognates in the production task again involves the same engagement of conceptual memory as all of the remaining translation conditions. In all then, the only exception to the general picture concerns forward translation of noncognates in the production task. The joint results thus point

towards the following two conclusions: (a) Backward translation production and backward translation recognition may be used interchangeably; (b) forward translation production and forward translation recognition may only be used interchangeably when cognates constitute the stimulus materials.

Recombining the above results leads to the conclusion that translation production shows a directionality effect—a larger role of semantic variables in forward than in backward translation of noncognates—whereas translation recognition does not show such an effect. Conceptually similar directionality effects in the translation-production task have been reported before. Such effects formed the basis for the asymmetry model of bilingual lexicosemantic organization proposed by Kroll and Stewart (1994; see our Fig. 1d). As set forth in the introduction, we can point out a number of processing differences between translation production and translation recognition. Any or all of these processing differences could underlie the present differential result of the two tasks, and further research should pinpoint its source. For now it suffices to say that, on the whole, the similarities between the effects obtained in the two tasks are more striking than the differences.

Concentrating on the recognition task solely, one can conclude that the composition of the stimulus set does not appear to play a crucial role. If only cognate translation pairs are presented, the role of semantic variables is independent of the type of nontranslation pairs—perceptually similar, dissimilar, or both. The only effect of manipulating the type of nontranslation pairs is on the overall response time: Response time is relatively short when all nontranslation pairs are perceptually dissimilar. From the fact that in this condition (Condition Cognates-Dissimilar) the semantic effects are as large as in the conditions that also or exclusively encompass perceptually similar nontranslation pairs (Conditions Cognates-Dissimilar/Similar and Cognates-Similar), we conclude that the effect on response time does not reflect shallower processing in Condition Cognates-Dissimilar. The actual process of translation retrieval in the memory structures of the types illustrated in Figure 1 appears not to differ between the conditions. Instead, the data suggest that the nature of the nontranslation pairs affects the duration of response selection and execution, that is, the stage following the actual translation-recognition process. So, even when the participants could theoretically rely completely on the perceptual similarity of the words within a stimulus pair, as in Condition Cognates-Dissimilar, they appear to access these memory structures and exploit them in the same way as when a response could not be based on perceptual information only.

Another potentially relevant manipulation in translation recognition was the blocking versus mixing of cognates and noncognates. From the statistical analyses we must conclude that this manipulation does not affect performance: The corresponding correlation coefficients in the blocked and mixed translation-recognition conditions were equally large statistically. Yet, an informal comparison of the size of the coefficients suggested that the semantic variables exert a slightly larger effect in the *blocked* condition. Particularly noncognates show this effect of blocking. So, if the purpose of an investigation is to maximize conceptual processing in translation recognition (as suggested by the effects of semantic variables), blocked presentation is preferable to mixed presentation. Finally, whether the L1 term of a translation pair preceded the L2 term or vice versa turned out to be immaterial.

The insights into translation recognition obtained here should not be thoughtlessly generalized to all bilingual studies that are to use this task. As already mentioned in the introduction, bilingual memory structure presumably depends on a number of factors (e.g., the participants' level of L2 proficiency, word type). From the assumed relation between bilingual memory structure on the one hand and translation-recognition performance on the other, one may conclude that with a change in structure, performance in translation recognition should also change (and, conversely, that different patterns of translation-recognition data reflect different memory structures). Indeed, that was the starting-point of two recent studies (de Groot & Hoeks, 1995; Talamas

& Kroll, 1994) that both dealt with the relation between L2 proficiency and the structure of bilingual memory, and that both used translation recognition as a research instrument (the studies indeed both suggested that translation-recognition performance responds to L2 proficiency and the structure of bilingual memory). In other words, had we tested participants from a different population, the data pattern might (indeed, should) have differed from the one obtained here. However, because of this feature of translation recognition—that it responds to experimental manipulations and, to a large extent, it does so the way translation production does—the task constitutes an inviting alternative to the much more laborious production task. So whoever, for whatever reasons, intends to use the translation task in an investigation of bilingualism, should consider opting for the recognition version of that task.

Revised version accepted 2 June 1995

Notes

[1]As will be seen below (e.g., Tables 1 and 5), nontranslation pairs take longer to respond to than translation pairs. They may therefore leave a stronger memory trace than translation pairs. With equal numbers of translation and nontranslation pairs this is likely to result in the participants' subjective experience that in fact more nontranslation pairs are presented, which may set up a response bias. To prevent this, more translation pairs than nontranslation pairs were presented.

[2]We also calculated correlations between the 13 predictor variables and the error scores for the 440 translation word pairs. We have not provided these correlations here (nor in Experiment 2), because they do not provide additional information.

[3]Another potentially interesting condition would have included only noncognate translation pairs and only perceptually similar nontranslation pairs (in this condition the bias would always work against the proper response). We did not include this condition because our exhaustible pool of participants necessitated that we make a choice among all the potentially interesting conditions. However, we believe that the combination of included conditions already provides us with a rather complete picture of what determines performance in translation recognition.

[4]We will not hazard an interpretation of this differential effect at this point, because subsequent analyses on subsets of the data of Experiment 1 and on

the data of Experiment 2 show that the effect of length, unlike the effect of the remaining variables, is rather precarious (cf. Tables 2, 6, and 7).

References

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

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

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. (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: Elsevier.

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

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., & Hoeks, J. C. J. (1995). *The development of bilingual memory: Evidence from word translation by trilinguals*. Manuscript submitted for publication.

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.

Hays, W. L. (1963). *Statistics*. London: Holt, Rinehart, & Winston.

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

Kerlinger, F. N., & Pedhazur, E. J. (1973). *Multiple regression in behavioral research*. New York: Holt, Rinehart, & Winston.

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.), *In practical aspects of memory: Current research and issues* (pp. 389–395). Chichester, England: John Wiley.

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.

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

Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 264–336). Hillsdale, NJ: Lawrence Erlbaum.

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

Paivio, A. (1968). A factor-analytic study of word attributes and verbal learning. *Journal of Verbal Learning & Verbal Behavior, 7,* 41–49.

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.

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

Talamas, A., & Kroll, J. F. (1994). *Form-related errors in second language learning: A preliminary stage in the acquisition of L2 vocabulary*. Unpublished manuscript.

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

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