Behavior Research Methods, Instruments, & Computers
1996, 28 (4), 491-503

Covariance structure analysis in experimental research: Comparing two word translation models

JANET G. VAN HELL, PAUL OOSTERVELD, and ANNETTE M. B. DE GROOT
University of Amsterdam, Amsterdam, The Netherlands

Covariance structure analysis is a statistical technique in which a theoretical model, or a covariance structure, is constructed, and the covariances predicted by the theoretical model are compared with those of the observed data. The adequacy of the model in reproducing the sample covariances is reflected by estimates of the parameters of the model and measures indicating the goodness of fit. Covariance structure analysis is frequently used for analyzing data obtained in nonexperimental or quasi-experimental research, but is seldom employed in experimental research. In this paper, the applicability of this technique in experimental research is discussed and illustrated by covariance structure analysis studies in which two models for word translation—the symmetrical model and the asymmetrical model—are described, refined, and contrasted.

The analysis of covariance structures is a statistical technique frequently used for nonexperimental data (e.g., in survey or quasi-experimental research).[1] However, data obtained from research usually labeled as experimental (i.e., research conducted under controlled circumstances in a laboratory) are rarely analyzed using covariance structure models (Bentler, 1980; Bollen, 1989; Royce, 1977). Typically, techniques like analysis of variance (ANOVA) are used. In this paper, we aim to demonstrate that covariance structure modeling can be a useful tool for researchers within the experimental cognitive field, both for hypothesis testing and for exploratory purposes during theory development or model generating (Jöreskog, 1993). We will illustrate the application of covariance structure analysis to this field by comparing two models of bilingual word processing. In doing so, it is not our intention to replace more conventional analysis techniques. Instead, we intend to augment the set of traditional techniques for data analysis in this domain with the analysis of covariance structures.

In short, the aim of covariance structure analysis is to minimize the difference between the observed covariances (i.e., the sample covariances) and the covariances predicted by the covariance structure model (i.e., the population covariances). The precision with which the model can reproduce the sample covariances is assessed using fit functions (described in more detail below) that yield estimates of the parameters of the model and several measures indicating the overall goodness of fit of the model to the observed data.

Before introducing the model specified to describe our data, a formal description of covariance structure models seems in order. Models for the analysis of covariance structures seek to explain the relationships among a set of observed variables in terms of unobserved, or latent, variables. Generally, these models comprise two components: a structural equation model and a measurement model. The structural equation model, or latent variable model (Bollen, 1989), specifies the causal relationships among the latent variables:

$$\eta = B\eta + \Gamma\xi + \zeta,$$

where η is a vector of latent dependent variables, ξ is a vector of latent independent variables, ζ is a vector of errors in equations, B is a matrix of coefficients relating the latent dependent variables to one another, and Γ is a matrix of coefficients relating the latent independent variables to the latent dependent variables. Hence, the structural equation model is a general matrix representation in which the assumed causal relationships between latent variables are described. For example, the element γ_{ij} of Γ describes the causal effect of the latent independent variable ξ_j on the latent dependent variable η_i. Similarly, by fixing the element β_{ij} of B to zero, it is assumed that the latent dependent variable η_i is not affected by η_j.

The measurement model specifies how the latent variables of the structural equation are measured in terms of the observed variables. The measurement model consists of a pair of (confirmatory) factor equations:

$$y = \Lambda_y\eta + \varepsilon$$
$$x = \Lambda_x\xi + \delta,$$

where y is a vector of the observed dependent variables, x is a vector of the observed independent variables, ε and

The authors thank Hanny Comijs, Lucia Dannenburg, John Hoeks, and Rik Poot for their assistance in collecting the data, and Conor Dolan for his help in reporting the analyses. The authors are grateful to Jeroen Raaijmakers, Robert W. Proctor, and two anonymous reviewers for providing insightful comments. Correspondence should be addressed to J. G. van Hell, Department of Psychology, University of Amsterdam, Roetersstraat 15, 1018 WB Amsterdam, The Netherlands (e-mail: pn_hell@ macmail.psy.uva.nl).

δ are vectors of unique factors (i.e., errors in measurement), and Λ_y and Λ_x are matrices of loadings of the observed y variables and the observed x variables on the latent η variables and the latent ξ variables, respectively. Hence, for example, the element λ_{ij} of Λ_x indicates the magnitude of the expected change in the observed variable x_i for a one-unit change in the latent variable ξ_j. Note that the equations of the measurement model in essence describe the multivariate regressions of y on η and of x on ξ.

SYMMETRICAL AND ASYMMETRICAL MODELS OF WORD TRANSLATION

The formal description of the covariance structure model will be exemplified by describing the predictions made by two models for word translation. Performance of subjects in word translation, a task in which bilinguals translate words between their two languages, has brought about two models describing the representation and processing of words in the bilingual mental lexicon: the symmetrical model and the asymmetrical model (de Groot, 1992; de Groot, Dannenburg, & van Hell, 1994; Kroll & Stewart, 1990, 1994; La Heij, Hooglander, Kerling, & van der Velden, in press; Sánchez-Casas, Davis, & García-Albea, 1992). Both models distinguish between a lexical and a conceptual level of representation in bilingual memory. Moreover, both models assume that the two words of a translation pair are stored in separate lexicons, whereas the semantic representations of the translation pair are assumed to be shared and stored in a common conceptual store. So, the symmetrical model and the asymmetrical model agree on the organization of the memory stores in bilingual memory. However, they disagree on their views on the nature of the links that connect the lexical and conceptual stores and, as a consequence, on the nature of processing in forward translation (from the native language, L1, to the second language, L2) and in backward translation (from L2 to L1). The covariance structure analyses to be reported in this paper focus on these divergent predictions.

According to the asymmetrical model, forward translation proceeds from the L1 lexicon via the shared conceptual store to the L2 lexicon. In backward translation, however, the conceptual store is not involved, since backward translation occurs directly from the L2 lexicon to the L1 lexicon. The asymmetrical model predicts that the processing of meaning, and, therefore, the effect of meaning variables, comes into play only in forward translation and not in backward translation. The basis for these predictions is that meaning aspects of words are stored in conceptual memory—so, only when this memory store is implicated will processing be influenced by meaning. Evidence favoring the asymmetrical model of word translation has been obtained by Kroll and Stewart (1990, 1994). They presented words in semantically categorized (all words belonging to the same semantic category, e.g., all garments) or in randomly mixed lists and found that when subjects translated from L1 to L2, the words in the categorized list took longer to translate than did those in the mixed list. In backward translation, however, words of both lists were translated equally fast. Assuming that the categorization/mixing manipulation induces a semantic manipulation on the stimulus materials, Kroll and Stewart's findings suggest that the meaning of words plays a role in forward but not in backward translation. In sum, forward translation proceeds via conceptual memory and backward translation via the lexical links.

Unlike the asymmetrical model, the symmetrical model does not assume directionality in word translation: forward translation and backward translation employ the conceptual and lexical links to the same extent. The symmetrical model predicts that forward and backward translation will be affected by meaning to the same degree. Results reflecting this symmetry have been found by La Heij et al. (in press). Using Stroop-like tasks in which the to-be-translated word was accompanied by a color or a picture, La Heij et al. found that semantic context effects were similar in the two translation directions. In addition, de Groot et al. (1994) found that word characteristics usually regarded as semantic variables (e.g., word imageability) had a strong, negative correlation with the time needed to translate words in forward direction and in backward direction. However, semantic variables played a somewhat weaker role in backward translation than in forward translation. Therefore, de Groot et al. proposed a weak version of the asymmetrical model, in which meaning influences forward and backward translation, though the latter to a lesser extent than the former.

In the covariance structure analyses to be reported here, the extreme versions of the asymmetrical and symmetrical model were tested. So, according to the (strong) asymmetrical model, meaning influences only forward translation but not backward translation, whereas the (strong) symmetrical model states that forward and backward translation will be equally affected by meaning.

