THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY
2006, 59 (6), 977–986

Psychology Press
Taylor & Francis Group

Short article

What do we know about psycholinguistic effects?

Michael B. Lewis and Matei Vladeanu
Cardiff University, Cardiff, UK

Identifying clear and unequivocal psycholinguistic effects for lexical retrieval tasks has been the aim of a significant proportion of recent research activity. Debates have erupted concerning the existence or otherwise of particular effects on particular lexical tasks. Here, it is suggested that the reason for these debates is that researchers exercise choice in what variables they consider in their analysis. It is further illustrated that methods that have been employed for comparing the size of these effects between tasks can only lead to inconclusive results. It is suggested that psycholinguistic data may be better analysed using structural equation modelling methodologies. An example of such an approach is presented.

What determines how fast we can read words or name pictures? Attempts to answer this question have been influential in the development of cognitive psychology. Indeed, there is a vast range of articles, both historically and recently, concerning the relative effects that a variety of word properties exert on various lexical tasks. It has been stated that this type of research has been central to the development of cognitive psychology. It is difficult to disagree with this argument given the number of pages of cognitive-psychology journals given over to this topic. While this research has generated a great deal of data, here we try to establish whether we really know anything about psycholinguistic effects based on these data.

In this paper, we illustrate some of the difficulties regarding the analysis of psycholinguistic data relating to lexical access. We look at the difficulties encountered when choosing which psycholinguistics variables to consider and suggest that much disagreement among researchers stems from these choices. We look at the difficulties associated with comparing the size of effects across different tasks. A method that has been employed to do this is investigated as to whether it can reveal the information asked from it. Finally, we offer a tentative solution to these problems by way of a structural equation modelling reanalysis of a large database of psycholinguistic data. This offers some hope for the future but also illustrates potential difficulties.

The importance of choosing variables

Psycholinguists have a vast range of independent variables to choose between when designing or

Correspondence should be addressed to Michael B. Lewis, School of Psychology, Cardiff University, Park Place, Cardiff CF10 3AT, UK. Email: LewisMB@cf.ac.uk

This research was funded by a Leverhulme Trust grant. We thank Todd Bailey, David Balota, Dennis Norris, Melvin Yap, and other anonymous reviewers for their comments on earlier versions.

http://www.psypress.com/qjep
DOI:10.1080/17470210600638076

analysing their studies. These variables need to be either controlled for or considered as manipulated variables. While some variables such as word frequency and length are almost ubiquitously employed, others such as age of acquisition or neighbourhood size are employed more selectively. The range of potential predictors of psycholinguistics outcome variables appears to be growing all the time. Indeed, Cutler (1981) suggested that with the rate of increase in variables it would become impossible to conduct psycholinguistic research in the future. Psycholinguistic research did continue, however, primarily by researchers selecting a particular limited number of variables and choosing to ignore others.

One can demonstrate that the choice of including a variable or not within an analysis can influence the results obtained and, more importantly, the conclusions based upon those results. Shibahara and Kondo (2002) presented a reanalysis of Yamazaki, Ellis, Morrison, and Lambon Ralph (1997) data on a Kanji naming-latency task. They demonstrated that the effect that Yamazaki and colleagues attributed to age of acquisition disappears if another variable, familiarity, is included in the analysis. In this case, the same data set was used to support different points of view showing that psycholinguistic effects established in this way remain open to differing interpretation.

Even when there is near universal agreement that a particular factor should be included within analysis, the researcher is left with a choice of which measure best represents that factor. Take, for example, word frequency. There are many measures of word frequency to choose between. What one would like to measure when using word frequency is the number of times a particular word is encountered by a particular person. We cannot know this so we have to settle for an estimate of the number of times that a word is encountered by an average person. We can obtain estimates of this property by counting word frequencies in particular corpuses. The formation of these corpuses will influence the accuracy of the frequency measure as an estimate of actual frequency. If, for example, many of the items in the corpus come from children's literature (as in the case of the Zeno, Ivens, Millard, & Duvvuri, 1995, measure) then earlier learnt words will be overrepresented. If the items in the corpus come from Internet newsgroups (as is the case for the Lund & Burgess, 1996, measure) then words common among Web users may be overrepresented. The measure of frequency that one uses can influence the outcome of a multiple regression analysis.

