

Another look at the relationship between phonetic symbolism and the feeling of knowing

ASHER KORAIAT

Hebrew University, Jerusalem, Israel

Koriat (1975), using a word-matching phonetic symbolism task, found a significant relationship between translation accuracy and the degree of subjective confidence associated with the response. This study examined the hypothesis that since responses on which subjects are in consensus are more likely correct than incorrect, confidence ratings may actually be related to the degree of consensus, regardless of the correctness of the response. The results strongly supported this hypothesis. For items with a consensually correct response, translation accuracy increased with degree of subjective confidence, while for items with a consensually incorrect response it decreased. Consensual responses were judged subjectively more compelling than nonconsensual responses regardless of their accuracy. Some implications of these findings are outlined.

One experimental approach to the study of phonetic symbolism is the word-matching technique originated by Tsuru and Fries (1933), which has enjoyed widespread use since. In its antonymic form (e.g., Slobin, 1968), this technique involves having English speakers match antonyms in a foreign noncognate language with their English equivalents. Significant success in matching is interpreted to suggest that sounds have intrinsic symbolic connotations that are shared alike by all humans, and that these have left a detectable residue in all natural languages.

Word-matching data, however, can also be analyzed in terms of intersubject agreement in matchings regardless of their accuracy. A significant consensus may indicate that a common conception of sound-meaning linkages is shared by the subjects. As far as the intersubject consensus is concerned, data thus far available indicate that English speakers tend to agree fairly well on the symbolic connotations of phonemes or nonsense syllables. The remarkable degree of intersubject consensus in most word-matching studies is, therefore, not particularly surprising, so that the major issue remains that of the correctness of the consensual translation. Consensus among English speakers is more readily interpretable in terms of a culturally acquired phonetic symbolism derived from experience with a common native language. Even Taylor and Taylor (1965), who strongly reject the idea of a universal phonetic symbolism, regard the phenomenon of phonetic symbolism in the sense implied by the intersubject

agreement as an undeniably pervasive feature of language behavior. However, accuracy, it demonstrated, points to the possibility of cross-cultural generality.

In most word-matching studies, indices of consensus are clearly stronger than indices of accuracy. This means that a subject's responses, on the average, agree better with the responses of other subjects than with the correct response. In a word-matching item with two response alternatives, consensus is demonstrated if the distribution of subjects' responses over the two alternatives departs from randomness in either direction, whereas accuracy is demonstrated by a departure favoring the correct response. This allows the classification of the items in a given study into three types: nonconsensual (NO items, for which none of the response alternatives is favored over the other; consensually correct (CC) items, in which subjects significantly favor the correct match; and consensually wrong (CW) items, in which it is the incorrect match which is consensually endorsed. The test of the possibility of a universal phonetic symbolism rests on the proportion of CC to CW items. The CW items may be said to represent a clash between the conception of sound-meaning linkages shared by the English-speaking subjects and that presumably embodied in some of the words of the foreign language. For these items, the correct match is felt by the observers to be "unreasonable" or "counterintuitive."

Recently (Koriat, 1975), a new parameter, also found to correlate with accuracy, was added to the study of phonetic symbolism. This parameter involves the subjective confidence with which the response in a word-matching task is endorsed. In Koriat's study, subjects were asked to match English antonyms with antonyms for three noncognate languages, and to indicate on a 4-point scale the degree of subjective confidence, or feeling of knowing (FOK), for each

This research was supported by grants from the Faculty of Social Sciences and the Human Development Center, Hebrew University of Jerusalem. I wish to thank Barbara Ziff and Karen Pardo for their assistance. Reprint requests should be addressed to Asher Koriat, Department of Psychology, Hebrew University, Jerusalem, Israel.

match. Although translation accuracy was in general better than chance, it was found to vary considerably with subjects' FOK ratings. Thus, across individuals and items, a match assigned the lowest FOK rating had a likelihood of 53% of being correct, compared to 6b% for a match assigned the highest FOK rating. These findings were interpreted to indicate that subjects can monitor their success in guessing the meaning of foreign words, and were taken to suggest that taking into account the degree of subjective confidence associated with a response might lead to stronger support for the idea of a universal phonetic symbolism.