In addition, the asymmetrical model and the symmetrical model diverge in their predictions regarding the influence of proficiency in the second language on translation. In general, it is assumed that as people become more proficient in their second language, the L2 lexicon will function increasingly independently from the L1 lexicon, and the links connecting the L2 words (represented in the L2 lexicon) and their accompanying concepts (represented at the conceptual level) will become increasingly strong (de Groot & Hoeks, 1995; Kroll, 1993; Kroll & Sholl, 1992). According to the asymmetrical model, a higher proficiency in L2 will be reflected in a growing influence of semantic variables on forward translation, but not on backward translation (Kroll, 1993). The symmetrical model, however, predicts that, with greater proficiency in L2, semantic variables will have a growing impact on both forward and backward translation.

In short, the main difference between the asymmetrical model and the symmetrical model is that the former predicts an influence of meaning in forward translation

but not in backward translation, whereas the latter states that both forward and backward translation will be equally affected by semantic variables. Furthermore, as more fluency in L2 is gained, the asymmetrical model predicts that meaning will influence forward translation to an increasing extent, whereas backward translation will remain unaffected. In contrast, the symmetrical model predicts that both forward and backward translation will be increasingly affected by semantic variables with growing proficiency in L2.

In the covariance structure analyses reported here, both models were tested on exactly these divergent predictions. To this aim, a covariance structure model was constructed (to be described in more detail below). This covariance structure model described the word characteristics that influence word translation (i.e., the determinants of word translation) and not the representation of words in bilingual memory. Hence, no mention will be made of a lexical level and a conceptual level. Instead, different (observed) word characteristics (e.g., imageability) were described in terms of latent variables, one of them being a semantic variable. Subsequently, the influence of these latent word characteristics—in particular, the semantic characteristics—on the speed of word translation was studied. Both forward and backward translation by (relatively) low- and high-proficient bilinguals was considered.[2] In this way, the above-mentioned predictions of the asymmetrical model and the symmetrical model were tested. Data collected in earlier studies (de Groot, 1992; de Groot et al., 1994) were used as the sample covariances against which both models were tested in the covariance structure analyses. So, before the covariance structure model is described in more detail, the experiments of de Groot (1992) and de Groot et al. (1994) will be discussed.

UNIT OF ANALYSIS: WORDS

The studies of de Groot (1992) and de Groot et al. (1994) were designed to reveal which word characteristics affect forward and backward word translation and to what extent. In addition, the predictions made by the asymmetrical model and the symmetrical model were tested. Two groups of 52 bilinguals with Dutch as their native, and dominant, language and English as their second language participated in these studies. All subjects were first-year psychology students of the University of Amsterdam who started to learn English at school around the age of 12 years for about 3–4 h a week until their enrollment in the university. The first group of subjects ("lower proficient") was tested about 4 months after admission, and the second group ("higher proficient") was tested about 11 months after entering the university. Since the curriculum required them to read mainly in English, the latter group of subjects could be expected to be more fluent in English. The subjects were asked to translate 440 words. Of both the lower and the higher proficient group, half of the subjects performed forward translation, and the other half performed backward translation.

The experimental procedure was as follows (for more details, see de Groot, 1992; de Groot et al., 1994). Upon presentation of a stimulus word on the computer screen, the subjects spoke out loud its translation. They had been instructed to do so as quickly as possible while making as few errors as possible. Reaction times (RTs) were measured from the onset of the stimulus until the onset of the subject's response. The onset of the subject's response was registered by a voice switch attached to the computer. After data collection, a mean forward-translation RT was calculated for each (Dutch) stimulus word, collapsed across the subjects who had provided a correct response for that word; this procedure was followed for the lower proficient and the higher proficient groups separately. Similarly, a mean backward-translation RT was calculated for each (English) stimulus word, for the lower proficient and higher proficient groups separately. In doing so, we obtained four groups of translation RTs for the same 440 words, which will be considered the observed dependent variables in the covariance structure analyses (see below): mean forward-translation RT of lower proficient subjects, mean forward-translation RT of higher proficient subjects, mean backward-translation RT of lower proficient subjects, and mean backward-translation RT of higher proficient subjects.

Prior to the collection of the translation data, norming studies were performed or existing corpora were consulted to obtain assessments of 11 characteristics of the 440 words to be translated. These 11 word characteristics were assumed to influence forward and/or backward translation to a lesser or greater extent and, hence, will be regarded as the observed independent variables in the covariance structure analyses (see below). First, the printed word frequency of each Dutch word and its English translation was assessed by consulting the CELEX frequency count (Burnage, 1990). Of these frequencies, the log transformations were taken, as is more often done in psycholinguistic research (e.g., Gordon, 1985; Kirsner, Smith, Lockhart, King, & Jain, 1984).

The ratings concerning the remaining 9 word characteristics were produced by nine different groups of subjects, drawn from the same population as those participating in the translation experiments. Care was taken that none of the subjects of the rating studies participated in the actual translation experiments. In the rating studies, subjects were presented with 440 words in a booklet and were asked to rate these words on a 7-point scale. Except for cognate status, in which both words of a translation pair were rated jointly (see below), all word characteristics were rated for the Dutch words and their dominant English translations in separate studies. The following instructions were used: "how easy or difficult it is to arouse a mental image of the word" (imageability); "how easy or difficult it is to come up with a particular context or circumstance in which the word might ap-

pear" (context availability); "how accurately the word can be defined" (accuracy of definition); "how often the word is heard or used" (familiarity); "how similar the words within the translation pair are" (cognate status; e.g., *huis–house* has a high cognate status, and *boom–tree* has a low cognate status). The instructions reported here are concise descriptions of the gist of the original instructions (cf. de Groot, 1992; de Groot et al., 1994). The exact wording of the scale points was tuned to the word characteristic to be rated but, generally, 1 meant *low* and 7 meant *high* on the variable in question. For example, in the context availability rating study 1 stood for *very hard to think of a context for that word*, whereas 7 meant *very easy to think of a context for that word*. After data collection, a mean score for each word in each norming study was calculated, collapsing across the subjects within a group.

In sum, for a series of 440 words, 11 word characteristics had been assessed in the norming studies of de Groot (1992) and de Groot et al. (1994). These will serve as the observed independent variables in the covariance structure analyses. Furthermore, these 440 words had been translated in forward direction and in backward direction by lower and higher proficient subjects, and the RTs of these words will serve as the observed dependent variables. For both the independent and the dependent variables, a single mean for each of the 440 words had been calculated, collapsed across subjects. These mean word scores will serve as the variables in the covariance structure analyses. In the next section, we will describe the covariance structure model that was constructed to describe these observed independent and dependent variables in terms of latent variables and that specifies the causal relationships between latent variables. With this covariance structure model, the predictions made by the symmetrical model and the asymmetrical model were tested.

COVARIANCE STRUCTURE MODEL

The data obtained in the nine rating studies and the log frequencies of the Dutch and English words comprised the measurement model of the observed independent variables. So, in total, there are 11 observed independent variables (see Figure 1): imageability of the Dutch word and its English translation; context availability of the Dutch word and its English translation; definition accuracy of the Dutch word and its English translation; familiarity of the Dutch word and its English translation; frequency of the Dutch word and its English translation; and the cognate status of the Dutch–English translation pair.

Initial factor analyses had revealed that, with the exception of cognate status, the observed independent variables loaded on two factors, or latent variables in terms of the covariance structure model (see Figure 1). The variables with a semantic component (imageability, context availability, and definition accuracy of the Dutch and English words) loaded substantially on one factor, referred to as *meaning*. The familiarity of the Dutch and English words, which theoretically can be regarded as subjective frequency measures, loaded, together with the (objective) frequency measures of the Dutch and English words, on a second factor, *familiarity*. Finally, cognate status was the only variable that loaded substantially on a separate, third factor. However, since cognate status was the only observed variable underlying this third variable, cognate status will be referred to as an observed variable and not a latent variable. So, in the measurement model the observed (i.e., rated) word characteristics were related to two latent variables, meaning and familiarity, whereas cognate status can be regarded a separate, yet observed variable.

