Journal of Experimental Psychology:
Learning, Memory, and Cognition
1992, Vol. 18, No. 5, 1001-1018

0278-7393/92/$3.00

Determinants of Word Translation

Annette M. B. de Groot
University of Amsterdam, Amsterdam, The Netherlands

Three experiments looked for the determinants of performance in 3 versions of the word-translation task. Experiment 1 contained the normal-translation version and the cued-translation version. In Experiment 2 Ss performed the translation-recognition task. In both experiments word frequency and word imageability were manipulated. Both affected performance in all 3 versions of the task. In Experiment 3 (normal translation), in addition to the effects of frequency and imageability, those of context availability, cognate status, definition accuracy, length of the stimulus words and of their translations, and familiarity were studied. All of them correlated with the performance measures, but only 4 variables accounted for unique translation variance: the frequency of the stimulus word, the frequency of the response word, cognate status, and context availability. These results are discussed in terms of bilingual memory structure.

In the study of language representation in bilinguals a large variety of tasks has been used. Among them are episodic-memory tasks, in which following some study task, subjects are tested on their retention of the study materials, and the language of the study and test materials is varied. Retention may be measured in an explicit-memory task, for instance, free recall or old–new recognition, or in an implicit-memory task, for instance, lexical decision or fragment completion. A popular question in many of these studies is whether the bilingual's knowledge of the two languages is stored in two separate memory systems (the *language-specific* or *dual-code* model) or integrated in a single store (the *language-independent* or *single-code* model). When performance during test is independent of the language of the study materials, this is seen as support for the language-independent model; when performance during test is dependent on the language of the study materials, this is seen as support for the language-specific model. Recent studies that exemplify this approach are those by Durgunoğlu and Roediger (1987) and Smith (1991).

Semantic-memory tasks have also been used, again often with the purpose of clarifying the above representational issue.[1] The most popular one presumably is the bilingual version of the common semantic-priming technique, in which a word in the one language is followed by a semantically related word in a second language, and the effect of the first word on processing the second word is assessed (e.g., Altarriba, 1991; de Groot & Nas, 1991; Kerkman, 1984; Kirsner, Smith, Lockhart, King, & Jain, 1984; Meyer & Ruddy, 1974; Schwanenflugel & Rey, 1986). If such an effect is obtained, this is regarded as support for the integrated-storage model. Its absence is taken to suggest language-specific storage. A second semantic-memory task used to investigate bilingual language representation is word association (e.g., Kolers, 1963; Taylor, 1976). The relevant question is whether the patterns of association obtained when bilingual subjects word-associate to words presented in their first and second languages are or are not the same. If they are the same, one concludes toward language-independent storage; if they are not, one concludes toward language-specific storage.

A further semantic-memory task used is word translation, in which subjects translate words presented in one language into another. But rather than being used on its own, this task is typically used in conjunction with other tasks, most often picture naming in a second language (e.g., Chen & Leung, 1989; Kroll & Curley, 1988; Potter, So, Von Eckardt, & Feldman, 1984). A comparison of response times in word translation and picture naming is meant to solve the issue of whether translation comes about by means of a direct connection between the lexical representations of the translations (the word-association hypothesis) or indirectly by means of an amodal conceptual representation that is shared by the two translations and the picture (the concept-mediation hypothesis). The word-association model predicts shorter response times (RTs) for translation than for picture naming in the second language, whereas the concept-mediation hypothesis predicts equal RTs in the two conditions. The data support the concept-mediation hypothesis when fluent bilinguals serve as subjects, but less fluent bilinguals show a pattern

A number of people contributed substantially to this article. The person I am most indebted to is Peter Starreveld, who skillfully programmed all experiments. Lucia Dannenburg and John Hoeks assisted in collecting and analyzing the data. Jeanette Altarriba, Anny Bosman, Jan Elshout, Judith Kroll, Marilyn Smith, Marcel Veenman, and an anonymous reviewer provided many helpful comments. Parts of this research were presented at the 31st and 32nd annual meetings of the Psychonomic Society in New Orleans (1990) and San Francisco (1991), respectively.

Correspondence concerning this article should be addressed to Annette M. B. de Groot, Department of Psychology, University of Amsterdam, Roetersstraat 15, 1018 WB Amsterdam, The Netherlands.

[1] When distinguishing between episodic- and semantic-memory tasks I am not implying the two can be strictly separated. Indeed, several authors (e.g., Brooks, 1987; Jacoby, 1983) have pointed at episodic influences in a large number of semantic-memory tasks. I merely call those tasks *semantic* that, given a separation of semantic and episodic knowledge in memory (but see, e.g., Hintzman, 1986), could be performed without consultation of episodic knowledge; in contrast, I call those tasks *episodic* that can only be performed when the trace of the relevant episode is contacted.

that is consistent with the word-association hypothesis (Kroll & Curley, 1988). Note that support for concept mediation obtained in these studies at the same time supports the language-independent model. But a restriction of these studies has been that only concrete words have served as stimulus materials (because only these can be easily pictured). These results may thus not generalize to other types of words. As discussed below, there indeed is reason to believe that abstract words are relatively often represented language specifically.

The starting point of the research reported in this article is the assumption that word translation on its own is a useful tool to study bilingual representation and processing. In comparison with some of the other tasks, it appears to be a relatively straightforward and natural instrument for this: Some more so than others (e.g., professional translators), but presumably all bilinguals, are occasionally engaged in word translation in realistic language settings. The purpose of this study is twofold: (a) to identify determinants of word translation from the subjects' native language (L1, here Dutch) to their second language (L2, here English) and (b) to explore different versions of the translation task.

In Experiments 1 and 2 the roles of word frequency (here, the frequency of a word's occurrence in print) and word imageability (the extent to which the referent of the word evokes a mental image) are studied. There are grounds to assume that both variables affect translation performance. Regarding word frequency, it is likely that words used often in monolingual settings also occur relatively often in translation settings. Translating a word will strengthen the memory connection (whether direct or indirect by means of the conceptual representation) between the translations, and the stronger this connection, the more skilled translating between these words will be. Researchers may thus expect shorter translation times and less translation errors for high-frequency words than for low-frequency words.

But to the extent that translation comes about by concept mediation, frequency of word occurrence in monolingual language use will also, indirectly, affect translation performance: Frequent use of an L1-word within monolingual communication settings in the subject's native language, here Dutch, will result in a strong connection between the lexical representation of this L1 word and its conceptual representation; frequent use of the corresponding L2 word within monolingual communication settings in this subject's second language, here English, will result in a strong connection between the lexical representation of this L2 word and its conceptual representation. If this L1 word and its L2 translation share their conceptual representation, there will thus exist a strong indirect (by conceptual memory) connection between them, by means of which translation can readily occur. Thus, word translation may benefit indirectly from frequent use of each word in monolingual settings.

An effect of the second variable, word imageability, may be expected when one considers Paivio's extension of his dual-coding theory to bilingual memory (Paivio, Clark, & Lambert, 1988; Paivio & Desrochers, 1980). This theory assumes the existence of two verbal systems, one for each of the bilingual's two languages, and one image system. Easily imageable, concrete words (word imageability and word concreteness are generally confounded[2]) are represented in both verbal systems and in the image system, whereas words that are difficult to imagine (abstract words) are only represented in the verbal systems. Translations of words that do have a representation in the image system are assumed to share this representation. A connection exists not only between corresponding representations in the two verbal systems but also between these representations and the corresponding representation in the image system. In other words, the translations of a concrete word are both directly connected (cf. the word-association link in Potter et al., 1984), and indirectly connected by their shared image representation (which clearly differs from the amodal conceptual representations assumed in the concept-mediation hypothesis of Potter et al.). But the translations of an abstract word would only be directly connected. It is plausible that in translation tasks concrete words will profit from the additional, indirect connection through the image system, because it provides an additional route to the response.

But an imageability effect on word translation does not necessarily involve image representations. Instead, it could be due to differences between amodal representations of concrete and abstract words. A plausible cause for such differences is that the meanings of corresponding words in two languages often do not completely overlap and that this especially appears to be the case with abstract words (Taylor, 1976). When one thinks of conceptual representations as each consisting of a single node in memory (local representations; Hinton, McClelland, & Rumelhart, 1986), this difference between concrete and abstract words could be reflected in a bilingual memory structure in which translations of concrete words would generally share a conceptual representation, whereas those of abstract words would relatively often have language-specific conceptual representations. In general, the more different the meanings of translations, the more often they would be stored in separate nodes in conceptual memory. Alternatively, one could think of conceptual representations as each consisting of a collection of nodes (distributed representations; Hinton et al., 1986), for instance, one for each of the various meaning elements that constitute the concept. Depending on the degree of meaning overlap between words and their translations, larger or smaller numbers of these nodes would be shared by translation pairs. Concrete words might thus share more of these nodes between languages than abstract words (see de Groot, in press, for a discussion of local and distributed bilingual conceptual memories).

An imageability (concreteness) effect on word translation could be accounted for in terms of either one of these two amodal-bilingual memories: Word translation can only utilize a common conceptual representation if such a representation exists. In the previous view of local representations many pairs of word translations may be represented in language-specific conceptual nodes. In these cases, alternative transla-

[2] In this study, word imageability and concreteness (the extent to which the word's referent can be experienced by the senses) are also confounded. This is why in this article high-imageability words and low-imageability words are occasionally referred to as concrete and abstract words, respectively.

tion routes will have to be used (e.g., translation by word association). Because especially abstract words are often represented language specifically, a concreteness effect may be expected. The distributed view could account for a concreteness effect if one assumes that translation performance is a function of the number of conceptual nodes shared by a pair of translations: The more shared nodes, the better performance will be.

But word frequency and word imageability are not the only variables that are likely to affect translation performance. Experiment 3 investigates the effect on word translation of a number of other variables. The importance of some of these variables was suggested by the results of Experiments 1 and 2. One of these variables is context availability, a variable that is known to be confounded with word imageability (e.g., Schwanenflugel, 1991; Schwanenflugel, Harnishfeger, & Stowe, 1988). One of the goals of Experiment 3 is to find out whether differential context availability of high- and low-imageability words might underlie the imageability effect.

In addition to identifying determinants of word translation, the second purpose of this study is to explore different versions of the translation task. The versions used are (a) normal word translation (Experiments 1 and 3); (b) cued word translation, in which the subjects are given a cue as to what the response should be (Experiment 1); and (c) translation recognition (Experiment 2), in which pairs of words are presented, each consisting of a Dutch and an English word, and the subjects have to decide whether or not the words within a pair consist of translations.

The reason to use cued translation and translation recognition as new tasks in addition to the more conventional normal-translation task is partly practical: When using unbalanced bilinguals as subjects and words of varying difficulty as stimuli (both are done here), quite a few of the stimulus words will not evoke a response. Although missing responses can be informative to the experimenter (indeed they constitute one of the dependent variables in this investigation), they can cause a lot of embarrassment for the subjects. This, in turn, may negatively affect performance. A solution could be not to present too many words that may be expected to cause problems. But then it would be inappropriate to generalize the emerging views on how word translation comes about to more difficult words. A better solution may be to think of ways to uncover knowledge that is present but not rooted strongly enough in memory to be elicited in the relatively hard normal-translation task. The cued-translation task and the translation-recognition task are meant to do just that.

But the use of the translation-recognition task is also of theoretical interest: Unlike the translation-production task, it does not require that the subject produces the translation out loud. Thus, by eliminating the production requirement, production is ruled out as the locus of any effect that will be observed. In other words, if a particular variable shows an effect both in translation production and in translation recognition, it can be concluded that (at least part of) the effect in the production task has to be localized in a processing stage prior to production. Furthermore, the translation-recognition task bypasses the translation-retrieval process required in translation production. As such, performance in this task may be better suited than that in translation production to be compared with performance in other tasks that do not include this retrieval process (e.g., word naming; Kroll and Stewart, 1990).