The purpose of the present study is to examine an alternative interpretation of these findings, namely, that they reflect the correlation of subjective confidence with the consensuality of a response, rather than with its accuracy. In word-matching studies yielding a better than chance correct translation, accuracy and consensuality are correlated in the sense that consensual translations are more often correct than wrong. This was true of Koriat's study as well, where consensually translated items were right three times as often as they were wrong. Evidently this consensuality-accuracy relationship is the basic datum on which the hypothesis of a universal phonetic symbolism rests. However, as far as confidence ratings are concerned, this same relationship may lead to the finding of an overall positive correlation between confidence and accuracy, although the real relationship is between confidence and consensuality. The understanding of the nature of confidence ratings in a word-matching task requires the creation of an artificial situation where the effects, of accuracy and consensuality are not confounded. This is possible with CW items where the consensual response is incorrect. We shall contrast two hypotheses regarding the nature of FOK ratings in a phonetic symbolism word-matching task: first, the "accuracy" hypothesis which states that in his FOK ratings the subject monitors the success of guessing the correct response; and second, the "consensuality" hypothesis, which states that in his FOK ratings the subject monitors the success of guessing the consensual response (i.e., the response favored by other subjects) regardless of its being right or wrong.

The general procedure involves the presentation of word-matching items which vary widely in terms of likelihood of eliciting the correct response, so that a large enough representation of items from each of the categories CC, NC, and CW is likely to exist. From the "accuracy" hypothesis, it follows that for each of the three sets of items a positive relationship should obtain between accuracy and FOK ratings. This means that (a) the stronger the subjective confidence associated with a translation, the better the likelihood that the response is correct, and (b) correct

translations will elicit stronger FOK ratings than incorrect responses. From the "consensuality" hypothesis, it follows that only for CC items should the confidence-accuracy relationship be positive; for NC items, no relationship should obtain; and for CW items, the relationship should be negative.

METHOD

Stimulus Materials

A list of 85 items was used, each item consisting of a pair of English antonyms and their correspondent translations in Czech (14 pairs). Hindi (13), Chinese (13), Thai (18), Kanarese (17), or Yoruba (10). The list was compiled from lists employed by Brown, Black, and Horowitz (1955), Slobin (1968), and Weiss (1963). The results obtained in these studies, as well as those obtained by Koriat (1975) were examined and compared in order to obtain for each item a rough estimate of the percentage of correct translation (PCT) likely to result in a visual administration of that item to a new group of subjects. Such an estimate appeared possible since the data available suggest that the percentage of correct translation associated with an item tends to be quite stable over different samples of English speakers. Antonyms were selected solely on the basis of these estimates, with the following considerations in mind: first, that the final list include items that be evenly distributed over a wide range of estimated PCT; and second, that the list include a large enough set of items for which the incorrect translation would be likely to be the consensual one. Since, in all the studies consulted, translations were more often correct than wrong, it was possible to secure more items with estimated PCT above 50% than items with estimated PCT below 50%. The final 85-item list therefore included proportionately more items of the former type.

A visual presentation of the list was employed. All the material was compiled in a booklet containing the instructions and the list of antonym pairs. Each English pair was matched by its corresponding foreign pair, printed in Roman characters, in the following format:

beautiful	chou
ugly	mei

Words within a pair were randomly arranged, and the order of the 85 items in the list was determined on a random basis without regard to the foreign language involved.

Subjects and Procedure

The procedure was similar to that used in the previous study (Koriat, 1975). Written instructions directed the subject to indicate the appropriate match for each of the items, and to rate on a 4-point scale the subjective confidence associated with each match. The subject was instructed to mark "1" when he felt his answer to be "based on a totally wild guess" and "4" when he felt the answer "reasonably likely to be right." and to try to make use of each of the four ratings at least a few times.