The two latent variables and cognate status were assumed to predict performance on four observed dependent variables: mean RTs for words translated in forward direction by lower and higher proficient subjects, and mean RTs for words translated in backward direction by lower and higher proficient subjects. Note that the mean ratings of word characteristics (the observed independent variables) were provided by groups of subjects different from those providing the mean translation times of these words (the observed dependent variables). Yet the stimuli (i.e., 440 words) were the same for all raters.

Formalizing our model in terms of covariance structure equations, the following system of equations emerges (see Figure 1). The four latent η variables (i.e., RTs of words translated in forward or backward direction by lower or higher proficient subjects) are assumed to exert no causal effect on one another. Therefore, the term $B\eta$ is crossed out in the structural equation, resulting in the following matrix expression:

$$\eta = \Gamma\xi + \zeta.$$

The implied structural equations are:

$$\eta_1 = \gamma_{11}\xi_1 + \gamma_{12}\xi_2 + \gamma_{13}\xi_3 + \zeta_1$$
$$\eta_2 = \gamma_{21}\xi_1 + \gamma_{22}\xi_2 + \gamma_{23}\xi_3 + \zeta_2$$
$$\eta_3 = \gamma_{31}\xi_1 + \gamma_{32}\xi_2 + \gamma_{33}\xi_3 + \zeta_3$$
$$\eta_4 = \gamma_{41}\xi_1 + \gamma_{42}\xi_2 + \gamma_{43}\xi_3 + \zeta_4.$$

where η_1 and η_2 are the RTs for words translated in forward direction by lower and higher proficient bilinguals, respectively, and η_3 and η_4 are the RTs for words translated in backward direction by lower and higher proficient bilinguals, respectively; ξ_1 is meaning; ξ_2 is familiarity; ξ_3 is cognate status; ζ_1, ζ_2, ζ_3, and ζ_4 are errors in equations for forward translation by lower and higher proficient bilinguals and backward translation by lower and higher proficient bilinguals, respectively.

The dependent variables—that is, the RTs of words translated in forward direction by lower proficient (η_1) and higher proficient (η_2) bilinguals and the RTs of words translated in backward direction by lower proficient (η_3) and higher proficient (η_4) bilinguals—are assumed to be directly measured by y_1, y_2, y_3, and y_4, respectively, so $\Lambda_y = \mathrm{I}$ and $\varepsilon = 0$. This results in[3]

using a maze similar to that of Holman. Johnson (1913) studied auditory discrimination in dogs in what was essentially a ϕ-maze, except that reinforcers were delivered in the return arms just outside the startbox. The mazes of Berger (1977) and Holman were automated; the mazes used by Lashley (1912), Hunter (1920), Levine et al., and Johnson required constant experimenter involvement.

Nakagawa (1993) has recently published an experiment in which a commercially available automated T-maze was employed. This automated T-maze, however, appears topologically more similar to a two-lever operant chamber than a T-maze, because the visual discriminative stimuli and the levers are located on the wall opposite the startbox. The apparatus appears to have great potential for visual discrimination studies; however, it requires the rat to press a lever to indicate its choice. This necessitates extensive shaping: Nakagawa indicated that 15 days of leverpress training were required before discrimination training could begin. It is likely that the location of the manipulanda distant from the feeder contributed to the need for extensive training; we have also found that shaping rats to press levers located at some distance from the feeder requires two shaping sessions under the best of circumstances, and sometimes more (Wilson & Cook, 1989). In the ϕ-maze, the only response required by the rat is locomotion, and shaping of the rat's behavior is not necessary.

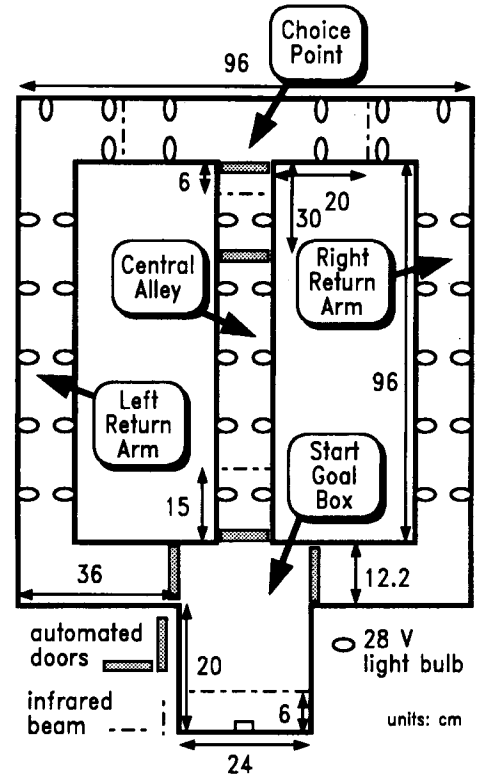

Figure 1. Plan of the ϕ-maze. Walls are black acrylic; floor is translucent white; and ceiling is clear.

THE ϕ-MAZE

Plan of the ϕ-Maze

The ϕ-maze consists of a start/goalbox (s/g box) from which a central alley leads to a choice point. From the choice point, left and right arms return to the s/g box. Figure 1 presents a plan of the maze. The maze is constructed of 0.25-in. (0.6 cm) acrylic plastic: the walls are black, the floor is translucent white, and the top is clear. Automated doors of translucent white acrylic separate the s/g box from the central alley and the two return arms. An automated mechanism in the s/g box can deliver appetitive reinforcers. Five infrared LED–photodarlington pairs located throughout the maze can record the rat's position. Banks of lights located beneath the floor of the central alley and the left and right arms can serve as visual cues. If desired, speakers could be mounted at the choice point or in the left and right arms to provide auditory cues. The total cost of materials for the ϕ-maze, including the parallel port interface for the computer, was less than $700 in the summer of 1991.

Alleyways

A cross section of an alleyway is presented in Figure 2. Each alley is 12.2 cm wide and 19.3 cm from floor to ceiling. Where they occur, the infrared LEDs and photodarlingtons are located 3.8 cm from the floor, on opposite sides of the alley. Beneath the floor of the central alley and the two return arms are 28-V lamps, located as shown in Figures 1 and 2.

Doors

The five guillotine doors are constructed of translucent white acrylic and slide within aluminum channels. The top of each door is attached by a short length of chain to a metal lever. The opposite end of the lever is counterbalanced with a lead weight sufficiently heavy to allow the door to close slowly by the force of gravity. A continuous-duty solenoid (Guardian Electric 4-C-24V) attached to the lever can open the door and hold it open indefinitely. Each door is normally programmed to close when the rat is not near it, but because it is closed by gravity and is carefully counterbalanced, it would not harm the rat if the rat happened to be under it when it closed. The doors open to a height of 6.5 cm in approximately 0.4 sec; they close in approximately 0.6 sec.

Feeder

Quartered Froot Loops (Kellogg's) are used as reinforcers. A Lehigh Valley Equipment pellet dispenser has been modified to deliver them automatically. The aluminum plate of the pellet dispenser has been replaced with an acrylic disk (19 cm in diameter) containing twelve 2-cm holes equally spaced around its perimeter. A 1.5-cm-deep bottomless plastic cup is cemented to each hole. Brief operation of the feeder results in a 30° rotation of the disk, which aligns one of the holes with an opening that allows the contents of the cup to drop into the feeder trough in the s/g box. Because of the increased momentum

$$x_9 = \lambda_{92}\xi_2 + \delta_9$$
$$x_{10} = \lambda_{10\,2}\xi_2 + \delta_{10}$$
$$x_{11} = \xi_3.$$

As was described above, the cardinal difference between the asymmetrical model and the symmetrical model concerns the influence of meaning on translation. First, the asymmetrical model states that meaning predicts the RTs of words translated in forward direction but has no direct influence on backward translation. On the other hand, the symmetrical model states that meaning predicts the RTs of forward and backward translation. Second, the asymmetrical model states that, as subjects become more proficient, the meaning of words will have a growing influence on the RTs of words translated in forward direction, whereas, despite an increased fluency in L2, meaning will (still) not predict backward translation. In contrast, the symmetrical model predicts a stronger effect of the meaning of words on their forward- and backward-translation times as proficiency in L2 is increased.