Experiment 1: Normal and Cued Translation

Method

Subjects. Forty-eight first-year psychology students from the University of Amsterdam, Amsterdam, The Netherlands, participated as subjects. They were all unbalanced bilinguals, with Dutch as their native language and English as their second language. They started to learn English at school, around the age of 12 and for about 3 to 4 hr a week. Their training at the university required them to read mainly in English. At the time of testing they were all quite fluent in English. Half of them took part in the normal-translation condition and the remaining half in the cued-translation condition. They received course credit for participation. On entering the laboratory the subjects were asked to rate on a 7-point scale their comprehension and production abilities in English (1 = *very low*; 7 = *same as in Dutch*). The subjects' mean comprehension ratings in the normal condition were $M = 5.42$, $SD = 0.83$, and in the cued condition were $M = 5.13$, $SD = 0.61$. The subjects' mean production ratings in the normal condition were $M = 4.71$, $SD = 0.91$, and in the cued condition were $M = 4.42$, $SD = 0.88$.[3]

Materials. The test materials consisted of 144 Dutch nouns, 96 of which were critical. The remaining 48 served as fillers. Among the critical words, word imageability (high vs. low; van Loon-Vervoorn, 1985) and printed word frequency (high vs. low; Uit den Boogaart, 1975) were orthogonally varied. Each of the four stimulus groups thus formed consisted of 24 words. The mean imageability ratings of the words in the high-imageability/high-frequency (HI/HF), high-imageability/low-frequency (HI/LF), low-imageability/high-frequency (LI/HF), and low-imageability/low-frequency (LI/LF) groups were $M = 6.58$, $SD = 0.28$; $M = 6.57$, $SD = 0.23$; $M = 3.30$, $SD = 0.81$; and $M = 3.27$, $SD = 0.71$, respectively (ratings were based on a 7-point scale). The mean word frequencies of the words in these groups were $M = 101.6$, $SD = 71.7$; $M = 7.4$, $SD = 3.8$; $M = 94.6$, $SD = 77.3$; and $M = 9.7$, $SD = 5.0$, respectively (based on a corpus of 620,000 words). I attempted to choose words of about equal length across the four groups, but this selection constraint could not be met: The selected HI/HF words were shorter than the words in the remaining groups ($p < .01$). The mean lengths in number of letters in the four groups were $M = 4.8$, $SD = 1.4$; $M = 6.3$, $SD = 2.2$; $M = 6.8$, $SD = 1.9$; and $M = 6.4$, $SD = 1.7$, in the preceding order. Similarly, the English translations of the HI/HF words were shorter than the translations of the words in the remaining groups ($p < .05$). The mean lengths in the HI/HF, HI/LF, LI/HF, and LI/LF groups were $M = 4.5$, $SD = 1.3$; $M = 5.5$, $SD = 1.7$; $M = 6.3$, $SD = 1.9$; and $M = 5.9$, $SD = 1.8$, respectively. A further constraint in selecting the materials was that a word and its translation must be noncognates, that is, words of which the Dutch form and its English translation are dissimilar in sound and spelling. Similarity ratings on the present

[3] The subjects in all but two of the remaining experiments and norming studies to be reported in this article were also drawn from this subject population. The majority of subjects participated in just one experiment or norming study. If a subject participated in two of the studies (never more than two), the experiments were as dissimilar as possible. For instance, no subject took part either in more than one of the actual translation studies or in more than one of the norming studies of Experiment 3.

pairs of translations were collected earlier (de Groot & Nas, 1991). A 7-point scale was used (1 = *very low similarity*; 7 = *very high similarity*). The mean ratings for the four groups of materials were $M = 1.71$, $SD = 0.90$; $M = 1.40$, $SD = 0.37$; $M = 1.77$, $SD = 0.93$; and $M = 1.66$, $SD = 0.88$, again in the previous order. Finally, the materials were selected such that the critical stimulus words either had just one translation in English or had a clearly dominant translation.

Of the 48 fillers, half were HI words (mean imageability rating: $M = 6.54$, $SD = 0.30$), whereas the other half were LI words (mean imageability rating: $M = 3.31$, $SD = 0.83$). The word frequencies of the fillers varied between 0 and 646 ($M = 59.8$, $SD = 99.3$).

Prior to presenting the test words, I introduced 15 nouns, all different from any of the test words, for practice.

Apparatus and procedure. The experiment was run on a Schneider PC 1640 DD, an IBM clone, in a normally lit room. Stimuli were presented in black lowercase letters against a light-grey background on the computer screen. A PASCAL program controlled the stimulus presentation and RT recording. A microphone that activated a voice-operated switch registered subjects' responses. Subjects sat facing the screen at a comfortable reading distance. The experimenter sat to the left of the subject, typing the subject's responses on the computer keyboard and monitoring the workings of the voice switch. Failures of the voice switch to respond to the subject's response and triggering by another sound were registered.

Prior to the experiment the subjects were instructed orally (in Dutch) to speak out loud the English translations of the Dutch stimulus words. They were asked to do so as quickly as possible, while making as few errors as possible and to remain silent in case they did not know the stimulus word's translation. Prior to the stimulus, a fixation stimulus (an asterisk) appeared on the screen for 1 s, slightly to the left and above where the word was to appear. Subsequently, in the normal-translation condition, the stimulus was presented in the middle of the screen. It remained on the screen until the voice switch registered the onset of the subject's response (or of any other sound). I measured RT from the onset of the stimulus. The experimenter typed the subject's response (which did not appear on the screen). Finally, the experimenter touched the RETURN key, and after 1 s the next trial started. The maximum presentation duration for a stimulus was 5 s. Whenever this duration expired, the experimenter typed the word *none* and called the next trial by touching the RETURN key. The stimuli were presented in random order, and the order was different for all subjects. The test materials were presented in groups of 24 stimuli each. After each group, the subject took a brief rest of minimally 10 s. The experimenter initiated the presentation of the next group.

The procedure in the cued-translation condition was identical to that in the normal-translation condition, except that simultaneously with the Dutch stimulus word the first letter of its English translation (or, in case of more than one translation, of its most dominant translation) and a dot for each of the remaining letters appeared on the screen. The stimulus word and cue were presented below one another; the letter part of the cue was in line with the first letter of the stimulus word. The experiment lasted about 25 min.

Results and Discussion

For each subject in both the normal and the cued condition, mean RTs were calculated for the four (critical) conditions that were formed by the two levels of the variables imageability (HI vs. LI) and frequency (HF vs. LF). Also, for both the normal and cued conditions, I calculated mean RTs for all stimuli, collapsed across subjects. In calculating these means, I excluded RTs associated with translation errors as well as voice-switch registration errors. The latter type of error occurred on fewer than 2% of the trials. A response was considered an error when it was not listed among the translations of the stimulus in two popular Dutch-to-English translation dictionaries (Martin & Tops, 1986; ten Bruggencate, 1978).

Table 1

Mean Response Times (RT; in Milliseconds), Omission Scores (OS; in Percentages), and Error Rates (ER; in Percentages) for all Imageability × Frequency Conditions From Experiment 1: Translation Production

	High imag			Low imag			Imag effect		
Frequency	RT	OS	ER	RT	OS	ER	RT	OS	ER
Normal translation									
High	948	1.2	1.4	1,196	8.0	2.4	248	6.8	1.0
Low	1,392	19.3	7.5	1,364	20.5	10.1	−28	1.2	2.6
Effect	444	18.1	6.1	168	12.5	7.7			
Cued translation									
High	849	0.3	0.5	1,028	5.7	0.7	179	5.4	0.2
Low	1,153	12.8	1.4	1,152	16.3	5.6	−1	3.5	4.2
Effect	304	12.5	0.9	124	10.6	4.9			

Note. Imag = imageability.

A 2 (frequency) × 2 (imageability) × 2 (translation condition; normal vs. cued) analysis of variance (ANOVA) was performed on the mean subject RTs, treating imageability and frequency as within-subjects variables and translation condition as a between-subjects variable. Also, the corresponding 2 (frequency) × 2 (imageability) × 2 (translation condition; normal vs. cued) ANOVA was performed on the mean RTs for the stimulus words, treating frequency and imageability as between-items variables and translation condition as a within-items variable. This same pair of analyses was also performed on the omissions per subject and per item, that is, the number of times no translation was given before the maximum stimulus duration (5 s) expired. Finally, this pair of analyses was done on the translation errors. The data are summarized in Table 1.[4]

The analyses of the RT data for subjects (F_1) and for items (F_2) revealed that all three main effects were significant: Responses were 179 ms slower in the normal condition than in the cued condition, $F_1(1, 46) = 10.04$, $p < .01$, $MS_e = 153{,}689$, and $F_2(1, 92) = 61.53$, $p < .001$, $MS_e = 26{,}707$; high-frequency words were responded to 260 ms faster than low-frequency words, $F_1(1, 46) = 93.26$, $p < .001$, $MS_e = 34{,}806$, and $F_2(1, 92) = 30.59$, $p < .001$, $MS_e = 125{,}031$; and high-imageability words were responded to 99 ms faster than low-imageability words, $F_1(1, 46) = 30.57$, $p < .001$, $MS_e = 15{,}542$, and $F_2(1, 92) = 6.91$, $p = .01$, $MS_e = 125{,}031$. The interaction between imageability and frequency was significant on the subject analysis but not on the item analysis, $F_1(1, 46) = 29.43$, $p < .001$, $MS_e = 21{,}244$, and $F_2(1, 92) = 2.41$,

[4] RTs reported in Tables 1 and 2 and in the text are all based on the analyses by subjects. RTs derived from the analyses by items are always somewhat different, but the differences are always small and never critical.

$p > .10$, $MS_e = 125{,}031$. It showed that an imageability effect only occurred for high-frequency words. A Newman-Keuls test on the subject analysis indicated that all differences between the interaction means were significant, except the one between the HI/LF condition and the LI/LF condition. The interaction between condition and frequency was marginally significant on both analyses, $F_1(1, 46) = 2.89$, $p = .10$, $MS_e = 34{,}806$; and $F_2(1, 92) = 3.03$, $.05 < p < .10$, $MS_e = 26{,}707$. This interaction reflects the fact that the frequency effect was larger in the normal-translation condition than in the cued-translation condition. The interaction between condition and imageability—and the three-way interaction between condition, imageability, and frequency—were not significant ($p > .10$ in all cases).

The analyses of omissions showed that more omissions occurred in the normal condition (12.2%) than in the cued condition (8.8%), but this main effect was only significant on the item analysis, $F_1(1, 46) = 2.47$, $p > .10$, $MS_e = 13.16$, and $F_2(1, 92) = 13.98$, $p < .001$, $MS_e = 2.33$. Furthermore, the results showed that fewer omissions occurred with high-frequency words (3.8%) than with low-frequency words (17.2%), $F_1(1, 46) = 87.37$, $p < .001$, $MS_e = 5.69$, and $F_2(1, 92) = 23.57$, $p < .001$, $MS_e = 21.10$. Finally, fewer omissions occurred with high-imageability words (8.4%) than with low-imageability words (12.6%), but this main effect was only significant on the subject analysis, $F_1(1, 46) = 15.81$, $p < .001$, $MS_e = 3.10$, and $F_2(1, 92) = 2.32$, $p > .10$, $MS_e = 21.10$. The interaction between imageability and frequency was significant on the subject analysis, $F_1(1, 46) = 4.11$, $p < .05$, $MS_e = 2.34$, but not on the item analysis, $F_2(1, 92) = 0.46$, $p > .10$, $MS_e = 21.10$. A Newman-Keuls test on the subject analysis showed again that all differences between the interaction means were significant, except the one between the HI/LF condition and the LI/LF condition. The interaction between condition and frequency was significant on the item analysis, $F_2(1, 92) = 4.14$, $p < .05$, $MS_e = 2.33$, but not on the subject analysis, $F_1(1, 46) = 1.69$, $p > .10$, $MS_e = 5.69$: Analogous to the RT analyses, the result suggested that the frequency effect was larger in the normal condition than in the cued condition. The interaction between condition and imageability was not significant, nor was the three-way interaction ($p > .10$ in all cases).