Rather than using the typical solution of arbitrarily selecting a word-frequency norm, a solution employed by Balota and colleagues (Balota, Cortese, Sergent-Marshall, Sieler, & Yap, 2004) was to identify which word frequency measure best predicted reaction times (i.e., the Zeno norms) and to use only that measure in the regression analysis. A potential problem with this "best predictor" method for selecting a word frequency measure is that it is biased in favour of any measure of frequency that also incorporates some other factor that influences reaction times. As the Zeno norms gathered data from a range of age-appropriate sources, the measure contains within it an element of this age effect with earlier learnt words showing a slightly higher frequency on this measure than they would on a norm based on only adult sources. As a result, the regression that includes the Zeno norms as a measure of objective frequency could be observing age of acquisition effects in addition to any effects of frequency to which the observed effects are attributed. In this way, effects attributed to frequency may contain another factor such as age of acquisition.

One further problem with choosing variables is the fact that few variables are exact measures. Consequently, it is possible to end up with the situation where one can have two independent measures that are in fact the same thing. To demonstrate: A multiple regression reanalysis of Balota and colleagues' (2004) lexical decision tasks as predicted by the Zeno frequency norms and the Lund and Burgess norms produces two independent frequency effects. In fact the same factor is probably producing both of these effects but it illustrates the point that multiple regression results can be open to different interpretations. It is possible that other, so-called, independent predictors are actually produced by the same underlying effect.

The lack of consensus for effects in psycholinguistic research stems from the fact that we cannot manipulate the factors that we call independent variables. A word has high frequency, not because we have manipulated its frequency, but because there is something that causes it to be produced more often than other words. When we identify word frequency effects, these effects are merely correlations between two dependent measures. We are no closer to knowing whether it is the thing that is causing a word to be produced more often that is also producing the effect or if it is word frequency itself. Zevin and Seidenberg (2002) highlighted this problem of cause and effect with regard to age of acquisition effects but the same is true for all psycholinguistic variables.

Comparison of effect sizes across tasks

Manipulations are required to infer causative effects. Psycholinguistic variables themselves cannot readily be manipulated but the tasks employed can be. For example, a study may employ an auditory and a visual lexical decision task (Turner, Valentine, & Ellis, 1998) with reaction times being faster for words presented visually than for auditory words. One can observe, therefore, real effects of task. Psycholinguists, however, do not tend to discuss effects such as the auditory-versus-visual presentation effect for lexical decision tasks. That is, we are more typically interested in the effect that a manipulation makes on the size of the psycholinguistic variables' effects. This is illustrated by the manipulation made by Balota and colleagues (2004). We focus upon this study because we believe that it is likely to become a seminal paper the methodology of which will be repeated if unchallenged.

Balota and colleagues (2004) made the task manipulation of whether a lexical decision task or a word-naming task was performed. They considered how the task affected the size of the psycholinguistic variables' effects by comparing two multiple regression designs. Indeed, the great range of different variables that showed differential effects across the two tasks is used to infer "different constellations of processes engaged by the two tasks" (p. 298). The implicit, and untested, assumption from the conclusions being drawn is that the analysis will reveal different standardized regression coefficients between tasks for only those factors that have different degrees of influence between those two tasks. This assumption, we posit, is incorrect (i.e., a variable can have equivalent effects on two tasks yet reveal different standardized regression coefficients). In order to demonstrate this, we describe an analysis similar to that performed by Balota and colleagues (2004) but performed on a "toy" dataset generated with known causal relationships.