This covariance structure model (see Figure 1) will be the theoretical model against which the predictions of the asymmetrical model and the symmetrical model will be tested. The models to be compared differed in one respect. In the asymmetrical model, the causal relationships between meaning and RTs for words translated in backward direction by lower proficient (γ_{31}) and higher proficient (γ_{41}) subjects were fixed to zero. In contrast, in the symmetrical model, these restrictions were not imposed. So, the difference between the asymmetrical model and the symmetrical model consists of differences in the parameter sets. More specifically, the asymmetrical model, in which γ_{31} and γ_{41} were fixed to zero, is nested in the symmetrical model, which had no constraints on the gamma parameters.

FITTING FUNCTIONS AND GOODNESS-OF-FIT INDICES

The predictions of the asymmetrical and symmetrical models were tested against the observed data in covariance structure analyses. The parameters of the models to be contrasted were estimated by minimizing the fitting function corresponding to generalized least squares (GLS):

$$F_{GLS} = 0.5 \text{ tr } \{[S - \Sigma(\Theta)]\, S^{-1}\}^2,$$

where S is the sample covariance matrix; $\Sigma(\Theta)$ is the population covariance matrix to be estimated; S^{-1} is the weight matrix by which the differences between S and Σ are being weighted; tr is the trace operator indicating the sum of the diagonal elements of a matrix (for more details, see Bollen, 1989; Jöreskog & Sörbom, 1988; Long, 1983). The estimates obtained by GLS are asymptotically normally distributed. That is, under the assumption of multivariate normality, these estimates approximate a normal distribution as the sample size increases. Moreover, GLS estimates are assumed to be robust, at least, against moderate deviations from non-normality (Bollen, 1989; Jöreskog & Sörbom, 1988; La Du & Tanaka, 1989; Marsh, Balla, & McDonald, 1988; Tanaka, 1987).

The overall fit of the models, as yielded by the fitting function GLS, was assessed by the frequently used goodness-of-fit index (GFI), and the adjusted goodness-of-fit index (AGFI; Jöreskog & Sörbom, 1988; Tanaka & Huba, 1985; for a review of several goodness-of-fit indices see, e.g., La Du & Tanaka, 1989, Marsh et al., 1988, or Tanaka, 1993). AGFI is adjusted for degrees of freedom (*df*s) and it seemed appropriate for our analyses, since the asymmetrical model had 2 *df*s extra due to the fixation of γ_{31} and γ_{41}. The values of GFI and AGFI range from 0 to 1, with 1 indicating a perfect fit. Furthermore, the results of the chi-square tests were used as indicators of the fit of the models (see, e.g., Gerbing & Anderson, 1993), where a low χ^2 relative to the *df*s indicates an adequate fit (all things equal, the magnitude of χ^2 and the number of *df*s decrease when parameters are added to the model).[4] Finally, the adequacy of the model was also assessed by estimates of the parameters, which should have a plausible sign and magnitude. Parameter estimates that are improper, or that deviate far from expectations, indicate that the model may be misspecified (Bollen, 1989).

RESULTS

Measurement Model of the Observed Independent Variables

First, analyses were performed in order to find an optimal measurement model for the independent latent variables, given constraints implied by the theory (see Lomax, 1982).[5] The results of these confirmatory factor analyses are presented in the left part of Figure 2.

The estimates of the lambdas ranged from .63 to .98, indicating moderate to high loadings of the observed independent variables on the latent independent variables. In general, the loadings on the latent variable meaning (ranging from .84 to .98) were somewhat higher than those on the latent variable familiarity (ranging from .63 to .90). The small standard errors of each lambda indicated that these parameter estimates were fairly precise, as was substantiated by the fact that all lambda parameters were highly significant (i.e., different from zero; *t* values ranged from 12.07 to 48.95; all *p*s < .001). Furthermore, the parameter estimates of the measurement errors of the indicators of the latent variable meaning were fairly low (except the .22 and .30 for the context availability and the definition accuracy of the English words), whereas the measurement errors for the indicators of the latent variable familiarity were rather high, ranging from .20 to .61. All delta parameters had small standard errors, indicating that they were precisely estimated, and all appeared to be significant (*t* values ranging from 4.26 to 10.86; all *p*s < .001). Finally, the estimates of the phi parameters were rather low, and only the relationships between the latent independent variables meaning and familiarity (ϕ_{12}), and meaning and cognate status (ϕ_{13})

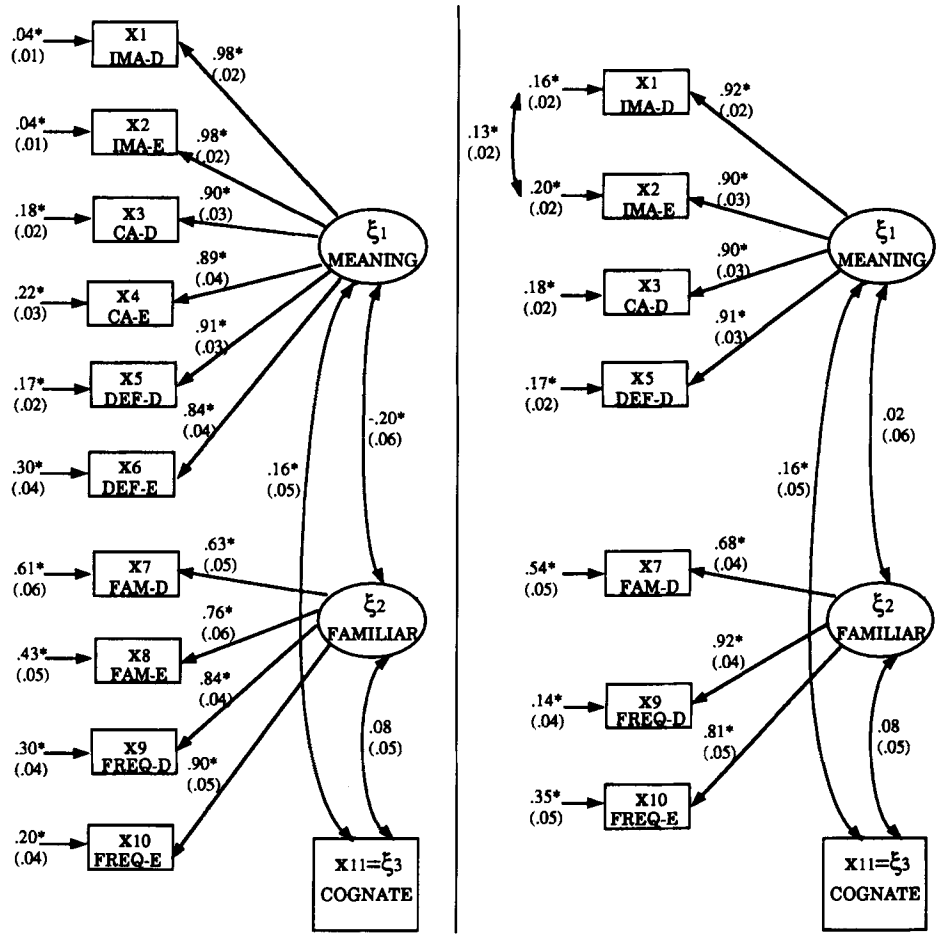

Figure 2. Path diagram with parameter estimates (standard errors) of the measurement models with 11 (left part) and 8 (right part) observed independent variables. IMA = imageability; CA = context availability; DEF = definition accuracy; FAM = familiarity; FREQ = log word frequency; D = Dutch words; E = English words; MEANING = latent word meaning; FAMILIAR = latent word familiarity; COGNATE = cognate status. All parameter estimates are standardized coefficients, and all standard errors are adjusted to the extent of non-normality and standardized. Parameter estimates indicated with * are significantly different from zero, $p < .05$ or better.

differed significantly from zero (ts = 3.48 and 3.38, respectively; both ps < .001). However, the sign of the parameter ϕ_{12} was negative, pointing at a negative relationship between meaning and familiarity. This negative relationship is inconsistent with earlier findings (see de Groot, 1992; de Groot et al., 1994) and might hint at a flawed hypothesis or a misspecification of the model (see Bollen, 1989).