The analyses of translation errors showed that overall more errors were made in the normal condition (5.3%) than in the cued condition (2.0%), $F_1(1, 46) = 38.55$, $p < .001$, $MS_e = 0.78$, and $F_2(1, 92) = 19.52$, $p < .001$, $MS_e = 1.54$. Furthermore, the analyses showed that fewer errors were made to high-frequency words (1.3%) than to low-frequency words (6.1%), $F_1(1, 46) = 96.49$, $p < .001$, $MS_e = 0.68$, and $F_2(1, 92) = 18.56$, $p < .001$, $MS_e = 3.52$. Finally, the results revealed that fewer errors were made to high-imageability words (2.7%) than to low-imageability words (4.7%), but on the item analysis this main effect was only marginally significant, $F_1(1, 46) = 14.74$, $p < .001$, $MS_e = 0.75$, and $F_2(1, 92) = 3.13$, $.05 < p < .10$, $MS_e = 3.52$. The interaction between imageability and frequency was significant on the subject analysis, $F_1(1, 46) = 10.16$, $p < .01$, $MS_e = 0.53$, but not on the item analysis, $F_2(1, 92) = 1.52$, $p > .10$, $MS_e = 3.52$. Further analysis of the normal and cued conditions separately showed that imageability and frequency only interacted significantly in the cued condition. The interaction between condition and frequency was statistically reliable on both analyses: $F_1(1, 46) = 16.28$, $p < .001$, $MS_e = 0.68$, and $F_2(1, 92) = 7.15$, $p < .01$, $MS_e = 1.54$. Again, the frequency effect was larger in the normal condition than in the cued condition. The interaction between condition and imageability and the three-way interaction between condition, imageability, and frequency were not significant ($p > .10$ in all cases).

The data so far can be summarized as follows: Subjects gave correct translations to high-frequency words faster than to low-frequency words, and fewer translation omissions and translation errors occurred on the former. This was the case in both the normal and the cued version of the translation task. Word imageability also affected translation time. In both versions of the task, subjects responded to high-imageability words faster than to low-imageability words, and fewer omissions and errors occurred on the former. But the imageability effect on the numbers of translation errors and omissions was not reliable on the analyses by items. Furthermore, the main effect of imageability was qualified by an interaction between frequency and imageability: The imageability effect on the RT and omission analyses only occurred for high-frequency words. However, this interaction was only reliable on the analysis by subjects.

The two versions of the translation task thus gave rise to about the same pattern of results. The variable condition (normal vs. cued) tended to interact with frequency in a way that the frequency effect was somewhat attenuated in the cued condition. But for the rest, the pattern of results was the same for the normal and cued translation conditions. The only robust effects of changing from normal to cued translation were that responding was sped up by about 200 ms, that about 60% fewer errors were made, and that about 30% fewer omissions occurred.

Although the data point out that word imageability determines the translation of high-frequency words, one aspect of the subjects' performance casts doubts on this. More specifically, it suggests that another variable, context availability (e.g., Schwanenflugel et al., 1988; Schwanenflugel & Shoben, 1983), might in fact underlie the imageability effect: After finishing the experiment, a number of subjects spontaneously reported that often they would retrieve an appropriate context for the stimulus word and then derive their response from this context. The use of this strategy sometimes leads to a complicated chain of processes, eventuating in the response, right or wrong. For example, the stimulus word *spijt* (*regret*) is often translated into *sorry* and *sorrow*. What appears to be happening is the following. After recognizing the stimulus word (Step 1), the subject thinks of an appropriate context for this word and retrieves the presumably most common one for *spijt* in Dutch, which is the phrase *het spijt me* (Step 2). After translating the Dutch phrase into English (*I am sorry*; Step 3), the subject strips off all the words surrounding the translation of the stimulus (Step 4). For a number of subjects processing ends here, and they produce *sorry* as the response. Some subjects realize in time, before responding, that *sorry* is not a noun as is the stimulus word. They then transform *sorry* into a word it closely resembles and that is a noun (Step 5).

This last process before responding results in the (incorrect) *sorrow* response (Step 6).

A further example suggests that the context that is first retrieved from memory may also be an episodic trace in which both stimulus and response are represented. A subject remarked that when presented with the word *plafond* (*ceiling*), he immediately thought of a song by Lionel Richie, titled *Dancing on the Ceiling*. This suggests that, when storing the song's title in memory, he stored its Dutch translation on the same trace. When in the translation task the Dutch word *plafond* was presented, it accessed its corresponding part on the memory trace, rendering the rest of the trace, including the word *ceiling*, available (cf. the encoding-specificity principle; Tulving & Thomson, 1973). In fact, this example suggests that Steps 2 and 3 in the previous example (retrieval of context and translation) may be replaced by a single step wherein the word *spijt* directly accesses a memory trace representing both the Dutch phrase *het spijt me* and the English phrase *I am sorry*. But a relevant similarity between the two examples remains: In both cases, word translation comes about with the retrieval of contextual information.

If contextual information is indeed at least occasionally used during word translation, the readiness with which it can be retrieved from memory will determine responding. It is well-known that word imageability and context availability are confounded variables (e.g., Schwanenflugel et al., 1988). Thus, in word translation, high-imageability words should have an advantage over low-imageability words. In other words, differential availability of contextual information may underlie the present imageability effect.

The role, among others, of context availability in normal-word translation is directly investigated in Experiment 3. But first a new task, translation recognition, is explored. In this task, subjects are shown word pairs consisting of a word in one language (Dutch) and a second word in another language (English). They have to decide whether or not the words within a pair are translations.

Experiment 2: Translation Recognition

Method

Subjects. Forty-eight first-year psychology students from the University of Amsterdam, Amsterdam, The Netherlands, participated as subjects. There were 24 subjects in each of two stimulus-onset-asynchrony (SOA) conditions (Condition 0 and Condition 240; see *Apparatus and procedure* section following). Prior to testing, the subjects were asked to rate their comprehension and production abilities in English on a 7-point scale (1 = *very low*; 7 = *same as in Dutch*). The mean comprehension ratings were $M = 4.89$, $SD = 0.90$ and $M = 5.39$, $SD = 0.78$ for Condition 0 and Condition 240, respectively. The mean production ratings were, in the same order, $M = 4.28$, $SD = 1.07$, and $M = 4.44$, $SD = 0.78$.

Materials. The test materials were 192 word pairs, each consisting of a Dutch and an English noun. The Dutch words in 96 of these word pairs were the critical stimuli of Experiment 1. The English words in these pairs were the translations of the Dutch words. The remaining 96 word pairs each consisted of a Dutch and an English noun that were not translations of each other. In half of them a concrete Dutch word was paired with a concrete English word, or an abstract Dutch word was paired with an abstract English word. In the other half a concrete Dutch word was paired with an abstract English word, or an abstract Dutch word was paired with a concrete English word. There were 24 trials of each of these four negative-trial types.

Prior to the test stimuli, 20 word pairs, formed in the same way as the test stimuli, were presented for practice.

Apparatus and procedure. The apparatus was the same as that used in Experiment 1, except that no voice switch was used. A key on the right side of the computer keyboard was assigned to *yes* responses. A key on the left side was assigned to *no* responses. During the experiment the subjects kept their forefingers on these keys and pushed one of them to give a response. The subjects were asked to respond as quickly and as accurately as possible. After monitoring the subject's performance during the practice session, the experimenter left the room. The experiment lasted about 25 min.

The word pairs were presented under two SOA conditions, Condition 0 and Condition 240. The sequence of events during a trial in Condition 0 was as follows: Prior to a word pair, a fixation stimulus (an asterisk) appeared on the screen for 1 s, slightly above and to the left of where the Dutch word was to appear. Twenty milliseconds after the offset of the fixation stimulus, the Dutch and English words appeared simultaneously in the middle of the screen, both in black lowercase letters against a light-grey background. The English word was always presented one line below the Dutch word. Immediately after the subject's response, one of the words *correct*, *slow*, or *wrong* appeared (in English), in uppercase letters, on the third line below the English word. *Slow* was shown when the response exceeded a 1,200-ms deadline. This feedback remained on the screen for 2 s, after which both word pair and feedback disappeared from the screen. The fixation stimulus reappeared 1 s later. RT was measured from the onset of the word pair. The word pairs were presented in random order, and the order was different for all subjects. The word pairs were presented in groups of 24. After each group the average RT and number of errors for that group were shown on the screen. Minimally 10 s later the subject initiated the presentation of the next group by pushing either response key.

The procedure in Condition 240 was the same as that in Condition 0, except that the English word appeared 240 ms after the onset of the Dutch word. When the English word appeared, the Dutch word remained on the screen. RT was measured from the onset of the English word.[5]

Results and Discussion

For each subject in both SOA conditions four mean RTs were calculated, one for each of the four critical word groups (requiring a yes response) formed by the two levels of the variables imageability and frequency. Also, for both SOAs,

[5] The reason to include Condition 240 was to substantiate de Groot and Nas's (1991) suggestion that a process of translation recognition could have confounded the cross-language repetition-priming (or translation-priming) effects observed in several lexical decision studies, including theirs: After having recognized the target as a word, but before executing the lexical decision response, the subjects may note the fact that prime and target are each other's translations. This could produce a bias toward the appropriate response, speeding it up. But such a process of translation recognition may only affect the lexical decision response if it is executed within the time required for lexical decision. This study could provide the relevant data, but only if they would be collected under circumstances similar to those of the earlier priming study. In that study the prime–target SOA had been 240 ms, hence the present choice of an SOA of 240 ms.

mean RTs for all 96 critical stimulus words, collapsed across subjects, were calculated. In calculating these means, RTs shorter than 100 ms and longer than 1,400 ms were excluded (less than 0.5% in all).

A 2 (frequency) × 2 (imageability) × 2 (SOA) ANOVA was performed on the mean subject RTs, treating imageability and frequency as within-subjects variables and SOA as a between-subjects variable (F_1). Also, the corresponding 2 (frequency) × 2 (imageability) × 2 (SOA) ANOVA was performed on the mean RTs of the stimulus words, treating imageability and frequency as between-items variables and SOA as a within-items variable (F_2). This same pair of analyses was also performed on the corresponding error scores. The data of the noncritical word pairs (requiring a no response) are not reported in detail.[6] The critical data are summarized in Table 2. The analyses of the RT data revealed that all three main effects were significant: High-frequency words were responded to 88 ms faster than low-frequency words, $F_1(1, 46) = 307.72$, $p < .001$, $MS_e = 1{,}220$, and $F_2(1, 92) = 57.72$, $p < .001$, $MS_e = 8{,}018$. High-imageability words were responded to 26 ms faster than low-imageability words, $F_1(1, 46) = 25.38$, $p < .001$, $MS_e = 1{,}233$, and $F_2(1, 92) = 4.76$, $p < .05$, $MS_e = 8{,}018$. Responding was slower in Condition 0 (742 ms) than in Condition 240 (570 ms), $F_1(1, 46) = 64.44$, $p < .001$, $MS_e = 21{,}944$, and $F_2(1, 92) = 1{,}111.11$, $p < .001$, $MS_e = 1{,}323$. Of the interactions, the one between frequency and SOA was significant, $F_1(1, 46) = 6.60$, $p < .05$, $MS_e = 1{,}220$, and $F_2(1, 92) = 10.92$, $p < .01$, $MS_e = 1{,}323$. The frequency effect was somewhat larger in Condition 0 (101 ms) than in Condition 240 (75 ms). The remaining interactions were not significant ($p > .10$ in all cases).