The dataset was generated as follows for a set of 10,000 hypothetical words. All variables were generated from independent random normally distributed data or simple scalar combinations of other variables. Three hypothetical psycholinguistic properties of these words were generated, called meaningfulness, frequency, and connectivity (these are just labels for the sake of argument). Connectivity and meaningfulness were generated as random variables. Frequency was generated by adding a random variable to the connectivity value, allowing the relative degrees of intercorrelation between the predictors to be different. Two variables were generated that refer to reaction times for two different tasks. The first task was generated such that it was predicted by frequency and meaningfulness. This was generated by adding a random variable to the sum of the frequency and meaningfulness values. The second task was generated such that it was predicted by frequency to the same extent as the first task but it was also further affected by connectivity. This was generated by taking a random variable and adding the frequency variable and also the connectivity variable. Figure 1 represents the pattern of associations within the data.

Two multiple regressions were conducted in a manner equivalent to that performed by Balota and colleagues (2004) on their psycholinguistic

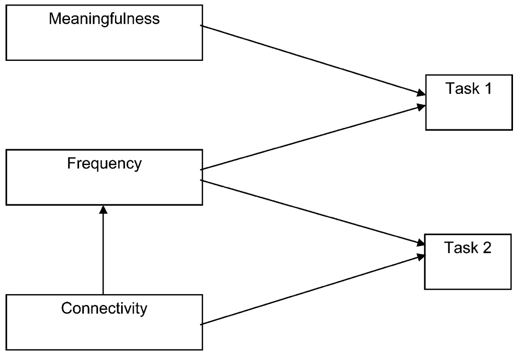

Figure 1. *Construction of the "toy" dataset. Connectivity and frequency are correlated with each other but meaningfulness is not correlated with either of the other two predictor variables. Meaningfulness and frequency affect Task 1 whereas connectivity and frequency affect Task 2. Note that the size of the effect of frequency is the same for Task 1 and for Task 2.*

data. Each regression had one of the tasks' reaction times as the dependent variable and three independent variables of meaningfulness, frequency, and connectivity. For the first task, reaction time was significantly "affected" by meaningfulness and frequency but not by connectivity. The standardized regression coefficients were .494 for meaningfulness, .718 for frequency, and .007 for connectivity. The second task showed significant effects for connectivity and frequency but not for meaningfulness, with standardized regression coefficients of .412 for connectivity, .573 for frequency, and .003 for meaningfulness. These figures reveal a large difference in the size of the coefficient for the frequency variable across the two tasks. Further, the zero-order or simple correlation between Task 2 and frequency was larger (r = .864) than between Task 1 and frequency (r = .706).

The results demonstrate the different size of the frequency effects (as measured by standardized regression coefficients) between the two tasks. The data set was generated such that the size of the effect of each variable should be the same across the two tasks. The results of the analysis, however, show that the standardized coefficient for frequency is considerably larger for Task 1 than for Task 2. Further, this difference would be significant if one employed the kind of by-participant analysis employed by Balota and colleagues (2004) on their psycholinguistic data.

This simulation demonstrates that a differential in the size of the mean standardized regression coefficients between two tasks does not necessarily mean that the two tasks are differentially affected by that property. In the current example, frequency equally affects the two tasks yet shows different standardized coefficients between the two tasks when analysed using the methods that have been employed by psycholinguists. The standardized frequency coefficients are different for the two tasks because the size of the intercorrelation between frequency and the predictors that do differentially affect the two tasks is different. This demonstrates that different standardized regression coefficients can be produced even when the predictor has the same effect on two different tasks. As a consequence, care must be taken when inferring that an effect is different on two different tasks as standard multiple regression is not able to evaluate this.

Structural equation modelling analysis of psycholinguistic data

Structural equation modelling allows for a set of relationships between one or more independent variables and one or more dependent variables to be examined. The resulting models can be discussed in terms of causality, but the causality as such is theory driven rather than an emergent property of the analysis. Also, several models can be compared in terms of goodness-of-fit (i.e., how well the data fit the hypothesized relationships between variables). Such techniques are widely employed in the fields of social psychology where one wishes to identify the independent impact of factors that one cannot directly manipulate. One feature of structural equation modelling is the use of latent variables (i.e., theoretical factors that cannot be measured directly, e.g., depression). If a person has depression then this will be expressed in a number of directly measurable ways (e.g., appetite loss, self loathing, or tiredness). When studying depression, we study these measurable variables in order to find out about the latent variable of depression. In psycholinguistics we are interested in word representations as affected by word frequency, imageability, and other factors. We cannot measure directly the strength of a word representation (it is latent) but we can use reaction times in particular tasks as indicators of it. There are many reasons for considering latent variables within an analysis that are particularly appropriate within psycholinguistic research (see Skrondal & Rabe-Hesketh, 2004). One reason is to reflect a true variable but one that has measurement error, which is often the case as psycholinguistics norms always contain error. A second reason is to reflect a hypothetical construct that cannot be directly measured (Bailey & Hahn, 2005, have employed latent variables of this nature in the domain of psycholinguistics). A third reason for using latent variables is to combine information about items from different sources.