In addition to the parameter estimates, the adequacy of the measurement model was assessed by the goodness-of-fit indices. As can be seen in Table 1, the overall fit of the measurement model, as assessed by GFI, AGFI, and χ^2, was rather low and necessitated a more thorough look at the adequacy of the measurement model.

To determine possible sources of the poor fit, the elements of the measurement model (e.g., parameters) were examined in more detail by inspecting the modification indices and estimates of standardized residuals. Hence, while GFI, AGFI, and χ^2 assess the fit of the general model, modification indices and standardized residuals can be used to examine the congruity of the model for each estimated element (both being zero in the case of a perfect fit).

Modification indices are measures associated with constrained and fixed parameters, whose size indicates the expected decrease in χ^2 if a particular constraint is freed. It appeared that the modification indices for the loadings of context availability and definition accuracy of the En-

Table 1
Fit Indices of the Measurement Models With 11 and 8 Observed Independent Variables, and the Symmetrical and Asymmetrical Models

	Measurement Model		Symmetrical	Asymmetrical
	11 Variables	8 Variables	Model	Model
GFI	.82	.96	.94	.92
AGFI	.72	.91	.86	.85
χ^2 (df)	383.67 (42)	79.32 (17)	151.84 (37)	180.21 (39)

Note—GFI = goodness of fit index; AGFI = adjusted goodness of fit index; df = degrees of freedom.

glish words on familiarity were 2.93 and 11.56, respectively. Since these connections were not assumed to exist, as can be seen in Figure 1, these lambdas were constrained to zero in the hypothesized measurement model. However, the modification indices of these parameters suggested that the observed data would be more accurately described by a measurement model with no constraints on these two lambda estimates—in particular, that of the definition accuracy of the English words on the latent variable familiarity.

The second measure, the standardized residuals, indicates the departure of the population (co)variances (as estimated by the model) from the observed sample (co)variances. Standardized residuals (i.e., fitted residuals divided by their asymptotic standard errors) are specified for each element of the model and are standardized for scaling differences of the observed variables. It appeared that the standardized residuals for the variances of the context availability, definition accuracy, and familiarity of the English words (9.36, 10.15, and 11.78, respectively) and the covariances between these variables (ranging from 9.46 to 11.47) were very high and differed significantly from zero, indicating that the model does not account for these (co)variances very well. This was substantiated by considerable correlations between the observed independent variables context availability and familiarity of the English words ($r = .55$) and between definition accuracy and familiarity of the English words ($r = .58$). In the model, however, both context availability and definition accuracy of the English words were assumed to be indicators of the latent variable meaning, and familiarity of the English words was assumed to be an indicator of the latent variable familiarity. Theoretically, a plausible explanation is that subjects, in assessing the context availability and the definition accuracy of words in their non-native language, contaminated the familiarity of a word with the specific word characteristic to be rated (see also de Groot et al., 1994). Therefore, it might be that the rated context availability and definition accuracy of the English words were not fully accurate measures of the latent variable meaning, and it was expected that the fit of the measurement model might improve by removing the observed variables context availability, definition accuracy, and familiarity of the English words (see also Alwin & Tessler, 1985).[6]

Finally, the modification index for the correlation between δ_1 and δ_2 (Θ^{δ}_{12}, errors in measurement of the imageability of the Dutch words and the English translations) was 41.89, suggesting that freeing this parameter might improve the fit of the measurement model (in the measurement model, all correlations between the δs were constrained to 0; see Figure 1). Theoretically, it can be understood that the errors in measurement (or, stated differently, the measure-specific components) of the imageability of the Dutch words and their English translations have something in common, since both words of each translation pair were rated on this same word characteristic.[7] Therefore, it seemed legitimate to free Θ^{δ}_{12}.

The aforementioned assertions were tested in a new confirmatory factor analysis with eight independent observed variables, meanwhile freeing Θ^{δ}_{12}. The parameter estimates are presented in the right-hand part of Figure 2, and the fit indices are presented in Table 1. As can be seen, the fit of the measurement model as assessed by GFI, AGFI, and χ^2 improved substantially. Also, the covariance between meaning and familiarity decreased, whereas the magnitude of the λ estimates remained, overall, unaffected.

Structural Equation Model

Having established a measurement model that described the observed data fairly accurately, the next step was to test the structural hypotheses of the symmetrical and asymmetrical models. Therefore, the measurement model with eight observed independent variables (identical for all comparisons to be made) was concatenated to the structural equation model. Two covariance structure analyses were performed.[8] In the first analysis, no constraints were imposed on the gamma parameters, thereby modeling the symmetrical model. In the second, the relationships between the latent independent variable meaning and the RTs of words translated in backward direction by lower proficient (γ_{31}) and higher proficient (γ_{41}) subjects were fixed to zero, conforming to the predictions of the asymmetrical model. The results of these analyses are presented in Figures 3 and 4.[9]

First, the estimates of the direct effects of the latent independent variable meaning on the observed dependent variables (i.e., the gamma parameters) will be discussed. Recall that the predictions of the asymmetrical model and those of the symmetrical model differed in two respects: the role of meaning in backward translation and the influence of proficiency in L2 on translation. For these comparisons, we will look at the data obtained in the analysis with no restrictions imposed on the gamma parameters, as presented in Figure 3.

The asymmetrical model predicts that the meaning of words will not influence the RTs of words translated in backward direction either by lower proficient (γ_{31}) or higher proficient (γ_{41}) subjects. In contrast, the model predicts that meaning will influence the RTs of forward translation by lower proficient (γ_{11}) and higher proficient (γ_{21}) bilinguals. Hence, the asymmetrical model predicts that γ_{31} and γ_{41} will be zero, whereas γ_{11} and γ_{21} will be different from zero. The symmetrical model, on the other hand, predicts that meaning does not differentially affect the latencies of words translated in forward direction (γ_{11} and γ_{21}) or in backward direction (γ_{31} and γ_{41}), by both lower and higher proficient subjects. So, the symmetrical model predicts that γ_{11}, γ_{21}, γ_{31}, and γ_{41} will all be different from zero. It appeared that (see Figure 3), as predicted by both models, γ_{11} and γ_{21} deviated significantly from zero ($ts = 10.02$ and 10.61, respectively; both $ps < .001$). However, consistent with the predictions of the symmetrical model, but contrary to those of the asymmetrical model, γ_{31} and γ_{41} were significantly different from