On the analyses of the error scores the only significant effects were the main effects of condition, $F_1(1, 46) = 4.50$, $p < .05$, $MS_e = 4.31$, and $F_2(1, 92) = 8.04$, $p < .01$, $MS_e = 2.41$; and of frequency, $F_1(1, 46) = 111.78$, $p < .001$, $MS_e = 2.98$, and $F_2(1, 92) = 21.49$, $p < .001$, $MS_e = 15.51$. Neither the main effect of imageability nor any of the interactions approached significance ($p > .10$ in all cases).

In summary, both an imageability effect and a frequency effect (on RT) materialized in both SOA conditions. Unlike in the production task the two variables did not interact.

Table 2
Mean Reaction Time (RT; in Milliseconds) and Error Rates (ER; in Percentages) for All Imageability × Frequency Conditions From Experiment 2: Translation Recognition

	High imag		Low imag		Imag effect	
Frequency	RT	ER	RT	ER	RT	ER
			Condition 0			
High	680	3.6	703	4.7	23	1.1
Low	786	15.3	800	17.9	14	2.6
Effect	106	11.7	97	13.2		
			Condition 240			
High	518	2.4	548	3.5	30	1.1
Low	590	12.3	626	12.7	36	0.4
Effect	72	9.9	78	9.2		

Note. Imag = imageability.

Furthermore, RT was shorter in Condition 240 than in Condition 0. This effect has a trivial cause: It must be because in addition to the time required for the actual translation-recognition process and for response execution, the RT in Condition 0 includes the time to recognize both the Dutch and the English word, whereas in Condition 240 the time to recognize the Dutch word is not encompassed (because RT was measured from the onset of the English word). Word recognition takes about 200–250 ms (e.g., Balota & Chumbley, 1985; Rayner & Pollatsek, 1989), so by the time the English word appears the Dutch word has already been recognized.

Finally, the imageability effect was equally large in the two SOA conditions, whereas the frequency effect was larger in Condition 0 than in Condition 240. The reason the frequency effect was relatively large in Condition 0 is again presumably that only in this condition does RT reflect the recognition time for both the Dutch and the English words. It is generally assumed that word recognition is affected by word frequency. Furthermore, the frequency of a word in one language and that of its translation in a second language correlate positively (e.g., $r = .78$ in Experiment 3). In other words, in Condition 0, when a low-frequency Dutch word and its (most of the time also low frequent) translation are presented, RT is negatively affected twice on a single trial. In contrast, in case a low-frequency word pair is presented in Condition 240, only the frequency of the English word will negatively affect the response, because the Dutch word will already have been recognized when RT measuring starts.

A salient aspect of the translation-recognition data is that, in comparison with the translation-production data, the effects were quite small. As pointed out previously, word recognition is presumably affected by word frequency. In fact, the finding that the frequency effect was larger in Condition 0 (in which the RTs encompass the recognition of two words) than in Condition 240 (in which the RTs include the recognition of only one word; see above) suggests the influence of word frequency on word recognition. A portion of the frequency

[6] Overall analyses on the data associated with the critical (positive) and noncritical (negative) trials showed a significant effect of response type (positive vs. negative) in both SOA conditions. Responses were 82 ms (Condition 0) and 60 ms (Condition 240) slower on negative trials than on positive trials ($p < .001$ in both cases). An additional set of analyses including only the data of the negative trials showed that responding to the 24 word pairs consisting of two concrete words (Condition 0: 788 ms; Condition 240: 597 ms) was significantly faster than responding to the 24 word pairs consisting of two abstract words (Condition 0: 864 ms; Condition 240: 670 ms; $p < .001$ in both cases). The mixed conditions, with either a concrete Dutch word combined with an abstract English word or vice versa, produced intermediate RTs (concrete Dutch word combined with abstract English word: 802 ms and 603 ms for Condition 0 and Condition 240, respectively; abstract Dutch word combined with concrete English word: 844 ms and 648 ms for Condition 0 and Condition 240, respectively). Note however that the words in these four groups of negative trials were not strictly matched across groups on a set of variables that may be expected to affect processing speed (e.g., word frequency). This should be taken into account when evaluating these RT differences.

effect may thus have to be localized in the word-recognition stage of processing.

There is a hint in the data that the word-imageability effect does not have to be attributed, not even partly, to word recognition: Unlike the word-frequency effect, the word-imageability effect is not larger (but, if anything, smaller) in Condition 0 than in Condition 240. Yet, on low-imageability trials in Condition 0, RT reflects the recognition of two low-imageability words (it may be expected that word imageability correlates positively between languages) rather than just one in Condition 240. If word imageability affects word recognition, a larger effect should have occurred in Condition 0. Converging evidence for the view that word recognition is not affected by word imageability is provided by Paivio and O'Neill (1970), who showed that word imageability does not affect word-recognition thresholds.

As already alluded to previously, translation recognition consists of more stages than the one of interest here. In addition to the actual translation-recognition process, it includes one or two (depending on the SOA) word-recognition processes. Translation production also includes more stages than the one of interest (translation retrieval)—namely, the recognition of the stimulus word—and a production component in which the product of retrieval is put into a word. A separate experiment was therefore run in which the roles of word frequency and word imageability on recognizing and producing the present set of words were directly investigated. It used the technique used by Balota and Chumbley (1985) to separate effects attributable to the recognition stage in word pronunciation from effects attributable to the production component in that task. The technique consists of comparing effects obtained under standard pronunciation instructions with those obtained under delayed pronunciation instructions. The delayed condition is the same as the normal condition, except that the subjects are asked not to start pronouncing the word until a signal to do so is presented. Care is taken that this signal only occurs after word recognition has been completed. RT is measured from the onset of the response signal. Effects that still materialize must then be attributed to the production stage. Effects observed under standard pronunciation instructions, but that cease to exist under delayed instructions, must be due to the recognition stage of processing. The use of this technique was thought relevant here because it allows an assessment of, first, what parts of the frequency and imageability effects in translation recognition are due to the word recognition component of that task and of, second, what parts of those effects in translation production are due to its word-recognition and word-production components. This information could then be used to assess the effect of these variables on the processes of interest, that is, translation recognition and the retrieval of the translation in translation production.

Normal and delayed pronunciation data were collected for both the English and the Dutch materials used in Experiment 2. Using a subtraction method I found the following: Imageability did not affect the recognition of the Dutch words and caused a small effect on the recognition of the English words. However, this latter effect was only about one third of the imageability effect obtained in the translation-recognition task. It thus seems fair to conclude that in the latter task the larger part of the effect (about two thirds) has to be localized in the actual translation-recognition stage of processing. Furthermore, a word frequency effect on word recognition occurred, but the effect attributable to word recognition was only about one fifth of the frequency effect obtained in translation recognition. It can thus be concluded that the larger part of the word frequency effect (about four fifths) in translation recognition has to be attributed to the translation-recognition stage itself. Finally, the data suggested that only a small portion of the imageability and frequency effects observed in translation production (about 20% of both effects) must be attributed to components of that task other than the actual translation-retrieval process. In summary, it appears that both word imageability (but see the following section) and word frequency affect the actual process of translation recognition (in the recognition task) and translation retrieval (in the production task).[7]

In Experiment 3 the thread that was left earlier is picked up. Recall that Experiment 1 concluded with the suggestion that differential context availability for high- and low-imageability words might underlie the imageability effect observed there (and, for that matter, in Experiment 2). The plausibility of that suggestion is explored in Experiment 3. Furthermore, the potential relevance of a number of other variables for word translation is investigated.

Experiment 3: Normal Translation

Introduction

The data of Experiment 1 showed that word frequency determines performance in translation production. They also suggested that word imageability affects it, but subjects' comments hinted at the possibility that differential context availability of high- and low-imageability words underlies the imageability effect. To pin down the source of the imageability effect, I added context availability as an additional independent variable in Experiment 3.

In the first of the examples illustrating the role of context in word translation (see *Results and Discussion* section of Experiment 1), the retrieved context was a verbal one (a common phrase). This may also have been the case in the second example (the title of a song), although there the relevant context may also have been a whole episode. This suggests that not context availability in general but the availability of verbal context may be the critical variable. To find out which of the two, if any, actually determines word translation, verbal context availability was also included as a variable in this new study.

Still other factors may affect word translation. One may be the cognate status of the translations, that is, whether they are similar in spelling, sound, or both. It is plausible that words

[7] A more detailed report of this study can be requested from A. M. B. de Groot.

similar in form to their translations are easier to translate than words dissimilar to their translations. This may be for either one or both of two reasons. One possible reason is that cognate and noncognate translations may be represented differently in memory. A second possible reason (but see further on) is that translating may occasionally involve a conscious search through the lexicon for a word in the other language that physically resembles the stimulus word. Such a strategy could only be effective in the case of cognate translations, and thus an effect of cognate status may be expected.

Yet another possibly critical variable in word translation was suggested by a number of subjects in Experiment 1. They noted that they experienced problems with abstract words, even very common ones, because they were often not exactly sure of their meaning. A number of researchers have suggested (e.g., the concept-mediation hypothesis by Potter et al., 1984) that translation often comes about through access of the meaning representation of the word. Not knowing the word's meaning could thus indeed hinder translation, because to be able to translate through the meaning representation, such a representation should exist. Therefore, definition accuracy was also included as a factor in Experiment 3.

A further variable that may determine translation performance, but for less interesting reasons, is the length of the translations. As pointed out before, normal translation (as measured in this study, in which the subjects produce the translations aloud) can be divided roughly into three stages: recognition of the stimulus word; retrieval of the translation; and production of the word, in which the product of retrieval is put into a word. The length of the to-be-translated word may affect translation time because it may be expected to affect the stimulus' recognition time. The length of the to-be-produced translation may affect translation time because it plausibly affects the production component of word translation. Such effects may be expected to be relatively small, but knowing their contribution to translation RT is nevertheless desirable because this knowledge will allow a purer assessment of the influence of the more interesting variables.

Context availability (general and verbal), cognate status, definition accuracy, and translations length were included as independent variables in Experiment 3, in addition to the earlier variables word frequency (but now the frequency of both the to-be-translated word and the to-be-produced translation; see the *Materials* section following) and word imageability. Word frequency, assessed by counting the occurrences of words in (most often) written text, is known to be correlated with the variable familiarity (e.g., Gernsbacher, 1984), obtained by having subjects indicate, for instance, on a 7-point scale how familiar they are with every word on a list. Familiarity has been shown to be a better predictor of performance than word frequency in at least one task, namely, lexical decision (Gernsbacher, 1984; Gordon, 1985). It is thus conceivable that subjective familiarity with words in fact underlies the word-frequency effects on word translation observed here. Word familiarity was therefore included as a final variable in Experiment 3. Of course, it was impossible to manipulate all these variables in a factorial experiment. Thus, a correlational approach was taken.

Method

Subjects. Forty-seven first-year psychology students from the University of Amsterdam, Amsterdam, The Netherlands, participated in the actual translation task. On entering the laboratory they were asked to rate their comprehension and production abilities in English on a 7-point scale (1 = *very low*; 7 = *same as in Dutch*). Their mean comprehension and production ratings were $M = 5.02$, $SD = 0.87$, and $M = 4.55$, $SD = 1.06$, respectively.

Materials. The test materials consisted of 458 Dutch nouns. These words varied widely both in word frequency and in word imageability. To assess their imageability I used the same corpus of word-imageability ratings as in Experiment 1 (van Loon-Vervoorn, 1985). To assess the words' frequency, I consulted the Centre for Lexical Information (CELEX) frequency count (Burnage, 1990).[8] In addition to a database of Dutch words (based on a count of 42.5 million printed words), CELEX also includes a database of English words (based on a count of 18.8 million printed words). This permitted the inclusion in the experiment of one more variable, namely, the frequency of occurrence (in English texts) of the words to be produced by the subjects.