One hundred students enrolled at the Hebrew University participated in this study. They were all native speakers of English who disclaimed any knowledge of the six foreign languages. The experiment was conducted in small groups.

RESULTS

Over all 85 items employed, the mean percentage of correct translation was 56.55, which is significantly better than chance. Evidently, in view of the item-selection procedure used in this study, this effect

can only be taken to indicate that "correct" items are overrepresented in the list compiled. Of the 85 items, 52 were translated correctly by over 50% of the subjects and 33 were translated correctly by less than 50% (no items had a PCT of 50%). The range of PCTs was remarkable as exemplified by the Chinese word pair *ching* (meaning light) and *chung* (meaning heavy) (PCT 94°7«) and the Hindi word pair *ranjida* (meaning sad) and *khush* (meaning happy) (PCT18%).

In order for a pair translation to be different from chance at the .05 level, it must have been made by at least 60% of the subjects. This criterion was used to divide the 85 items into three types: consensually correct (CO items, with PCTs of 60% or more, consensually wrong (CW) items, with PCTs of 40% or less, and nonconsensual (NC) items. The CC type included 38 items with a mean PCT of 74.42 ($a = 8.73$), the NC type included 28 items with a mean PCT of 50.07 ($a = 5.60$), and the CW type included 19 items with a mean PCT of 30.37 ($o = 5.45$).

Likelihood of Correct Translation as a Function of Confidence

We shall first examine the relationship between likelihood of correct translation and degree of subjective confidence. The most convenient summary of the data pertaining to this relationship appears in Table 1. For each subject, each of the pair translations was classified in terms of both the accuracy of the translation (correct or incorrect), and the FOK rating associated with it. The contingency frequencies were then pooled across all individuals and all items within a given set. Table 1 presents the pooled frequencies for each of the items separately, and for all items combined.

Two features of the results presented in Table 1 are immediately apparent. First, when the data for the full list are considered, a positive relationship obtains

Table 1
Frequencies of Correct and Incorrect Translations for Responses With Different FOK Ratings: for CC, NC, and CW Items and for All Items Combined

FOK Rating	CC Items				NC Items			
	1	2	3	4	1	2	3	4
Incorrect	253	306	304	109	284	464	423	227
Correct	441	793	967	627	315	422	441	224
Total	694	1099	1271	736	599	886	864	451
Percentage Correct	63.5	72.2	76.1	85.2	52.6	47.6	51.0	49.7

FOK Rating	CW Items				All Items			
	1	2	3	4	1	2	3	4
Incorrect	254	411	431	227	791	1181	1158	563
Correct	154	191	154	78	910	1406	1562	929
Total	408	602	585	305	1701	2587	2720	1492
Percentage Correct	37.7	31.7	26.3	25.6	53.5	54.3	57.4	62.3

Table 2
Mean Individual Translation Accuracy Scores for Responses Assigned Different FOK Ratings: for CC, NC, and CW Items, and for All Items Combine

FOK Rating	CC Items		NC Items		CW Items		All Items	
	Mean	N	Mean	N	Mean	N	Mean	N
1	61.15	94	49.99	97	34.09	94	50.95	99
2	73.43	100	47.77	99	32.61	99	54.35	100
3	76.32	97	51.31	96	25.45	96	57.64	97
4	85.01	93	47.12	87	23.62	80	67.72	95