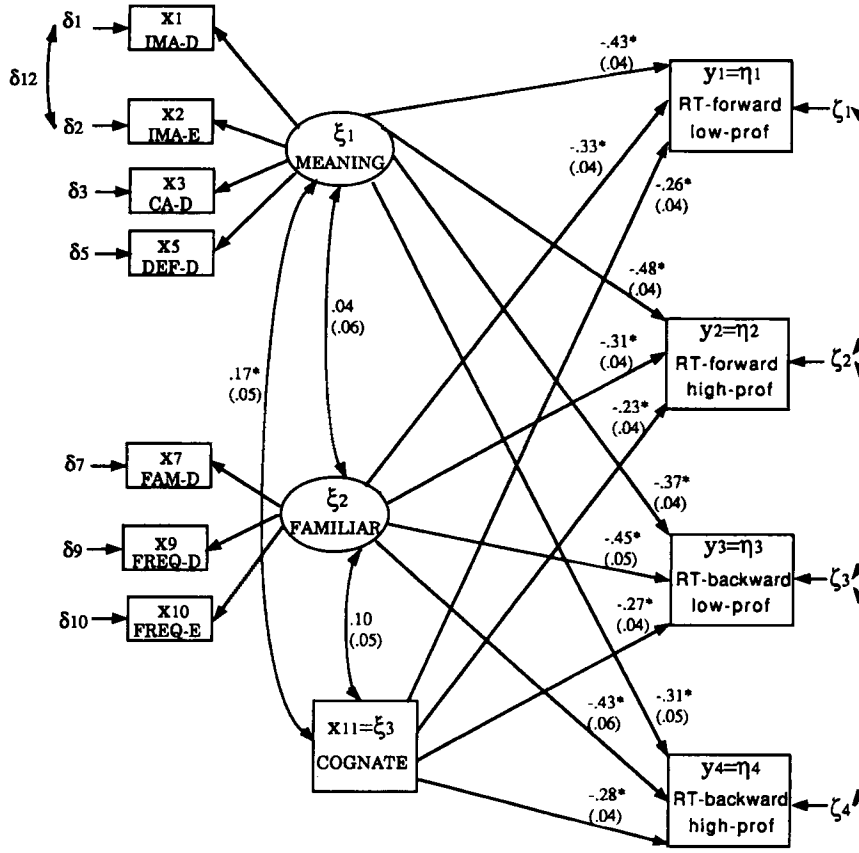

Figure 3. Path diagram with parameter estimates (standard errors) of the symmetrical model. IMA = imageability; CA = context availability; DEF = definition accuracy; FAM = familiarity; FREQ = log word frequency; D = Dutch words; E = English words; MEANING = latent word meaning; FAMILIAR = latent word familiarity; COGNATE = cognate status; RT-forward = reaction times of words translated in forward direction; RT-backward = reaction times of words translated in backward direction; low-prof = lower proficient bilinguals; high-prof = higher proficient bilinguals. All parameter estimates are standardized coefficients, and all standard errors are adjusted to the extent of non-normality and standardized. Parameter estimates indicated with * are significantly different from zero, $p < .05$ or better.

zero (ts = 8.04 and 6.37, respectively; both ps < .001). This implies that the RTs of words translated in forward direction and in backward direction, by either lower or higher proficient subjects, were predicted by their meaning. These findings are in accordance with the predictions of the symmetrical model, yet contrary to those of the asymmetrical model.

However, a closer look at the gamma estimates hints at a weak asymmetry, because it seems that meaning plays a larger role in forward translation than in backward translation (see Figure 3). In order to compare the size of these parameters (γ_{11} with γ_{31}, and γ_{21} with γ_{41}), confidence intervals were calculated for each parameter. If one parameter value falls outside the confidence interval of the parameter that it is being compared with, it can be concluded that two parameter estimates are significantly different. However, if a parameter falls within the confidence interval of a second parameter, their values are not significantly different (see, e.g., Hays, 1963). The limits of the 95% confidence intervals were calculated by subtracting from (lower limit) and adding to (upper limit) the unstandardized parameter estimate the product of the Z value associated with the 95% confidence level (i.e., 1.96) and the unstandardized standard errors. So, for example, the unstandardized estimate for γ_{11} was -166.86, and its confidence interval was $-166.86 \pm 1.96 \times 16.65$, yielding the interval -199.49 to -134.23. Similarly, the unstandardized estimates of γ_{21}, γ_{31}, and γ_{41} were -157.91, -108.91, and -96.54, respectively. The confidence intervals ranged from -187.07 to -128.75 for γ_{21}, from -135.47 to -82.34 for γ_{31}, and from -126.24 to -66.83 for γ_{41}. Hence, the estimate of γ_{11} fell outside the confidence interval of γ_{31}, and vice versa. Similarly, the estimate of γ_{21} fell outside the confidence interval of γ_{41}, and vice versa. So, it can be concluded that the RTs of words translated in forward direction were affected more by meaning than were those of words translated in backward direction, both for lower proficient (γ_{11} and γ_{31}) and

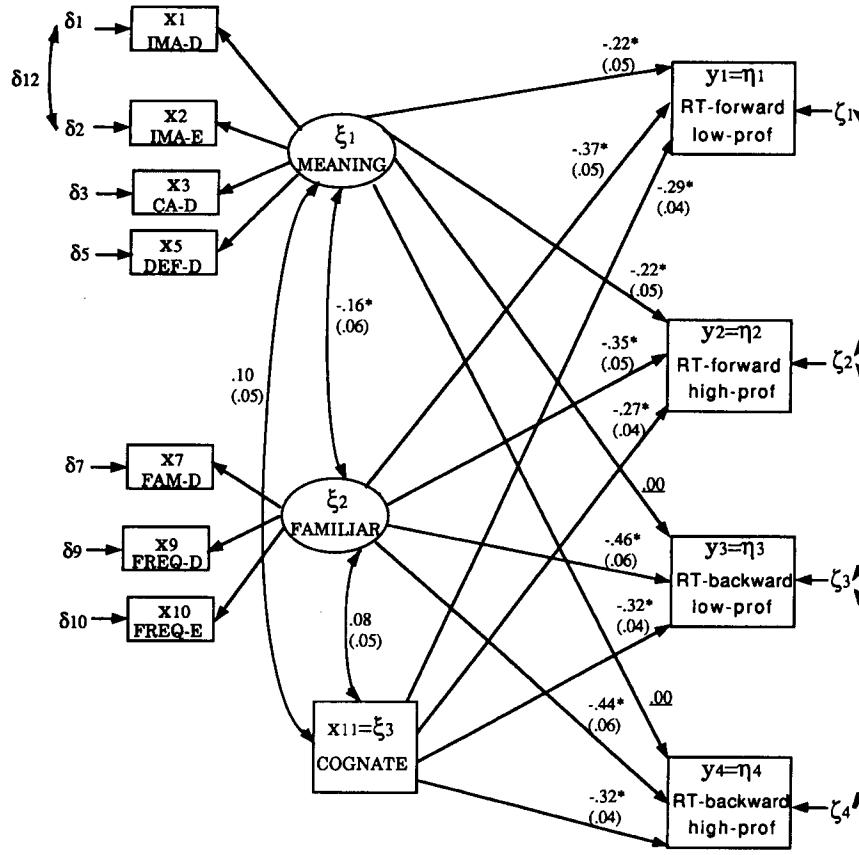

Figure 4. Path diagram with parameter estimates (standard errors) of the asymmetrical model. IMA = imageability; CA = context availability; DEF = definition accuracy; FAM = familiarity; FREQ = log word frequency; D = Dutch words; E = English words; MEANING = latent word meaning; FAMILIAR = latent word familiarity; COGNATE = cognate status; RT-forward = reaction times of words translated in forward direction; RT-backward = reaction times of words translated in backward direction; low-prof = lower proficient bilinguals; high-prof = higher proficient bilinguals. All parameter estimates are standardized coefficients, and all standard errors are adjusted to the extent of non-normality and standardized. Parameter estimates indicated with * are significantly different from zero, $p < .05$ or better. The underlined parameters have been constrained to equal the given value.

higher proficient (γ_{21} and γ_{41}) subjects. An alternative method for comparing parameter estimates is by adding equality constraints to these parameters and examining the effect on χ^2. To see whether the differences in χ^2 were significant, chi-square difference tests (or likelihood ratio tests) that are applicable to nested models were performed (Bollen, 1989). It appeared that equating γ_{11} with γ_{31} yielded a chi-square test statistic of 3.88 ($p < .05$, 1 *df*), indicating that relaxing the equality constraint on γ_{11} and γ_{31} will lead to a significant improvement in general fit. Similarly, equating γ_{21} with γ_{41} increased χ^2 with 5.54 ($p < .02$, 1 *df*), implying that relaxing the equality constraint on γ_{21} and γ_{41} will result in a better fit. Hence, the results of the chi-square difference tests converged on those of the confidence interval tests.