In addition to the earlier constraint that the stimulus words should differ along the frequency and imageability dimensions, a further selection constraint was that about equal numbers of stimulus words should have a cognate and a noncognate relation with their translations. Finally, as in Experiment 1, I attempted to select only words with either an unambiguous translation or with a clearly dominant translation. This was particularly relevant now because when translations other than the intended (by the experimenter), but nevertheless correct, are produced, it is virtually impossible to determine accurately the effects of a number of the new variables on translation: Cognate status and the length and frequency of the response word can only be easily determined if there is just one correct response.

Norming studies. To assess the cognate status of the translations, I used data from an earlier cognate-rating study (see de Groot & Nas, 1991, for details) and performed a new rating study (with 25 subjects). Together, the old study and new study contained all 458 Dutch stimulus words paired with their English translations. The subjects in these two studies were presented with a subset of these word pairs, and they were asked to rate on a 7-point scale how similar they thought the words within each pair were (7 = *very high similarity*; 1 = *very low similarity*). The subjects were told that a rating should reflect a combined assessment of both spelling and sound similarity of the word pair under consideration. In both the old and the new rating study about half of the word pairs consisted of cognates (e.g., *grond–ground*), and the remainder consisted of noncognates (e.g., *lichaam–body*), as initially judged by the author. To assess the reliability of the similarity ratings, 134 word pairs occurred both in the old and in the new study. A mean cognate rating for each of the translation pairs, collapsed across subjects, was calculated. The old and new ratings for the 134 pairs common to both studies correlated highly ($r = .98$).

In four other norming studies, the context availability, verbal context availability, definition accuracy, and familiarity scores for all 458 Dutch stimulus words were collected. In all four of the studies subjects (different subjects in different studies; see also Footnote 3) were randomly assigned to two groups. Each group rated 229 out of

[8] This CELEX frequency count only became available at this point of the present investigation. Because it is based on a much larger corpus of texts, this count is likely to provide a better estimation of the frequency of word occurrence than does the Uit den Boogaart (1975) corpus, used to select the materials for Experiments 1 and 2.

the 458 stimulus words. The assignment of words to groups was based on an alphabetical ordering of all 458 words: The first word in alphabetical order was assigned to Group 1, the second to Group 2, the third again to Group 1, and so on. The number of subjects within a group varied between 22 and 32 across the four studies. The words were presented to the subjects in booklets, eight words per page, all words underneath one another, and pages reshuffled in every new booklet. Words were randomly assigned to a page, but care was taken that each page contained both common and relatively rare words, and both concrete and abstract words. In order to have an assessment of the intergroup reliability of the ratings, 38 words were presented to both groups within each study.

In the context availability norming study the instructions given to the subjects were those used by Schwanenflugel and Shoben (1983, as reported in Schwanenflugel et al., 1988) but were translated into Dutch, providing the subjects with different example words. The subjects in this task were asked to rate the words on "how easy it is to come up with a particular context or circumstance in which they might appear" (Schwanenflugel et al., 1988, p. 502). Ratings were made on a 7-point scale (1 = *very hard to think of a context for that word*; 7 = *very easy to think of a context for that word*).

In the verbal context availability norming study the subjects were instructed as follows (but in Dutch):

> It is plausible that words differ on how easy it is to think of a sentence or a sentence fragment in which they occur. For instance, I find it easy to think of a sentence fragment given the word *spijt* and *haak*. I immediately think of *het spijt me* and of *aan de haak slaan*. I find it more difficult with the word *taak*, but eventually I think of *tot taak stellen*. Please indicate on a scale of 1 to 7 how easily you could come up with a sentence or sentence fragment for each of the following words. A 7 represents *very easy*; a 1 represents *very difficult*. So I would rate the words *spijt* and *haak* with a 6 or a 7, and the word *taak* with a 1 or a 2. Please circle the scale value that you have chosen for the word, use the full range of the scale, and check before you leave that you have not skipped words or whole pages.

The Dutch words in these instructions translate to *regret*, *hook*, and *task*, respectively. The associated sentence fragments translate to *I am sorry*, *hook* (as in *hook a fish* or *hook a customer*), and *set a task*.

In the definition accuracy norming study the instructions to the subjects were as follows (but in Dutch):

> Words differ on how accurately they can be defined. Please indicate on a scale of 1 to 7 how accurately you think you could define each of the following words. You should choose 1 if you think you could hardly define the word, and 7 if you think you could define it very accurately. Please circle the scale value that you have chosen for the word, use the full range of the scale, and check before you leave that you have not skipped words or whole pages.

In the familiarity norming study the instructions to the subjects were those of Noble (1953; but in Dutch) but included four other practice words. A further difference was that Noble used a 5-point rating scale, whereas here, as in the above norming studies, a 7-point scale was used (1 = *never seen, heard, or used the word*; 7 = *seen, heard, or used the word nearly every day*).

After data collection a mean score for each of the words in each norming study was calculated, collapsed across the subjects within a group. Also, for each of the norming studies the correlation was calculated between the Group 1 and the Group 2 scores for the 38 words presented to both groups. These two scores highly correlated within the context availability, the definition accuracy, and the familiarity study ($r = .94$, $r = .91$, and $r = .94$, respectively) but only moderately within the verbal context availability study ($r = .58$).

Apparatus and procedure. The apparatus was the same as that used in Experiment 1. The 458 test words were presented in two groups, the same groups of 229 words each as used in the norming studies. The first subject first translated the words of Group 1 and then those of Group 2; the second subject first translated the words of Group 2 and then those of Group 1; the third subject again started with the words of Group 1, and so on. The within-group presentation of the words was randomized and different for all subjects. The procedure was the same as in the normal-translation condition of Experiment 1. After presentation of the first group of words, the subject paused for about 5 min before presentation of the second group started. The test words of both groups were preceded by 15 practice words, which were the same words for both groups.

Results and Discussion

A mean translation RT of correct responses (collapsed across subjects) and a mean error score were calculated for each stimulus word. In calculating the mean RTs, those associated with voice-switch registration errors were excluded from the data. The latter occurred on fewer than 3% of the trials. Furthermore, for each stimulus word an omission score was calculated, that is, the number of subjects who did not respond within 5 s after stimulus onset. Finally, a combined measure of RT, accuracy, and omissions was calculated for each stimulus word (cf. Gernsbacher, 1984, p. 276).[9] Seven stimulus words were removed from further analyses. Five of them had a dominant translation that was not the expected one, and thus no cognate ratings were available for them. The remaining two were not properly presented on the screen: Only 12 character positions were available for the presentation of the stimulus on the screen, but these two words were 13 and 14 letters, causing the last letters to be inadvertently dropped during presentation. The remaining analyses were thus performed on 451 words. The correlation matrix of the 14 variables of Experiment 3 (10 independent variables and 4 dependent variables) is shown in Table 3.

As can be seen in Table 3, the four dependent variables (RT; OS, for omission score; ER, for errors; and CPM, for the combined performance measure) correlated positively with one another. The correlations between the independent and dependent variables were all in the expected direction: Familiarity (FAM), cognate status (CS), imageability (IMA), context availability (CA), verbal context availability (VCA), definition accuracy (DEF), log word frequency of the Dutch stimulus words (LOGF-D), and log word frequency of the English response words (LOGF-E) all correlated negatively with the dependent variables; length of the Dutch stimulus words (LEN-D) and length of the English response words (LEN-E) correlated positively with the dependent variables.

The correlations between IMA and LOGF-D on the one hand and the dependent variables on the other replicate the

[9] Gernsbacher's (1984) combined performance measure was based on z score transformations of the mean RT and number of errors per item. She added these two z scores per item to form the combined performance measure (M. A. Gernsbacher, personal communication, May 17, 1991). Here the same procedure was followed, except that the combined performance measure included the z score transformations of three variables (RT, error score, and omission score).

Table 3
Correlation Matrix of the Variables in Experiment 3

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. RT	—													
2. OS	.82	—												
3. ER	.50	.45	—											
4. CPM	.91	.89	.76	—										
5. FAM	−.38	−.44	−.27	−.42	—									
6. CS	−.39	−.34	−.26	−.38	.10	—								
7. IMA	−.35	−.32	−.27	−.37	.05	.17	—							
8. CA	−.44	−.42	−.25	−.43	.19	.14	.81	—						
9. VCA	−.35	−.31	−.25	−.35	.51	.13	.13	.23	—					
10. DEF	−.32	−.31	−.22	−.33	.05	.12	.83	.81	.08	—				
11. LEN-D	.25	.23	.23	.28	−.09	−.27	−.37	−.25	−.23	−.28	—			
12. LEN-E	.32	.29	.28	.35	−.06	−.23	−.42	−.35	−.12	−.34	.53	—		
13. LOGF-D	−.43	−.49	−.36	−.50	.66	.10	−.02	.08	.45	.01	−.14	−.09	—	
14. LOGF-E	−.40	−.49	−.31	−.47	.56	.13	−.04	.05	.34	.01	−.17	−.24	.78	—

Note. RT = reaction time; OS = omission scores; ER = error rates; CPM = combined performance measure; FAM = familiarity; CS = cognate status; IMA = imageability; CA = context availability; VCA = verbal context availability; DEF = definition accuracy; LEN-D = length of the Dutch stimulus words; LEN-E = length of the English response words; LOGF-D = log word frequency of the Dutch stimulus words; LOGF-E = log word frequency of the English response words.

main findings of Experiment 1. Further notable aspects of the data are, first, that IMA, CA, and DEF highly correlated with one another (*r*s between .81 and .83; the high correlation between IMA and CA replicates the results of Schwanenflugel et al., 1988); second, that the three correlated variables IMA, DEF, and CA were all weakly negatively correlated with LEN-D and LEN-E (words that are easy to define, imagine, or create a context for tend to be shorter than words that are hard to define, imagine, or create a context for); and, third, that VCA correlated higher with FAM, LOGF-D, and LOGF-E than with CA. This latter finding suggests that subjects instructed to rate words on how easy it is to think of a verbal context perform some assessment of the frequency of the words' occurrence instead.

Length of the Dutch words correlated positively with length of the English words. Furthermore, weak negative correlations between CS on the one hand and LEN-D and LEN-E on the other occurred, indicating that cognates tend to be shorter than noncognates. Finally, and not surprisingly, LOGF-D and LOGF-E were highly correlated ($r = .78$).

Multiple regression analyses. To assess the relative contribution of the various predictor variables to the efficacy of prediction, I first performed four multiple regression analyses, each one with RT, OS, ER, or CPM as a criterion variable and including the total set of 10 independent variables as predictors. On the analyses with RT, OS, and CPM as criterion variables, the variance accounted for varied between 46% and 52%; RT: $R = .68$, $F(10, 440) = 37.23$, $p < .0001$; OS: $R = .69$, $F(10, 440) = 39.71$, $p < .0001$; and CPM: $R = .72$, $F(10, 440) = 46.84$, $p < .0001$. In these analyses the same three variables were significant predictors of the criterion, these were CS, CA, and LOGF-D. LOGF-E was significant on two of these analyses (OS and CPM) and marginally significant ($p = .06$) on the RT analysis. When only these four variables were entered into the regression analyses, the variance accounted for was already between 45% and 50%; RT: $R = .67$, $F(4, 446) = 89.27$, $p < .0001$ (now the effect of LOGF-E was significant; $p < .05$); OS: $R = .68$, $F(4, 446) = 96.68$, $p < .0001$; and CPM: $R = .71$, $F(4, 446) = 110.30$, $p < .0001$. In two of the analyses including all 10 predictors (RT and CPM), LEN-E also contributed significantly to prediction, but additional analyses adding this predictor to CS, CA, LOGF-D, and LOGF-E showed that LEN-E only accounted for one more percentage of the variance at the most. Finally, in the analysis with OS as the criterion, FAM contributed significantly ($p < .05$), but an analysis adding this predictor to CS, CA, LOGF-D, and LOGF-E increased the variance accounted for by only 0.5%.