While there are similarities between structural equation modelling and multiple regression (e.g., they are based on analysis of intercorrelations) there are fundamental differences. Multiple regression provides simple definitive results about which independent variables produce significant effects on which dependent variables. In structural equation modelling, the analysis requires a theoretical model that can be tested against the available data. It can be determined whether this model is consistent with the data and also how good a fit it is.

The following analysis represents an initial analysis of psycholinguistic data using structural equation modelling. It takes as its data set the data collected and made available by Balota and colleagues (2004).

The model that was tested had eight latent variables. The first latent variable was "voice key delay", which was a hypothetical construct that had as indicators the phonemic properties of the words likely to affect the relationship between the onset of speech and the timing of the voice key response. The second latent variable was "imageability", which was supposed to reveal how simple or complex a concept was such that "bread" would be simple whereas "creed" would be complex. This latent variable reflected a true variable with measurement error. As such, indicators of this latent variable were measures of imageablity. The third latent variable was "frequency", with indicators of four measures of frequency, which again was a true variable with measurement error. The use of this latent variable meant that the model was not restricted to a single measure but was able to model most appropriately all available frequency data with regard to the frequency norms and the rest of the data in the model. The fourth latent variable was "size", which combined information about letter length and phoneme length and hence had these measures as its indicators. The fifth and sixth latent variables were "onset consistency" and "rime consistency", which again combined information and had indicators of the appropriate feedforward and feedback measures. The seventh latent variable was the "representation strength" for the word, which is a hypothetical construct that is meant to be how easy a word is to access or to recognize as being a word. This latent variable had indicators of the reaction times for the lexical decision times. The final latent variable was "word production", which again is a hypothetical construct reflecting ease of word access and had as indicators the reaction times for the word-naming tasks.

The latent variables were connected with the following theoretically justifiable causal links. Voice key delay was connected such that it had a causal influence on word production. This is justifiable because words that have sounds that take longer to produce a response on a voice key will produce longer word production times. Representation strength was connected such that it had a causal effect on word production. This is theoretically justified because words that are most susceptible to brain damage are also those that are read more slowly.[1] Size was connected to word production because it has been suggested that we have to build the whole word before we start to pronounce

[1] We thank Sachiko Kinoshita for suggesting this justification.

it. Onset and rime consistency were connected to word production because words that have fewer competitors for pronunciation could be produced faster. Imageability and frequency were linked to representation strength. The former link is because it has been argued that the imageability of a word affects how quickly it is processed (e.g., Morrison & Ellis, 2000; Shibahara, Zorzi, Hill, Wydell, & Butterworth, 2003), and the latter link is because the frequency of a word is likely to improve the strength of the representation through some form of incremental learning.

The model, as described above (and shown in Figure 2a), is consistent with the findings reported by Balota and colleagues from their regression analysis. The constellations of processes are reflected in the differing connections to representation strength and word production. Structural equation modelling evaluation of this model shows that it is a borderline fit to the data. That is, the model is not inconsistent with the data collected: $\chi^2(145) = 650.16$, $p < .001$; comparative fit index (CFI) = .939; root mean square error of approximation (RMSEA) = .066 (N.B., a CFI > .900