Second, for higher fluency in L2, the asymmetrical model predicts a higher influence of meaning on forward translation, but not on backward translation. In contrast, the symmetrical model predicts that, as proficiency in L2 grows, both forward translation and backward translation will be influenced increasingly by meaning. In terms of the covariance structure model, the asymmetrical model predicts that $\gamma_{11} < \gamma_{21}$, and $\gamma_{31} = \gamma_{41}$. The symmetrical model, however, agrees that $\gamma_{11} < \gamma_{21}$, but predicts that $\gamma_{31} < \gamma_{41}$. A comparison of the size of these estimated parameters employing the confidence intervals calculated above showed that the estimation of γ_{11} did not differ from that of γ_{21}, nor did γ_{31} differ from γ_{41}. On a statistical level, this indicates that, for both proficiency levels, meaning equally influenced the RTs of words translated in forward direction and in backward direction. However, this conclusion should be considered with some caution, because it might be the case that the difference in L2 proficiency between the two groups providing the translation times was not sufficiently large to induce different processing strategies, in either forward translation or backward translation.

Besides this detailed look at the fit of the component parts of the models, their general fit will be compared. To this end, the fit indices of the analysis without restric-

tions on the gamma parameters, modeling the symmetrical model, were compared with the fit indices yielded by the analysis in which γ_{31} and γ_{41} were fixed to zero, thereby modeling the predictions of the asymmetrical model. As can be seen in Table 1, GFI and AGFI of the symmetrical model were somewhat higher than those of the asymmetrical model. This difference in overall fit was substantiated by the lower χ^2 of the symmetrical model relative to that of the asymmetrical model. The concomitant chi-square difference test statistic was 28.37 with 2 *df*s ($p < .001$), indicating that relaxing the constraints on γ_{31} and γ_{41} (getting the symmetrical model) will lead to a significant improvement of fit.

In sum, the variances and covariances predicted by the structural hypotheses of the symmetrical model more closely resembled the observed (co)variances than did those estimated by the asymmetrical model. It appeared that the RTs of words translated in forward direction and in backward direction, regardless of the proficiency level of the bilinguals providing the translations, were both influenced by the meaning of the words being translated. So, contrary to the predictions of the asymmetrical model, this suggests that conceptual memory is involved in both forward translation and backward translation. However, the gamma estimates suggested that meaning plays a larger role in forward translation than in backward translation. In other words, a weak version of the asymmetrical model is supported by the data.

DISCUSSION

As is clear from the exposition above, the application of covariance structure analysis compels scrutiny of theoretical premises. Implicit operationalizations of theoretical constructs (e.g., the word characteristic "imageability" is a semantic variable) must be made explicit by the researcher, because they must be modeled in the covariance structure: The researcher has to define explicitly how the latent variables are measured in terms of the observed variables and has to specify the causal relationships among the latent constructs. The adequacy of the so-constructed theoretical model is tested against the observed data in the covariance structure analysis: The more adequate the theoretical model, the better the fit. As such, the technique of covariance structure modeling can be useful in the development and subsequent refinement of theoretical models. Moreover, we think that covariance structure analysis can be a valuable addition to more traditional techniques for data analysis frequently used in experimental research (e.g., ANOVA or multiple regression analysis). We will illustrate this by comparing information obtained by "traditional" data analysis (i.e., de Groot, 1992; de Groot et al., 1994) with that obtained by the covariance structure analyses reported here.

First, in ANOVA or multiple regression analysis, no distinction is made between an observed variable and a latent, theoretical construct: It is assumed that the observed (or manipulated) variable is isomorphic with the unobserved latent variable. However, in many cases, the assumed perfect equivalence between the observed variable and the unobserved latent variable will not reflect reality (Alwin & Tessler, 1985). In covariance structure analysis, the relationship between observed and latent variables is explicitly modeled, yielding a direct test of the equivalence assumption. This point can be illustrated by comparing the original study of de Groot et al. (1994) with the covariance structure analyses reported here. De Groot et al., using a correlational design, assumed that the imageability, context availability, and definition accuracy of Dutch words and their English translations reflected meaning, and they examined the correlations between these word characteristics (observed independent variables) and forward- and backward-translation latencies (observed dependent variables). Using the data of this study, in covariance structure analyses, we found (see Figure 2) that the loadings of these six observed variables on the latent variable meaning ranged from .84 to .98. So, though these factor loadings were fairly high, the observed variables were not isomorphic to the latent variable that they were assumed to reflect.

A second, related point is that ANOVA and multiple regression analysis presuppose no errors in the measurement of the observed variables. However, in most experiments, measurement errors will be the rule rather than the exception, conceivably influencing the effects of the manipulation of independent variables or the power of the analysis (Alwin & Tessler, 1985; Blalock, 1985; Bollen, 1989). By way of illustration, de Groot (1992), using an ANOVA design, orthogonally manipulated the imageability and the log frequency of Dutch words (observed independent variables) and tested the influence of this manipulation on forward-translation latencies (observed dependent variables). In the confirmatory factor analyses reported in this paper, however, the measurement errors of the observed imageability and log frequency of Dutch words were .04 and .30, respectively (see left-hand part of Figure 2). Although the imageability and frequency ratings adapted from de Groot et al. (1994) and used in this paper were different from those used by de Groot (1992), the acuteness of the manipulation or the power of the analysis could, theoretically, have been influenced or confounded by the measurement errors. In covariance structure analysis, however, two components are distinguished in the variances of the observed variables: a unique variance (or measurement error) and a common variance, related to the latent underlying variable.

Finally, the aforementioned merits of covariance structure analysis are related to a third, more theoretical, point. Usually, experiments are designed to test causal relationships between variables or constructs in a theory, and not between observed measures. Yet, these observed measures are being modeled in ANOVA or multiple regression analysis. Though, as was discussed before, the a priori assumptions that observed measures and theoretical constructs are isomorphic, or that observed variables are measured without error, may be faulty. In co-

variance structure analysis, however, structural equations are modeled using the error-free latent variables. Stated differently, in covariance structure analysis, the causal relationships between theoretical constructs are being modeled, operationalized by measures of these constructs.

The analyses reported in this paper were intended to serve as demonstrations of the application of covariance structure modeling on data obtained in experimental research. Of course, covariance structure analysis has a wider scope than can be demonstrated within the limits of one single experiment, and we will glance at some other applications of this technique below. The covariance structure analyses reported in this paper were based on our data from earlier research, in which we used a correlational design and, among others, performed multiple regression analyses. The measurement scale of our independent variables was ordinal (except for the Dutch and English word frequencies). However, in experiments, particularly in ANOVA designs, the independent variables are often orthogonally manipulated and categorized in mutually exclusive, nominal classes. Still, the data of these designs can be modeled using covariance structure analysis, as has been elegantly demonstrated by Alwin and Tessler (1985) and Bagozzi (1977). These authors emphasized explicit modeling of the effectiveness of the manipulation of independent variables (i.e., by constructing manipulation checks). Furthermore, the analyses reported in this paper were all "one-group" analyses. However, in experimental research, the focus of interest is often on differences between groups (e.g., an experimental and a control group), and data are obtained from different samples, necessitating a multigroup analysis. Such a simultaneous analysis of different samples can be performed using structural equation modeling, with the restriction that independent random samples are available. Bollen (1989), Jöreskog and Sörbom (1988), and Lomax (1983) provide detailed explanations and guidelines on how to perform multisample structural equation modeling.

We consider covariance structure modeling a useful technique for researchers within the experimental cognitive field. A practical hindrance in using covariance structure analysis might be that this technique demands a fairly large number of cases (Bollen, 1989; La Du & Tanaka, 1989; Marsh et al., 1988; Tanaka, 1987). In our analyses, we were not hindered by this constraint: Ensuing from our research topic, words were the unit of analysis, and the sample size of the words was large enough. Although, as yet, there is no agreement on the exact number of cases needed for covariance structure analysis, a sample size of 75–100 cases seems to be a minimum. Hopefully, the constraint imposed by sample size can be circumvented because of recent developments in bootstrapping techniques (see Bollen & Stine, 1993), in which relatively small samples can be analyzed reliably.