In the analysis with 10 predictors and ER as the criterion, the variance accounted for was only 25%: $R = .50$, $F(10, 440) = 14.82$, $p < .0001$. CS and LOGF-D again contributed significantly to prediction, but CA and LOGF-E did not. Instead, the contributions of IMA and LEN-E were significant. To see whether the effects of CA and LOGF-E could have been suppressed by other variables, I again ran an analysis with only CS, CA, LOGF-D, and LOGF-E as predictors. The variance accounted for was now 22%: $R = .47$, $F(4, 446) = 30.84$, $p < .0001$. In addition to CS and LOGF-D, CA now contributed significantly ($p < .0001$). The contribution of LOGF-E remained insignificant ($p > .10$).[10]

[10] In addition to the subjective measure of cognate status used here, two objective measures were calculated. One of them defined similarity in terms of position-dependent letter overlap between the translations. The second measure was the same, except that the scores were now transformed to a 7-point scale and that the length of the translations was taken into account. The correlations between CS and RT and CS and OS were −.26 and −.20, respectively, for the first of these measures and −.38 and −.33, respectively, for the second. Multiple regression analyses including these measures of cognate status and with both RT and OS as criterion variables showed essentially the same results as the analyses based on the subjective measure of cognate status. In all cases cognate status was a significant predictor of the criterion, in addition to CA, LOGF-D, and LOGF-E.

Context availability, imageability, and definition accuracy. The preceding analyses suggest that context availability is a better predictor of word translation than is imageability. To substantiate this, I performed two series of analyses. In the first series partial correlations between context availability and the dependent variables were calculated, partialing out imageability. These partial correlations were compared with those between imageability and the dependent variables, partialing out context availability. The second series was a set of multiple regression analyses with, again, cognate status and the log frequencies of the stimulus and response words as predictors, but instead of context availability, imageability was now entered as a fourth predictor.

When IMA was partialed out, CA still correlated significantly with translation RT ($r = -.28$), OS ($r = -.29$), and CPM ($r = -.25$), but when CA was partialed out, IMA no longer correlated with these dependent variables; RT ($r = .00$), OS ($r = .04$), CPM ($r = -.03$). When the same procedure was followed with ER serving as the dependent variable, the correlation between ER and CA, with IMA partialed out, was no longer significant ($r = -.06$, $p > .05$), and the one between ER and IMA, with CA partialed out, became very small ($r = -.12$, $p < .01$). The majority of these analyses suggest that context availability is a better predictor than is imageability.

The set of multiple regression analyses lead to the same conclusion: Except for the analysis with ER as criterion, the inclusion of IMA instead of CA led to a smaller percentage of accounted variance than when CA was entered: RT = 41% (was 45%); OS = 43% (was 46%); CPM = 48% (was 50%); but ER = 24% (was 22%).

Imageability correlates strongly not only with context availability but also with definition accuracy. The earlier analyses including all 10 predictor variables suggested that context availability is a better predictor of translation performance than is definition accuracy. To consolidate this suggestion, I again performed the same two sets of analyses as before, but now comparing context availability and definition accuracy.

When DEF was partialed out, CA still correlated significantly with translation RT ($r = -.32$), OS ($r = -.30$), CPM ($r = -.30$), and ER ($r = -.12$). With CA partialed out, DEF no longer correlated with these dependent variables; RT ($r = .07$), OS ($r = .06$), CPM ($r = .04$), and ER ($r = -.03$). In the set of multiple regression analyses with CS, LOGF-D, LOGF-E, and DEF as predictors and RT, OS, and CPM as criterion variables, the percentage of variance accounted for was smaller than when CA instead of DEF was included; RT = 39% (was 45%); OS = 42% (was 46%); and CPM 46% (was 50%). With ER as a criterion, prediction was equally good when DEF or CA was included (22% in both cases).

The majority of these analyses thus converge on the conclusion that context availability is a better predictor of translation performance than either imageability or definition accuracy. In summary, the results so far suggest that the most relevant predictors of word translation are cognate status, context availability, and the two log frequency measures.[11]

Differential analyses. Recall that the data of Experiment 1 showed an interaction between frequency and imageability of the stimulus words, an effect of imageability occurring only for high-frequency words. To see whether this interaction also occurred in Experiment 3, I ordered the stimulus words according to their log frequency. Correlations were then calculated between imageability and the dependent variables for the 200 stimulus words of highest frequency and the 200 stimulus words of lowest frequency separately. The correlations between imageability on the one hand, and RT, OS, and CPM on the other hand, were significantly higher ($p < .05$ or better) for the high-frequency words (−.56, −.50, and −.55, respectively) than for the low-frequency words (−.26, −.29, and −.32, respectively), again suggesting a Frequency × Imageability interaction. However, all correlations, including those concerning the low-frequency words, were significant, evidencing an imageability-effect for low-frequency words. This finding is not inconsistent with that of Experiment 1 (in which no effect of imageability whatsoever occurred for low-frequency words), because in Experiment 3 the frequencies of the low-frequency words varied between very infrequent and reasonably frequent, whereas all low-frequency words of Experiment 1 were of very low frequency. The combined data of Experiments 1 and 3 thus suggest that imageability affects word translation, except when the stimulus words are quite rare.

But if, as several of the earlier analyses suggest, context availability in fact underlies the imageability effect, an interaction similar to the one between imageability and frequency should occur between context availability and frequency. Indeed, context availability on the one hand, and RT and CPM on the other hand, correlated significantly higher ($p < .05$ or better) within the group of high-frequency words (−.59 and −.54, respectively) than within the group of low-frequency words (−.32 and −.38, respectively). Context availability and OS correlated equally highly in both groups of words (−.47 and −.41, respectively).

To compare the size of the frequency effect obtained in the normal translation condition of Experiment 1 (see Table 1) with the one obtained in Experiment 3, I calculated the overall mean RT for all 200 stimulus words of highest log frequency and that for all 200 stimulus words of lowest log frequency. Furthermore, the corresponding omission scores and error scores were calculated. The results of these analyses are presented in Table 4. Similar analyses were performed on further divisions of the data according to the other variables that were

[11] The correlational analyses reported thus far are not totally pure. Recall that one of the constraints in selecting the stimulus words was that they should have just one or a clearly dominant translation. However, although by far the majority did, not all subjects produced the intended translations to the latter group of stimuli. In the reported analyses this fact has been ignored: The CS-, LOGF-E-, and LEN-E-values (these predictors are relevant in this respect) were based on the intended translations. Including the data associated with unintended but nevertheless correct responses introduces an inaccuracy in the analyses. To see whether this may have affected the results, I did all analyses reported thus far once more, but now encompassing only the data associated with stimuli to which all subjects gave the intended response. These analyses concerned 284 out of the earlier 451 words. With one exception, all results of these analyses replicate those of the analyses on the total set of stimuli. The exception is the multiple regression analysis with OS as the criterion. Here LOGF-D does not contribute significantly to the prediction.

Table 4
Mean Reaction Times (RT; in Milliseconds), Omission Scores (OS; in Percentages) and Error Rates (ER; in Percentages) for Divisions of the Data According to (a) the Log Word Frequency of the Dutch Stimulus Words (LogF-D), (b) the Log Word Frequency of the English Response Words (LogF-E), (c) Cognate Status (CS), (d) Context Availability (CA), and (e) Word Imageability (Imag) From Experiment 3: Translation Production

Variable	RT	OS	ER
High LogF-D	1,150	6.4	3.2
High LogF-E	1,148	6.4	3.3
High CS	1,142	8.3	4.2
High CA	1,131	7.1	4.2
High imag	1,211	10.9	4.6
Low LogF-D	1,540	24.0	10.2
Low LogF-E	1,569	25.2	10.6
Low CS	1,494	19.8	8.5
Low CA	1,523	21.8	9.2
Low imag	1,517	20.9	9.4
LogF-D effect	390	17.6	7.0
LogF-E effect	421	18.8	7.3
CS effect	352	11.5	4.3
CA effect	392	14.7	5.0
Imag effect	306	10.0	4.8

shown to be particularly relevant for word translation (the log frequency of the response words, cognate status, and context availability; further details about the cognate status analysis are presented in the General Discussion section that follows). Finally, again for comparison with Experiment 1, these analyses were done on a division of the stimulus words according to word imageability. Table 4 also shows the results of these analyses.

Retrospective analyses of Experiments 1 and 2. By now substantial evidence has been collected that supports the view that context availability is a stronger predictor of normal word translation than is imageability. An issue not addressed yet is whether that can also be said about cued translation (Experiment 1) and translation recognition (Experiment 2). To answer this question, and to assess the role of the other new variables in cued translation and translation recognition, I correlated the mean RTs of all critical stimulus words (collapsed across subjects) in both translation production conditions of Experiment 1 (cued and normal) and in the two translation-recognition conditions of Experiment 2 (Condition 0 and Condition 240) with all but one of the independent variables of Experiment 3. This could be done because all 96 critical stimulus words of Experiments 1 and 2 were included in Experiment 3, so ratings on the various independent variables were available for all of them. The excluded variable was cognate status (recall that all stimulus materials of Experiments 1 and 2 were selected to be noncognates). The correlation matrix is shown in Table 5. The correlation coefficients for the normal-translation (NT) condition of Experiment 1 were very similar to the corresponding correlation coefficients obtained in Experiment 3 (Table 3). The largest difference between corresponding correlation coefficients concerns the correlation between VCA and RT (RT-NT in Table 5; −.35 vs. −.19), but even this difference was not statistically reliable ($p > .10$). It can thus be concluded that Experiment 3 and the normal-translation condition of Experiment 1 closely replicate each other.

But more interesting is that the corresponding correlations in the four different translation conditions are on the whole quite similar. At first sight it appears that FAM, LOGF-D, and LOGF-E correlate higher with RT in the translation-recognition conditions than in the translation-production conditions. In contrast, it appears that CA correlates higher with RT in the translation-production conditions than in the translation-recognition conditions and that LEN-E correlates relatively highly with RT in the NT condition. However, statistically, the correlations between each of these variables and RT were the same across all four conditions. With respect to the remaining variables, some differential results across

Table 5
Correlation Matrix for the Retrospective Analyses of Experiments 1 and 2

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13
1. RT-Cond0	—												
2. RT-Cond240	.83	—											
3. RT-NT	.62	.69	—										
4. RT-CT	.74	.70	.77	—									
5. FAM	−.55	−.53	−.41	−.33	—								
6. IMA	−.16	−.27	−.29	−.28	.08	—							
7. CA	−.29	−.37	−.44	−.40	.21	.84	—						
8. VCA	−.31	−.16	−.19	−.23	.42	.10	.26	—					
9. DEF	−.25	−.35	−.38	−.37	.12	.87	.82	.12	—				
10. LEN-D	.07	.12	.28	.21	−.10	−.27	−.17	−.11	−.22	—			
11. LEN-E	.20	.18	.33	.22	−.04	−.29	−.22	−.14	−.29	.46	—		
12. LOGF-D	−.57	−.52	−.48	−.46	.71	.03	.17	.42	.09	−.16	−.15	—	
13. LOGF-E	−.64	−.58	−.44	−.45	.66	−.08	.04	.31	.03	−.17	−.22	.83	—

Note. RT-Cond0 = reaction time for Condition 0; RT-Cond240 = reaction time for Condition 240; RT-NT = reaction time for normal translation; RT-CT = reaction time for cued translation; FAM = familiarity; IMA = imageability; CA = context availability; VCA = verbal context availability; DEF = definition accuracy; LEN-D = length of the Dutch words; LEN-E = length of the English words; LOGF-D = log word frequency of the Dutch words; LOGF-E = log word frequency of the English words.

conditions occurred: The correlation between IMA and RT was not significant in Condition 0, the correlation between LEN-D and RT was not significant in Condition 0 and in Condition 240, and the correlation between VCA and RT was not significant in Condition 240. In all remaining conditions these variables correlated significantly with RT.