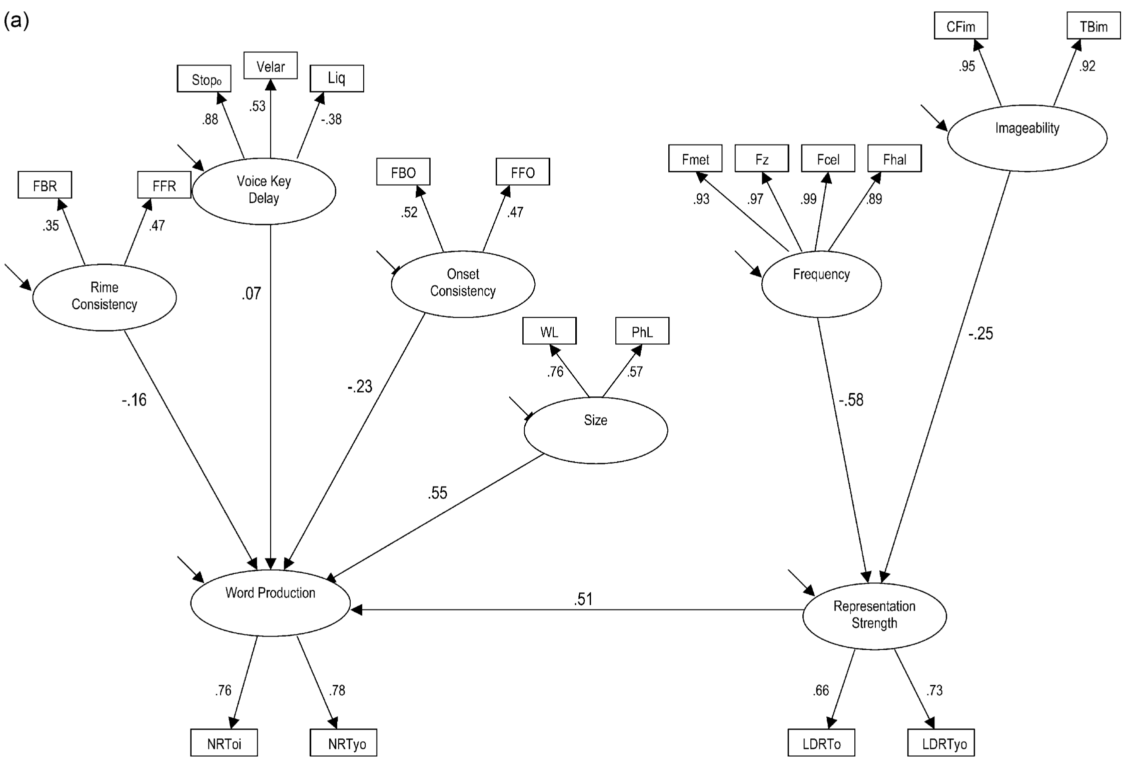

Figure 2a (above), and Figures 2b and 2c (facing pages). *Three different versions of the structural equation modelling analysis of some of the data presented in the Balota and colleagues (Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004) dataset. Measured variables are in rectangles whereas latent variables are in ovals. Connections represent the pattern of effects being investigated in each model. Stop, Velar, and Liq(Liquid glide) are three phonemic properties. FBR = feedback rime consistency; FFR = feedforward rime consistency; FBO = feedback onset consistency; FFO = feedforward onset consistency; WL = word length; PhL = phonological length; Fmet = the MetaMetrics frequency norms (MetaMetrics word frequency counts database, MetaMetrics Inc., 2003, as cited in Balota et al., 2004). Fz = the Zeno, Ivens, Millard, and Duvvuri (1995) frequency norms; Fcel = the Celex frequency norms (Baayen, Piepenbrock, & van Rijn, 1993); Fhal = the HAL frequency norms (Lund & Burgess, 1996); TBim = Toglia and Battig (1978) imeagability narms; CFim = Cortese and Fugett (2003) imageability norms; NRTo/y = Balota et al.'s (2004) naming reaction times for old/young participants; and LDRTo/y = Balota et al.'s (2004) lexical decision reaction times for old/young participants.*