between subjective confidence and likelihood of correct translation. This finding replicates that of our previous study. Second, however, examination of the data obtained for each set of items separately reveals that the confidence-accuracy relationship differs from one set to another. For CC items, likelihood of correct translation increases monotonically with degree of subjective confidence, for CW items it decreases, and for NC items no apparent relationship emerges. Thus, it appears at this point that the positive confidence-accuracy relationship demonstrated in the previous study (and in the present study, with all items combined) is probably due to the overrepresentation of consensually correct items in the sample of items employed. The data of Table 1 do not lend themselves to a chi-square analysis in view of the dependencies involved. When all responses are combined, the total effects probably represent a mixture of contributions from both individuals and items. Thus, individuals displayed a great and consistent variability in the mean confidence ratings employed, probably reflecting differences in personal standards. To evaluate appropriately the significance of the accuracy-confidence relationships, four translation accuracy scores were calculated for each individual, consisting of the percentages of correct translations for responses that he assigned FOK ratings of 1, 2, 3, and 4, respectively. This was done for each of the three sets of items separately, and for all items combined. For some individuals, none of the items in a given set was assigned a certain FOK rating, and therefore the respective translation accuracy scores could not be computed for these individuals. The means and Ns of the individual accuracy scores are presented in Table 2. It can be seen that the trends are similar to those observed in the data of Table 1. For CC items, translation accuracy increases with degree of subjective confidence, for CW items it decreases, and for NC items no systematic change is evident. One-way analyses of variance for repeated measures were carried out for each of the three sets of items, using for each those subjects for whom all four translation accuracy scores were available. The results were as follows: for CC items, $F(3.258) = 32.61, p < .001$; for NC items, $F(3.246) = .65, n.s.$; and for CW items, $F(3.216) = 1.97, n.s.$

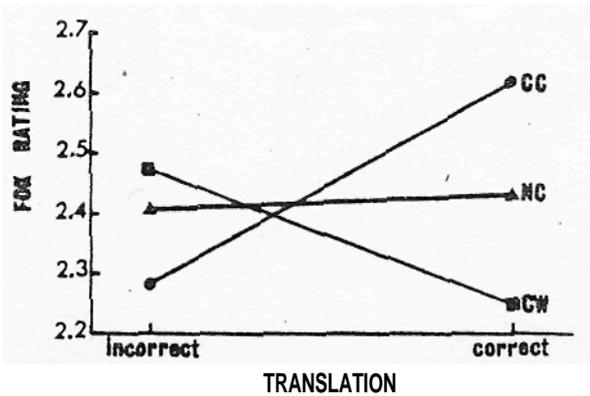


Figure 1. Mean FOK ratings for correct and incorrect translation, for CC, NC, and CW items.

The results, on the whole, appear to be more consistent with the "consensuality" hypothesis than with the "accuracy" hypothesis, although the monotonic negative trend observed for CW items did not prove significant. This latter result, it is suspected, may be due to the fact that, since the CW set includes only 19 items, too many of the individual translation accuracy scores calculated for this set had to be based on as few as a single item. To evaluate this possibility, the FOK categories were regrouped into weak FOK (ratings 1 and 2) and strong FOK (ratings 3 and 4), and individual translation accuracy scores were calculated separately for weak and strong FOK responses for each of the three sets of items. The means of these translation accuracy scores for weak and strong FOK responses, respectively, were as follows (including only individuals for whom both scores were available): 69.63 and 79.55 for CC items ($t = 6.27, df = 96, p < .001$), 48.46 and 50.54 for NC items ($t = .95, df = 95, n.s$), and 34.45 and 24.22 for CW items ($t = 4.27, df = 95, p < .001$). Thus, for CW items, stronger FOK ratings are associated with significantly lower accuracy.

On the whole, the results clearly indicate that the degree of subjective confidence is positively related to the likelihood of making the consensual response regardless of its accuracy.

Confidence Ratings of Correct and Incorrect Translations

In the previous report, the confidence-accuracy relationship was shown to be symmetrical in that responses which proved correct were associated with significantly stronger FOK ratings than responses which proved incorrect. To examine this relationship with the present data, mean FOK ratings for correct and incorrect translations were determined for each individual. This was done for each set of items separately and for all items combined. Over all items, the mean FOK ratings for incorrect and correct responses were 2.403 and 2.521, respectively ($t =$

5.29, $df = 99, p < .001$ }, thus replicating the finding of the previous study. This analysis, however, hides a more complex relationship, as shown in Figure I, which presents the results for the three item sets separately. For CC items, correct translations were assigned higher FOK* ratings than incorrect translations ($t = 8.26, df = 99, p < .001$), whereas for CW items, it is the incorrect translation that is endorsed with a stronger degree of confidence ($t = 4.64, df = 99, p < .001$). The difference obtained with the NC set is not significant ($t = .70, df = 99$). These results indicate that the individual's feeling of confidence is attuned to the consensuality of the response regardless of its accuracy. It is the consensual, rather than the correct, response which is felt to be compelling.