A final set of eight analyses again directly addressed the issue of whether imageability or context availability is the better predictor of performance in the various translation tasks. Multiple regression analyses were performed, with RT in the two translation-production conditions of Experiment 1 and in the two translation-recognition conditions of Experiment 2 as the criterion variable and, first, including CA, LOGF-D, and LOGF-E as predictors. Subsequently, this same set of four analyses was run once more, but with IMA instead of CA as a predictor. In all cases the inclusion of CA as a predictor resulted in a larger percentage of accounted variance than did the inclusion of IMA (the difference was 3.2% on average). With CA included, the variance accounted for varied between 35% and 48%. With IMA included, the variance accounted for varied between 31% and 45%.

In summary, it appears legitimate to generalize the conclusion that context availability is a better predictor of translation performance than is imageability to all versions of the translation task explored in this study.

General Discussion

The purpose of this investigation was twofold: To identify determinants of word translation and to explore different versions of the translation task. In Experiments 1 and 2 the only variables to be manipulated were the frequency and imageability of the Dutch stimulus words. Both variables affected translation production (Experiment 1), and they did so in the same way in the two versions of that task. The effects were about equally large in the normal and cued translation tasks and in both task versions they interacted (on the analyses by subjects) in the sense that the imageability effect only occurred for words of high frequency. Frequency and imageability effects were also obtained in the translation recognition task (Experiment 2), but here they did not interact.

In addition to the frequency and imageability of the stimulus words, Experiment 3 (normal translation production) included the following variables: familiarity, context availability (general and verbal), and definition accuracy of the stimulus words; frequency of the response words; lengths of the stimulus and of its translation; and, finally, cognate status of the translations. All of these variables correlated significantly with the dependent variables, but additional analyses showed that on the majority of the analyses only four variables each accounted for unique translation variance. These variables were (a) frequency of the stimulus words, (b) frequency of the response words, (c) cognate status of the translations, and (d) context availability. Further analyses suggested that the latter may underlie the effects both of imageability and of definition accuracy, two variables that both correlate highly with context availability.

After splitting the data of Experiment 3 into two parts, one associated with high-frequency words and one with low-frequency words, and running correlational analyses on the separate data sets, a similar interaction between imageability and frequency emerged as obtained in Experiment 1. However, an interaction between context availability and frequency also occurred, again suggesting that context availability may underlie the imageability effect. Finally, retrospective correlational analyses of Experiments 1 and 2 pointed out that context availability may also underlie the effects of imageability in translation recognition and cued translation production.

Because the normal and cued translation production tasks produce very similar patterns of data, it seems warranted to conclude that they can be used interchangeably in the study of bilingualism. The cued translation task may extend the scope of bilingual investigations to the inclusion of more difficult stimulus materials and less proficient bilinguals as subjects. Of course, for the cue to facilitate performance, the stimulus words need to be selected such that, as was the case here, they have only one or a clearly dominant translation. If this constraint is not met, the cue could even have a detrimental effect, putting the subject on the wrong track.

In trying to explain the previously mentioned effects, I start here with a version of a simple but well-known model of bilingual word and concept representation (e.g., Chen & Leung, 1989; Kroll & Stewart, 1990; Potter et al., 1984). I then specify the model on some points, so that it may better account for the data. Figure 1 depicts the model. It shows the representation of a pair of Dutch–English translations. Words are represented at the lexical level of representation; their meanings at the conceptual level. Translations are both directly connected (Link T1) and indirectly connected by the conceptual representation that they share (Links T2a and T2b). Potter et al. (1984) concluded against the existence of the direct T1 links, but data by Kroll and Stewart (1990) and de Groot and Nas (1991) clearly suggest their existence. Links T1 and T2 provide two different translation routes (cf. the word-association and concept-mediation hypotheses discussed previously).

If one assumes that link strength determines translation performance, these memory structures can account for most of the reported effects. The explanation of the two frequency effects has already been given (see the introduction to this article). They can be interpreted in terms of the strengths of the T2a and the T2b links. Frequent use of a word within monolingual communication situations in the bilingual's native and second languages, Dutch and English, will result in strong T2a and T2b connections, respectively. The stronger

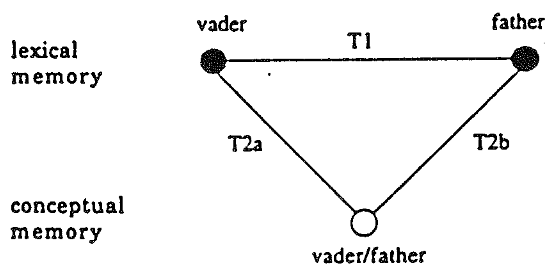

Figure 1. A representation in bilingual memory.

both of these links are, the more successful the translation by conceptual memory will be. An additional cause for the frequency effects was also mentioned previously: Words often used in monolingual settings will also occur relatively often in bilingual translation settings. Each translation act will strengthen the link between the lexical representations of the translations that it traces (T1 or T2). Therefore, translating between these words will gradually become more skilled.

The interaction between word frequency and word imageability—or for that matter, between word frequency and context availability—can be understood in terms of different translation routes for high- and low-frequency words. Recall that beginning bilinguals translate by tracing the T1 links (supporting the word-association hypothesis), whereas fluent bilinguals typically translate by Route T2 (supporting the concept-mediation hypothesis; Kroll & Curley, 1988). It thus seems that at some point during learning their second language, bilinguals switch to conceptually mediated translation. If practice determines this switch (affecting link strength), the use of these varied translation routes by different types of bilinguals may be mimicked within the mental lexicon of a single fluent bilingual by manipulating word frequency. High-frequency words, often practiced in both languages and therefore with strong T2 connections, will relatively often be translated by Route T2, whereas low-frequency words, with much weaker T2 links, will more often be translated by Route T1.

But only if a word is translated by conceptual memory may one expect the effects of variables that reflect meaning aspects of words. Word imageability and context availability are such variables. Effects of imageability and context availability should thus predominantly show up in the case of high-frequency words, translated by conceptual memory. This is in fact what was found in this study. The absence of an interaction between frequency and imageability in translation recognition suggests that the relatively weak T2 links of low-frequency words are still strong enough to be exploited in that task.[12]

Two causes of an effect of cognate status were briefly mentioned before: Cognates and noncognates could be represented differently, or translation might occasionally involve a conscious search through the lexicon for an L2 word physically resembling the stimulus word. In fact, the presence of as many as 50% cognates among the stimulus materials of Experiment 3 could have encouraged the subjects to exploit such a strategy. If this strategic process is the source of the effect, one might expect it to be larger when the proportion of cognates among the stimulus materials is relatively large. This turns out not to be the case.[13] It thus seems warranted to conclude that the effect is due to representational differences between cognates and noncognates.

One plausible way that the representation of cognates and noncognates differs is that the T1 links could be stronger in the case of cognates. Collins and Loftus (1975) already suggested that orthographically and phonologically similar words are linked in lexical memory. Although they never explicitly considered the memory structure of a bilingual, there is no reason that this type of link should be restrained to words within the same language. In other words, in addition to being translations, cognates have an extra reason to be linked in lexical memory. This could be reflected in relatively strong T1 links.

If cognates indeed have stronger T1 links, their translation will occur relatively often by this route. According to the same logic as applied above when explaining the interaction between imageability or context availability and word frequency, it can be predicted that effects of variables that reflect

---

[12] Kroll and her colleagues (Kroll, Altarriba, Sholl, Mazibuko, & Stewart, 1991; Kroll & Stewart, 1990) showed that, at least for some types of words, an asymmetry exists in translation from the subjects' first to their second language, and vice versa. Indications for this were that translating from L1 to L2 took longer than translating in the reverse direction and that the L1-to-L2 translation process was influenced by conceptual variables, whereas translating from L2 to L1 was not. This asymmetry may provide an explanation for the hint of an Imageability × SOA interaction (although not statistically significant) in the translation recognition experiment (Experiment 2), with the imageability effect being smaller in Condition 0 (19 ms) than in Condition 240 (33 ms). It is possible that in Condition 0, both the Dutch word and its English translation being present simultaneously, the subject is able to use the L2-to-L1 translation route. Consequently, smaller conceptual effects would be predicted for this condition than for Condition 240. I thank J. F. Kroll for pointing out this explanation to me.

[13] To find out whether the cognate effect is a strategic one, I collected translation data from 28 additional subjects. Fourteen of them participated in a condition with a high proportion of cognates (HPC); the remaining 14 were tested in a condition with a low proportion of cognates (LPC). The subjects in both conditions translated 200 words from Dutch to English. Out of these 200 words, 150 were the critical materials for both groups of subjects. The remaining 50 were fillers. Out of the 150 critical stimuli, 100 were noncognates, and 50 were cognates. In the HPC group, all fillers were cognates; in the LPC group they all were noncognates. Thus, the proportion of cognates to noncognates in the HPC group was 50:50 (as in Experiment 3), whereas it was 25:75 in the LPC group. The cognate ratings for the cognates varied between 5.1 and 7.0 on a 7-point scale; those for the noncognates varied between 1.0 and 1.4. If the effect of cognate status is a strategic effect, one might expect CS to be a better predictor of translation performance in Condition HPC than in Condition LPC. In fact, the correlations between CS and RT and CS and OS (based on the data of the 150 critical stimuli) were the same in both conditions (HPC: −.45 and −.34, respectively; LPC: −.44 and −.38, respectively). Converging evidence that the effect of cognate status is not strategic comes from a set of ANOVAs run on the same data. For each subject in both proportion conditions three mean RTs were calculated, one for the 50 critical cognates and two for the 100 critical noncognates (randomly split into two groups of 50). Also, for both proportion conditions, mean RTs for all 150 critical stimuli were calculated, collapsed across subjects. On these means, two 3 (word groups: noncognates 1, noncognates 2, and cognates) × 2 (proportion condition) ANOVAs were performed, one by subjects and one by items. This same pair of analyses was also performed on the omission scores and on the error scores. If the effect of cognate status is a strategic effect, an interaction between word groups and proportion condition should occur, with a larger difference between cognates and the two groups of noncognates in Condition HPC than in Condition LPC. However, on none of the six ANOVAs was the interaction statistically significant. A main effect of cognate status occurred on all analyses. Particularly noteworthy was the size of the effect on the RT analyses: It was well over 300 ms in the various conditions.

meaning aspects of words (e.g., imageability and context availability) should be smaller in the case of cognates than in the case of noncognates. The data indeed contained some indications consistent with this prediction: All stimulus words were ordered according to their cognate status, and the correlations were calculated between the independent and dependent variables for the 200 words with the highest cognate ratings and the 200 words with the lowest cognate ratings separately. The correlations between context availability on the one hand and three of the dependent variables on the other hand turned out to be significantly lower for the cognates (−.31, −.15, and −.35 for the correlations between CA and OS, CA and ER, and CA and CPM, respectively; all $ps < .05$ or better) than for noncognates (−.52, −.34, and −.53, respectively; all $ps < .01$). The correlation between CA and RT was equally high within both groups of words. The same pattern was observed for the correlation between imageability and the dependent variables, but here the differences between the corresponding correlation coefficients were somewhat smaller and failed to be significant.

In addition to differing in the strength of T1 links, cognates and noncognates may also be represented differently at the conceptual level. In a study on semantic priming within and between languages, de Groot and Nas (1991) obtained differential interlingual priming effects for cognates and noncognates. In fact, when the prime was degraded, such that it could not be identified by the subjects, the interlingual effect for noncognates disappeared altogether. At the time they concluded that noncognate translations do not share a representation at the conceptual level, whereas cognates do. This conclusion may have been too strong. An alternative to this view is one in terms of the distributed conceptual representations discussed before (see the introduction to this article): Cognate translations may share more of the nodes in these distributed representations than noncognates (cf. an interpretation of the concreteness effect that follows). Recently, I (de Groot, in press) discussed a number of reasons why such might be the case.