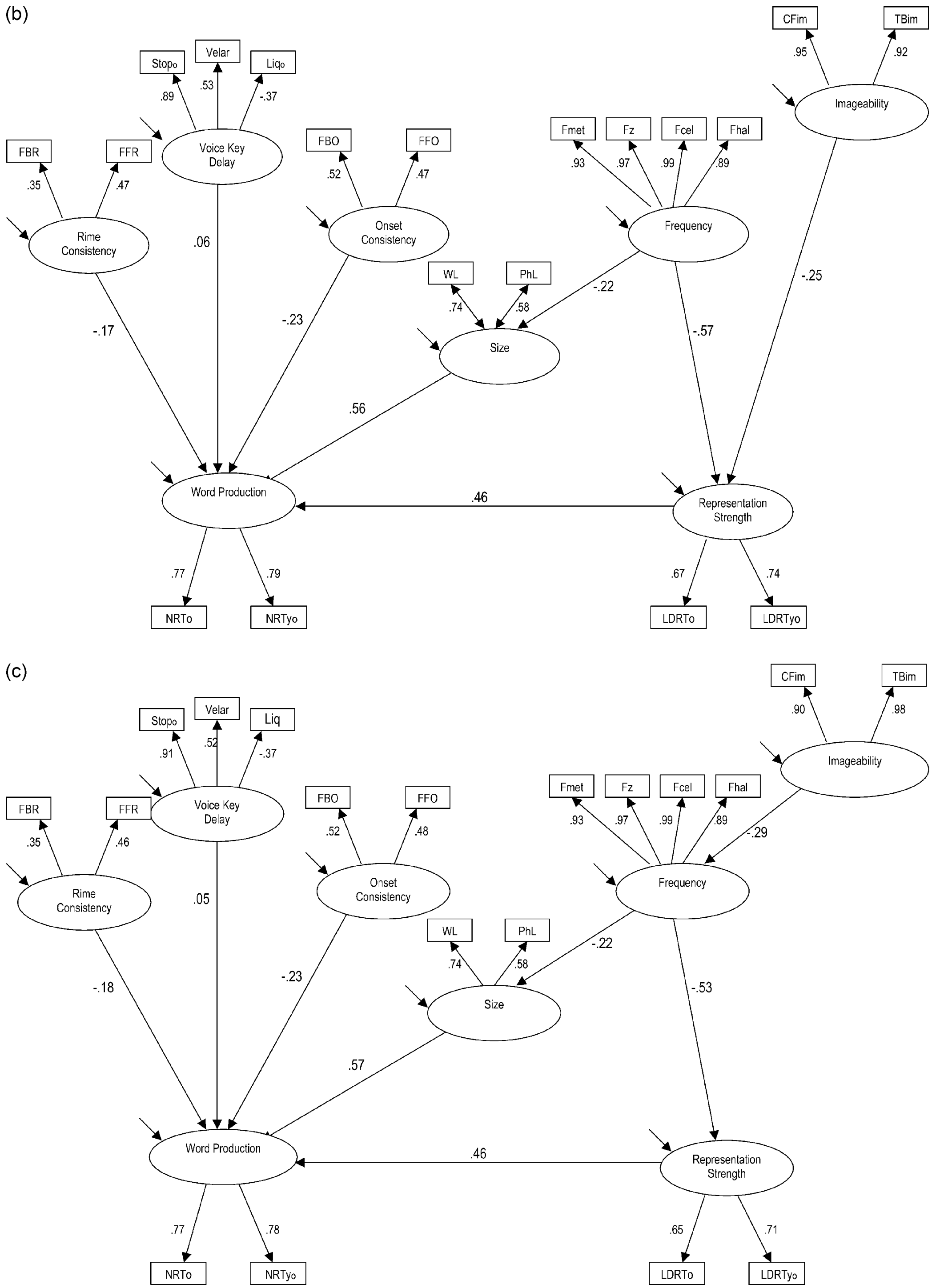

Figure 2 (Continued).

represents a good fit, Bentler, 1992). Figure 2a also shows the coefficients of the connections between the variables included in the model. These show strong effects of frequency and imageability on representation strength. It also shows strong effects of size and consistency on word production. Representation strength also had a large effect on word production.