Another way to look at this matter is to examine the relationship, across items, between the percentage of correct translations of an item and the mean FOK rating associated with it. This correlation is .63 for the 38 CC items, .13 for the 28 NC items, and -.43 for the 19 CW items. Thus, the average FOK rating associated with an item is related to the degree of consensuality in responding to that item.

DISCUSSION

On the basis of the results obtained in the previous study, it was concluded that in a phonetic symbolism task subjects are able to monitor their success in guessing the meaning of words from foreign languages. The results of the present study, however, indicate that this finding is most likely a statistical by-product of the significant relationship which obtained in the previous study between the consensuality of a response and the likelihood of its correctness. This relationship, of course, is what should be expected to obtain if the idea of a universal phonetic symbolism is valid. However, the intentional inclusion in this study of a sufficient number of items for which the consensuality-accuracy relationship breaks down clearly revealed that subjective confidence is actually related to response consensuality rather than to response accuracy. The conclusions of the previous report should thus be revised to read as follows: In a word-matching phonetic symbolism task, the observer's subjective degree of confidence monitors the success of making a consensual response: the stronger the consensuality of a response, the more subjectively compelling it is felt to be.

The confidence-consensuality relationship in itself probably has little bearing on the issue of a universal phonetic symbolism. This relationship might be taken to suggest that, although the accuracy data support the idea of a universal substratum for sound-meaning connections, ratings of subjective confidence in the

accuracy of word matchings are in accord with culture-specific aspects of these connections.

A possible explanation for the confidence-consensuality relationship is that both response choice and confidence ratings are based upon the same features of the words being matched: the stronger the support in a given match, the stronger the confidence rating associated with it. Consensuality and confidence may thus be regarded as alternative indicators of the strength of a given match. The question then arises as to the nature of the word features determining the strength of a match. One possibility is that the strength of a match depends on similarities between the words being matched in obvious structural features such as word length, initial letters, and the like. If the strength of a match depends exclusively on the degree to which it is consistent with similarities in anatomical word features relative to the consistency of the alternative match, then it would be superfluous to invoke the concept of "phonetic symbolism" in connection with the confidence-consensuality relationship found in this study.

Several analyses were carried out to evaluate this possibility, the procedure and the results of which will not be reported in full in view of the detailed technicalities involved. The results can be summarized to indicate that although similarities in

anatomical features seem to affect both response choice and confidence ratings, the consensuality-confidence relationship is obtained even when this effect is controlled. Thus, consensually endorsed matches were found to be associated with higher FDK ratings than were nonconsensual matches, whichever match was favored by anatomical word features. These results suggest that the relationship between confidence and consensuality is primarily due to factors more appropriately subsumed under the concept of phonetic symbolism.

REFERENCES

- BROWN, R., BLACK, A. H., & HOROWITZ, Z. A. Phonetic symbolism in natural languages. *Journal of Abnormal and Social Psychology*. 1955. 50. 388-393
- KORLAT, A. Phonetic symbolism and feeling of knowing. *Memory & Cognition*. 1975. 3. 545-548.
- SLOBIN, D. I. Antonymic phonetic symbolism in three natural languages. *Journal of Personality and Social Psychology*. 1968. 10. 301-305.
- TAYLOR, I. K., & TAYLOR, M. M. Another look at phonetic symbolism. *Psychological Bulletin*. 1965. 64. 413-427
- TSURU, S., & FRIES, H. S. A problem in meaning. *Journal of General Psychology*. 1933. 8. 281-284.
- WEISS, J. H. Role of "meaningfulness" versus meaning dimensions in guessing the meaning of foreign words. *Journal of Abnormal and Social Psychology*. 1963. 66. 541-540.

(Received for publication March 31, 1975;
revision received September 23, 1975.)