What remains is to explain why, when words are translated by conceptual memory (high-frequency words; see above), performance is better for concrete than for abstract words, or if context availability is the relevant variable, why translation performance is better for words for which it is easy to retrieve a context (or, at least, subjects think it is easy to do so) than for words for which it is difficult to retrieve a context. Possible explanations of a concreteness effect were suggested previously. One of them was in terms of Paivio's dual-coding theory (e.g., Paivio & Desrochers, 1980), but the fact that imageability fares worse than context availability in the majority of the analyses provides a serious problem for this interpretation. The second explanation was in terms of amodal representations. Two versions of an interpretation in terms of amodal representations were suggested, one assuming local conceptual representations, the other assuming distributed conceptual representations. In terms of the local view it was suggested that translations of concrete words generally share a conceptual representation, whereas translations of abstract words are represented in separate conceptual representations. In terms of the model (Figure 1), for concrete words two routes to the translation response exist (T1 and T2), whereas for abstract words there is just one route (T1). If additional routes are beneficial to performance, this representational difference may explain the concreteness effect.

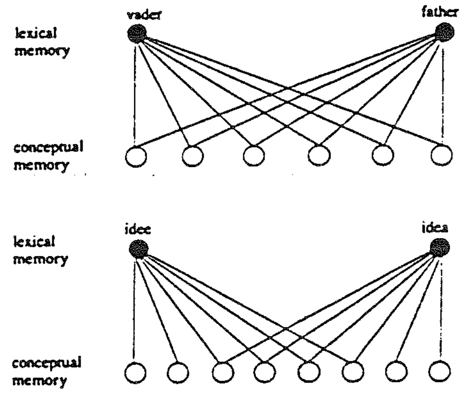

Figure 2. Distributed conceptual representations in bilingual memory.

In terms of the distributed view it was suggested that translations of concrete words share more nodes of their distributed conceptual representation between languages than do abstract words. Figure 2 depicts this situation (T1 links are not shown). A single concept is now represented in a number of nodes in conceptual memory, each one representing, for instance, an individual meaning element of the word. The translations of the concrete word *father* share all of the nodes of their conceptual representation, but the translations of the abstract word *idea* share only a subset of these nodes. The process involved in translation (by the connections between nodes within memory structures of the types depicted in Figures 1 and 2) presumably is spreading activation (see, e.g., Collins & Loftus, 1975). The more conceptual elements are shared by a pair of translations, the more activation will spread from the lexical node of a word to that of its translation and, hence, the more available the translation response will be.

Another representational difference between concrete and abstract words may also be relevant in this context. From the data of a continued-word-association experiment I concluded earlier (de Groot, 1989) that conceptual representations of concrete words contain more information than those of abstract words. In terms of the structures depicted in Figure 2, the conceptual representations of concrete words consist of more nodes than those of abstract words. It is plausible that this difference could also account for the concreteness effect, because the more conceptual nodes there are for a word, the more there are to be shared between languages. And again, the more shared elements, the more activation will spread from the lexical node of the word to that of its translation.[14]

[14] As in Experiments 1 and 2, in my word-association study I also manipulated word frequency and word imageability orthogonally. In addition to the continued association data, I collected discrete word-association data, with RT being one of the dependent variables. Although in some ways translation production and discrete word association are similar tasks, a striking difference between the two studies occurred: The imageability effects were considerably larger in

But what if context availability in fact underlies the concreteness effect, and why are the variables concreteness and context availability so highly correlated? It is not clear what subjects actually do when rating context availability of words on a scale, but assuming the representations of Figure 2, it may be that when performing this task they access these structures. When there is a lot of information to be found there, as is often the case when the word is concrete (de Groot, 1989), they may assign the word a high-context availability rating. When there is relatively little information there, as is often the case with abstract words, they may assign the word a low rating. In other words, the context availability measure may reflect the distributed conceptual memory structures assumed here. If correct, this view on the task thus provides an explanation of the high correlation between concreteness and context availability. Performance in the accuracy of definition task may be explained similarly: Presented with a word, subjects access the word's conceptual representation. Finding a lot of information there causes subjects to think the word could be defined accurately; finding little information there causes subjects to think the word could not be defined accurately.

In brief, what I am suggesting is that the number of conceptual elements in these memory structures may determine context availability and accuracy of definition ratings, and (back to word translation) it also determines translation performance. However, if this view is correct, accounting for the earlier examples of contextual information retrieval in word translation (translating *spijt* to *sorry* by *I am sorry*, and *plafond* to *ceiling* by *dancing on the ceiling*) seems to demand that one allows nonlexical (e.g., episodic) knowledge to be contained in these memory representations.

In summary, most of the observed effects can be explained straightforwardly in terms of both the local and the distributed version of the present model of bilingual word and concept representation. However, the explanation of the effect of context availability needs to be substantiated further. The most parsimonious interpretation of this effect is the one presented here, which attributes the effect to information contained in the same memory structures that were regarded as the locus of the remaining effects. If correct, it seems that these structures not only contain generalized knowledge, abstracted from episodes, but also episodic knowledge itself. An alternative but less parsimonious interpretation of the effect would be one in terms of a third route to the translation response that bypasses the present memory structures and instead exploits other knowledge sources, for instance, episodic knowledge.

This study has provided only indirect evidence against Paivio's dual-coding theory as the correct framework for explaining imageability effects in word translation by showing that context availability better accounts for the data than does word imageability. Future work on word translation could provide a more direct test of this theory by applying the techniques used by Schwanenflugel and her colleagues in other domains (the reading of sentences: Schwanenflugel & Shoben, 1983; lexical decision: Schwanenflugel et al., 1988). These techniques are the addition of appropriate contextual information to the words to be responded to and the matching of high- and low-imageability words on context availability. If imageability effects are in fact due to differential context availability of words, the former should disappear when providing context for the words to be translated or when matching these words on context availability.

Finally, this study clearly suggests that different words are represented differently in bilingual memory. The differences involve stronger versus weaker links between the memory nodes of the translations, and shared conceptual nodes for some types of words versus language-specific conceptual nodes for others (or, in terms of the distributed view, differential amounts of overlap between the conceptual representations of translations). As such, the present data converge with those of other studies that manipulated word type in other bilingual processing tasks (interlingual word association and interlingual semantic and repetition priming; de Groot & Nas, 1991; Jin, 1990; Kolers, 1963; Taylor, 1976). It thus seems that neither the language-specific model nor the language-independent model of bilingual representation is correct but that the correct model takes an intermediate position.

discrete word association than in the present word translation production task (about 400 ms vs. about 100 ms here). Also, in word association the effect was equally large for high-frequency and low-frequency words. These findings suggest that word association responses are always based on accessing the conceptual level. As argued in this study, this is not the case with word translation responses. The frequency effects also differed between the studies, but in the reverse direction: The frequency effect was absent in word association, whereas it was quite large (about 250 ms) in translation production. This difference between the two tasks is not surprising given the fact that the frequency effect in translation was attributed to the strength of links that are not exploited in word association.

References

Altarriba, J. (1991, June). *Is there automatic semantic priming across a bilingual's two languages?* Paper presented at the 3rd Annual Meeting of the American Psychological Society, Washington, DC.

Balota, D. A., & Chumbley, J. I. (1985). The locus of word-frequency effects in the pronunciation task: Lexical access and/or production? *Journal of Memory and Language, 24*, 89–106.

Brooks, L. R. (1987). Decentralized control of categorization: The role of prior processing episodes. In U. Neisser (Ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization* (pp. 141–174). Cambridge, England: Cambridge University Press.

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

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

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

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, and Cognition, 15*, 824–845.

de Groot, A. M. B. (in press). Bilingual lexical representation: A closer look at conceptual representations. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning.* Amsterdam: Elsevier.

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

Durgunoğlu, A. Y., & Roediger, H. L., III. (1987). Test differences in accessing bilingual memory. *Journal of Memory and Language, 26,* 377–391.

Gernsbacher, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of Experimental Psychology: General, 113,* 256–281.

Gordon, B. (1985). Subjective frequency and the lexical decision latency function: Implications for mechanisms of lexical access. *Journal of Memory and Language, 24,* 631–645.

Hinton, G. E., McClelland, J. L., & Rumelhart, D. E. (1986). Distributed representations. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 1. Foundations* (pp 77–109). Cambridge, MA: MIT Press.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review, 93,* 411–428.

Jacoby, L. L. (1983). Remembering the data: Analyzing interactive processes in reading. *Journal of Verbal Learning and Verbal Behavior, 22,* 485–508.

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

Kerkman, J. P. M. (1984). Woordherkenning in twee talen [Word recognition in two languages]. In A. J. W. M. Thomassen, L. G. M. Noordman, & P. A. T. M. Eling (Eds.), *Het leesproces* [The reading process] (pp. 139–152). Lisse, The Netherlands: Swets & Zeitlinger.

Kirsner, K., Smith, M. C., Lockhart, R. S., King, M. L., & Jain, M. (1984). The bilingual lexicon: Language-specific units in an integrated network. *Journal of Verbal Learning and Verbal Behavior, 23,* 519–539.

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

Kroll, J. F., Altarriba, J., Sholl, A., Mazibuko, T., & Stewart, E. (1991, June). *Asymmetric connections in bilingual memory.* Poster presented at the 3rd Annual Meeting of the American Psychological Society, Washington DC.

Kroll, J. F., & Curley, J. (1988). Lexical memory in novice bilinguals: The role of concepts in retrieving second language words. In M. Gruneberg, P. Morris, & R. Sykes (Eds.), *Practical aspects of memory: Vol. 2.* New York: Wiley.

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.

Martin, W., & Tops, G. A. J. (1986). *Van Dale Groot Woordenboek Nederlands–Engels* [Van Dale large Dutch–English dictionary]. Utrecht, The Netherlands: Van Dale Lexicografie.

Meyer, D. E., & Ruddy, M. G. (1974, April). *Bilingual word-recognition: Organization and retrieval of alternative lexical codes.* Paper presented at the meeting of the Eastern Psychological Association, Philadelphia, PA.

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

Paivio, A., Clark, J. M., & Lambert, W. E. (1988). Bilingual dual-coding theory and semantic repetition effects on recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 163–172.

Paivio, A., & Desrochers, A. (1980). A dual-coding approach to bilingual memory. *Canadian Journal of Psychology, 34,* 388–399.

Paivio, A., & O'Neill, B. J. (1970). Visual recognition thresholds and dimensions of word meaning. *Perception & Psychophysics, 8,* 273–275.

Potter, M. C., So, K.-F., Von Eckardt, B., & Feldman, L. B. (1984). Lexical and conceptual representation in beginning and proficient bilinguals. *Journal of Verbal Learning and Verbal Behavior, 23,* 23–38.

Rayner, K., & Pollatsek, A. (1989). *The psychology of reading.* Englewood Cliffs, NJ: Prentice-Hall.

Schwanenflugel, P. J. (1991). Why are abstract concepts hard to understand? In P. J. Schwanenflugel (Ed.), *The psychology of word meanings* (pp. 223–250). Hillsdale, NJ: Erlbaum.

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

Schwanenflugel, P. J., & Rey, M. (1986). Interlingual semantic facilitation: Evidence for a common representational system in the bilingual lexicon. *Journal of Memory and Language, 25,* 605–618.

Schwanenflugel, P. J., & Shoben, E. J. (1983). Differential context effects in the comprehension of abstract and concrete verbal materials. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9,* 82–102.

Smith, M. C. (1991). On the recruitment of semantic information for word fragment completion: Evidence from bilingual priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 234–244.

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.

ten Bruggencate, K. (1978). *Engels woordenboek* [English dictionary]. Groningen, The Netherlands: Wolters-Noordhoff.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80,* 352–373.

Uit den Boogaart, P. C. (Ed.). (1975). *Woordfrequenties in Geschreven en Gesproken Nederlands* [Word frequencies in written and spoken Dutch]. Utrecht, The Netherlands: Oosthoek, Scheltema, & Holkema.

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

Received July 17, 1991
Revision received December 27, 1991
Accepted January 30, 1992 ■