While this model is consistent with the conclusions drawn by Balota and colleagues (2004) and also with the data, there are alternative models that can be tested and evaluated. One model to consider is one that includes a causative link between frequency and size such that higher frequency causes words to be shorter. It may seem strange to suggest such a causative effect in psycholinguistic terms because the length of a word can be measured very accurately, and it is difficult to see a cognitive explanation for such an effect. The model being presented, however, is not just a model of cognitive psychological effects but of all causative effects between correlated psycholinguistic variables. The language we use has been generated over many years, and there are reasons why words have the particular form they do. One of these reasons may be the frequency with which a concept is used. It is easy to think of examples of concepts that have developed shorter names in the last century as the concept has become more common (e.g., "automobile" became "car"; "telephone" became "phone"). It is possible that the same pressure to use shorter terms for more commonplace items leads to a language structure where frequency arguably affects the size of a word (either in letters or phonemes). The second model presented (Figure 2b) is the same as the first model but includes a causal link between word frequency and size. Analysis of this model shows that it too is consistent with the dataset $\chi^2(144) = 624.82$, $p < .001$; CFI $= .942$; RMSEA $= .064$. Further, this model is a significantly better fit to the data than is the previous model, $\Delta\chi^2(1) = 25.34$, $p < .001$.

One further adjustment that we would like to suggest is a change in the nature of the causal influence of imageability in the tasks. The models so far have a direct causal link between imageability and representation strength. This link was placed there because of arguments put forward in the literature regarding the effects of imageability on lexical processing (e.g., Shibahara et al., 2003). This hypothetical effect was based on small-scale studies of the kind that can produce illusory effects (see Lewis, 2006). An alternative model is suggested where imageability has a causal effect on frequency only and no direct causal effect on representation strength. This is justified by the speculation that the words that refer to more basic concepts are those that we are going to use more often. A simple concept like "bread" will occur in natural language more often than a complex concept like "creed". In this way, it is the simplicity of the concept that affects how often a word is used. The deleting of the causal link between imageability and representation strength is justified because it is difficult to determine a direct causal reason why a simple concept should have a stronger representation except for the fact that there may be a mediating effect of word frequency. This model is represented in Figure 2c. The model is consistent with the data: $\chi^2(144) = 600.09$, $p < .001$; CFI $= .945$; RMSEA $= .063$. Further, this model is a significantly better fit to the data than is either of the previous two models: $\Delta\chi^2(1) = 24.73$, $p < .001$. The conclusion that can be drawn is that the effects that imageability has on lexical processing that have previously been observed are indirect correlations and as such not real effects. Making a concept more imageable will not, in itself, make a lexical decision task faster unless the frequency is also increased.

While this last model is consistent with the data and better than the previous models, there is nothing to say that a better model is not possible. The models presented and tested are based on a combination of theoretical justification and existing research findings. As demonstrated by the last model, previous research that does not take into consideration indirect causation can be possibly misleading and, we argue, should not be used as justifications for links of causation in structural equation modelling analysis.

The models presented above have all been described as consistent with the data. This might lead one to believe that almost any model is likely to be consistent but there are at least some models that were found to be inconsistent. Subjective frequency was included in the models described above either as an indicator of frequency (as this is what raters were asked to estimate) or as an indicator of representation strength (as this is what raters might have actually used in providing their ratings). The former models led to great increases in the chi-squared term indicating a poor fit whereas the latter models could not be resolved due to negative covariance matrices (indicative of either a poor fit or a high degree of multicolinearity, Gerbing & Anderson, 1987). This is more problematic for a simple interpretation of what subjective frequency is actually measuring than for the models reported here.

Indeed, negative covariance matrices provide even greater problems for the more general application of this technique to psycholinguistic data. There are many additional links that psycholinguists may like to see within the models that are not currently there. At least some of these links, however, increase the degree of colinearity and lead to models with negative covariances. Consequently, these variations on the models cannot be evaluated. It may ultimately be the case that the correlations between psycholinguistic variables are too strong even for structural equation modelling to be able to effectively test or isolate distinct psycholinguistic effects.