In conclusion, we think that the technique of covariance structure modeling is a valuable addition to the more traditional techniques for data analysis used in experimental research, and it can be a valuable tool for the development and refinement of theories and for contrasting alternative (or competing) models.

REFERENCES

Alwin, D. F., & Tessler, R. C. (1985). Causal models, unobserved variables, and experimental data. In H. M. Blalock (Ed.), *Causal models in panel and experimental designs* (pp. 55-88). New York: Aldine.

Bagozzi, R. P. (1977). Structural equation models in experimental research. *Journal of Marketing Research*, **14**, 209-226.

Bentler, P. M. (1980). Multivariate analysis with latent variables: Causal modeling. *Annual Review of Psychology*, **31**, 419-456.

Bentler, P. M. (1989). *EQS structural equations program manual*. Los Angeles: BMDP Statistical Software.

Blalock, H. M. (1985). *Causal models in panel and experimental designs*. New York: Aldine.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.

Bollen, K. A., & Stine, R. A. (1993). Bootstrapping goodness-of-fit measures in structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 111-135). Newbury Park, CA: Sage.

Brown, R. L. (1986). A comparison of the LISREL and EQS programs for obtaining parameter estimates in confirmatory factor analysis studies. *Behavior Research Methods, Instruments, & Computers*, **18**, 382-388.

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

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

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

de Groot, A. M. B., & Hoeks, J. C. J. (1995). The development of bilingual memory: Evidence from word translation by trilinguals. *Language Learning*, **45**, 683-724.

Gerbing, D. W., & Anderson, J. C. (1993). Monte Carlo evaluations of goodness-of-fit indices for structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 40-65). Newbury Park, CA: Sage.

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

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

Jöreskog, K. G. (1993). Testing structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 294-316). Newbury Park, CA: Sage.

Jöreskog, K. G., & Sörbom, D. (1988). *Lisrel 7: A guide to the program and applications* [Computer program manual]. Chicago: SPSS.

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 & Verbal Behavior*, **23**, 519-539.

Kroll, J. F. (1993). Accessing conceptual representation for words in a second language. In R. Schreuder & B. Weltens (Eds.) *The bilingual lexicon* (pp. 53-81). Amsterdam/Philadelphia: John Benjamins.

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

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

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

La Du, T. J., & Tanaka, J. S. (1989). Influence of sample size, estimation method, and model specification on goodness-of-fit assessments

in structural equation models. *Journal of Applied Psychology*, **74**, 625-635.

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

Lomax, R. G. (1982). A guide to LISREL-type structural equation modeling. *Behavior Research Methods & Instrumentation*, **14**, 1-8.

Lomax, R. G. (1983). A guide to multiple-sampling structural equation modeling. *Behavior Research Methods & Instrumentation*, **15**, 580-584.

Long, J. S. (1983). *Covariance structure models: An introduction to LISREL*. Beverly Hills, CA: Sage.

Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin*, **103**, 391-410.

Royce, J. R. (1977). Guest editorial: Have we lost sight of the original vision for SMEP and MBR? *Multivariate Behavioral Research*, **12**, 135-141.

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

Tanaka, J. S. (1987). "How big is big enough?": Sample size and goodness-of-fit in structural equation models with latent variables. *Child Development*, **58**, 134-146.

Tanaka, J. S. (1993). Multifaceted conceptions of fit in structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 10-39). Newbury Park, CA: Sage.

Tanaka, J. S., & Huba, G. J. (1985). A fit index for covariance structure models under arbitrary GLS estimation. *British Journal of Mathematical & Statistical Psychology*, **38**, 197-201.

NOTES

1. Several computer programs are available for performing covariance structure analysis. Frequently used programs are LISREL (Jöreskog and Sörbom, 1988) and EQS (Bentler, 1989). See Brown (1986) for a comparison of applying the LISREL and EQS programs in confirmatory factor analysis.

2. The qualifications "low" and "high" should be considered relative terms; the difference in L2 fluency between our two proficiency groups was rather small, and certainly substantially smaller than those of other studies aimed at studying the influence of proficiency on translation (see, e.g. de Groot & Hoeks, 1995; Kroll & Sholl, 1992). In fact, relative to those of many other studies, our low-proficient subjects can be regarded as quite fluent in English. For these reasons, we will henceforth refer to our subject groups as *lower* proficient and *higher* proficient groups.

3. A more elaborate measurement model for the dependent variables could have had four latent variables: performance on words translated in forward direction by lower proficient and higher proficient bilinguals, and performance on words translated in backward direction by lower proficient and higher proficient bilinguals. Each of these four latent dependent variables had been measured by three observed variables: RTs, errors, and omissions (de Groot et al., 1994). However, the distribution of the errors and omissions was skewed. Furthermore, we thought that this complex measurement model would not be helpful in clarifying the essence of covariance structure analysis. Therefore, we decided to use the directly measured y_1, y_2, y_3, and y_4 as the dependent variables.

4. Since the distribution of some observed variables deviated somewhat from normality, we used the Satorra–Bentler scaled χ^2 (see Bollen, 1989), as suggested by one of the reviewers of this article. The Satorra–Bentler scaled χ^2 is available in EQS (Bentler, 1989), though not in LISREL (Jöreskog & Sörbom, 1988).

5. The analyses of the measurement model were based on 438 words. Because of missing values, 2 words of the original corpus were "listwise deleted."

6. Other solutions could be proposed, all based on the original model described in Figure 1. Model A: addition of a path between context availability of the English word and latent familiarity, and one between definition accuracy of the English word and latent familiarity. Model B: addition of a path between familiarity of the English word and the latent variable meaning. Model C: combination of Models A and B. Model D: Addition of a fourth latent variable on which context availability, definition accuracy, and familiarity of the English words load (meanwhile deleting the original relationships between these observed and latent variables as described in Figure 1). However, none of these alternative measurement models leads to a satisfactory fit. More details of these analyses can be requested from the first author. In addition, in Model E, we freed the correlations between measurement errors of Dutch and English versions of the same word characteristic, hence δ_1 with δ_2, δ_3 with δ_4, δ_5 with δ_6, δ_7 with δ_8, and δ_9 with δ_{10}. In Model F, we correlated all errors related to the Dutch and all errors related to the English word characteristics, hence δ_1, δ_3, δ_5, δ_7, and δ_9, as well as δ_2, δ_4, δ_6, δ_8, and δ_{10}. However, both Model E and Model F were not identified, implying that the parameters of these models can not be uniquely determined. Hence, estimation of these models is arbitrary and useless, since it is possible to find an infinite number of values for the parameters, with each set being consistent with the covariance equation (see Bollen, 1989; Long, 1983).

7. Freeing Θ^{δ}_{12} and, hence, allowing δ_1 and δ_2 to correlate, in fact introduces a new and, in this case, not explicitly defined latent variable. However, we considered it justified to free Θ^{δ}_{12} because the measure-specific components of the imageability of the Dutch words and their English translations have something in common: In essence, the same word has been rated twice (once in Dutch and once in English) using the same instruction. For this reason, we thought it legitimate to improve the accuracy of the measurement model by freeing Θ^{δ}_{12}. Incidentally, the modification indices for $\Theta^{\delta}_{\text{ca-d/ca-e}}$ (Θ^{δ}_{34}), $\Theta^{\delta}_{\text{def-d/def-e}}$ (Θ^{δ}_{56}), and $\Theta^{\delta}_{\text{fam-d/fam-e}}$ (Θ^{δ}_{78}) were 29.34, 54.72, and 15.33, respectively. However, since the observed context availability, definition accuracy, and familiarity of the English words were removed in the final measurement model, freeing these deltas was unnecessary.

8. The analyses of the structural equation model were based on 436 words. Because of missing values, 4 words of the original corpus were "listwise deleted."

9. For easy reference, the estimates of the lambda and delta parameters are not presented in Figures 3 and 4, because these estimates were comparable to those presented in the right-hand part of Figure 2.

(Manuscript received July 5, 1994;
revision accepted for publication September 28, 1995.)