CONCLUSIONS

Traditional analyses of psycholinguistic data lead to findings of effects that are open to differing interpretation. It is suggested that structural equation modelling analysis may offer a solution. The reason for conducting the structural equation modelling of the data reported above was not to present a definitive account of causal patterns between psycholinguistic variables. Rather, we wished to demonstrate the way that this method of data analysis can be brought to bear on the data that have been collected to test whether the conclusions drawn are, first, consistent with the data and, second, the only or best way to explain the data. The conclusions that can be drawn from this analysis are, on the whole, similar to the conclusions that Balota and colleagues (2004) drew.

In the data analysed here, it is shown that the effect of imageability, widely reported in small-scale studies and further supported by Balota and colleagues' (2004) multiple regression design, could be an illusory effect brought about by a causal relationship between imageability and word frequency. No matter how well controlled a small-scale study may be, it would not be capable of identifying whether an effect of imageability is illusory or not. Similarly, large-scale multiple regression designs cannot confirm such effects because of the inaccuracies in measurement. Structural equation modelling, however, may offer an effective insight into whether theoretically justified effects are real or illusory although even this methodology has its limitations regarding uncovering causality.

Original manuscript received 7 December 2005
Accepted revision received 13 February 2006

REFERENCES

Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1993). *The CELEX lexical database* [CD-ROM]. Philadelphia: University of Pennsylvania, Linguistic Data Consortium.

Bailey, T. M., & Hahn, U. (2005). Phoneme similarity and confusability. *Journal of Memory and Language*, *52*, 339–362.

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General*, *133*, 283–316.

Bentler, P. M. (1992). On the fit of models to covariances and methodology to the Bulletin. *Psychological Bulletin*, *112*, 400–404.

Cortese, M. J., & Fugett, A. (2003). Imageability ratings for 3,000 monosyllabic words. *Behaviour Research Methods, Instruments & Computers*, *36*, 384–387.

Cutler, A. (1981). Making up materials is a confounded nuisance, or: Will we be able to run any psycholinguistic experiments at all in 1990? *Cognition*, *10*, 65–70.

Gerbing, D. W., & Anderson, J. C. (1987). Improper solutions in the analysis of covariance structures: Their interpretability and a comparison of alternate respecifications. *Psychometrika*, *52*, 99–111.

Lewis, M. B. (2006). Chasing psycholinguistic effects: A cautionary tale. *Visual Cognition*, *13*(7/8), 1012–1026.

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavioral Research Methods, Instruments & Computers*, *28*, 203–208.

Morrison, C. M., & Ellis, A. W. (2000). Real age of acquisition effects in lexical retrieval. *British Journal of Psychology, 91*, 167–180.

Shibahara, N., & Kondo, T. (2002). Variables affecting naming latency for Japanese Kanji: A re-analysis of Yamazaki et al. (1997). *Perceptual and Motor Skills*, *95*, 741–745.

Shibahara, N., Zorzi, M., Hill, M. P., Wydell, T., & Butterworth, B. (2003). Semantic effects in word naming: Evidence from English and Japanese Kanji. *Quarterly Journal of Experimental Psychology*, *56A*, 263–286.

Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: Multilevel, longitudinal and structural equation models*. London: Chapman & Hall/CRC.

Toglia, M. P., & Battig, W. F. (1978). *Handbook of semantic word norms*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Turner, J. E., Valentine, T., & Ellis, A. W. (1998). Contrasting effects of age of acquisition and word frequency on auditory and visual lexical decision. *Memory & Cognition, 26*, 1282–1291.

Yamazaki, M., Ellis, A. W., Morrison, C. M., & Lambon Ralph, M. A. (1997). Two age of acquisition effects in the reading of Japanese Kanji. *British Journal of Psychology*, *88*, 407–411.

Zeno, S. M., Ivens, S. H., Millard, R. T., & Duvvuri, R. (1995). *The educator's word frequency guide*. Brewster, NY: Touchstone Applied Science.

Zevin, J. D., & Seidenberg, M. S. (2002). Age of acquisition effects in word reading and other tasks. *Journal of Memory and Language*, *47*, 1–29.