

# AN EXPERIMENTAL STUDY OF LANGUAGE AND NON-LINGUISTIC BEHAVIOR<sup>1</sup>

HOWARD MACLAY

THIS EXPERIMENT was designed originally to test the Sapir-Whorf hypothesis that the structure of a language conditions non-linguistic behavior and thought.<sup>2</sup> That it is something less than a clear test of this proposition is related to the fact that neither Sapir nor Whorf stated the hypothesis in a form that could be proved or disproved in a single experiment.

Most previous investigations have consisted of an *ad hoc* demonstration that, in a given situation, some aspects of language could be shown to be congruent with some, usually very general, aspects of non-linguistic behavior. The point at issue is whether accurate predictions about relatively concrete and specific behavior in an experimental situation can be derived from the examination of language structure. The present study thus differs from most earlier work in three major respects: (1) the method of investigation is experimental rather than observational; (2) the predictions are made from present language to future behavior; (3) the behavior studied consists of specific instances rather than large scale patterns.

The general proposition that the language of a group is not functionally independent of non-linguistic factors would certainly meet with little objection. Disagreements arise at the point where attempts are made to specify the nature and direction of this relation. Both Sapir and Whorf usually wrote as if language were the initiating factor. Given this view, a convenient basis for prediction lies in the well-known fact that languages categorize the universe and further that referents united in one language are separated in another. This being the case it should fol-

---

1 This paper is revised from a University of New Mexico Ph.D. dissertation in Anthropology written under the direction of Stanley Newman to whom I am indebted for both specific advice and general encouragement. The personnel of the United Pueblos Agency and the Albuquerque Indian School were most cooperative and made a special effort to procure subjects with the proper linguistic background in addition to providing interpreters and space for the administration of the experiment. Charlie De Jolie acted as the Navaho linguistic informant and provided the formal subject interview which contributed materially to the experiment. Roger W. Weldon offered many helpful suggestions on the experimental design and the statistical interpretation of the results. This investigation has also benefited substantially from discussion with the author's colleagues on the Southwest Project in Comparative Psycholinguistics sponsored by the Social Science Research Council.

2 J. B. Carroll (ed.), *Language, Thought, and Reality: Selected Writings of Benjamin L. Whorf* (New York, 1956); D. G. Mandelbaum (ed.) *Selected Writings of Edward Sapir* (Berkeley, 1949). A recent evaluation of this position is found in Harry Hoijer (ed.), *Language in Culture* (Chicago, 1954).

low that referents classified together linguistically are likely to be classified together non-linguistically. This can be stated more precisely as follows:

If it is highly probable that referents *x* and *y* will elicit the same linguistic response, this will increase the probability that they will elicit the same non-linguistic response. Conversely, if they elicit different linguistic responses this should lower the probability that they will elicit the same non-linguistic response.

Extending this to a cross-language comparison to develop a testable hypothesis:

If Language A unites referents *x* and *y* and Language B separates them, then speakers of Language A should be more likely than speakers of Language B to exhibit the same non-linguistic response to them.

In this formulation a "linguistic response" involves the oral production of speech by an informant while a "non-linguistic response" refers to any other behavior. It is not obvious that such a rough distinction can readily classify every instance of human activity but it does serve to separate the particular behaviors used experimentally. "Elicit" refers to a procedure where an informant is shown potential referents and asked to respond verbally (i.e. "name them") or non-verbally (i.e. "divide them into groups"). Referents that elicit the same response are said to be "united," either linguistically or non-linguistically, while those which elicit different responses are "separated." Linguistic responses are the "same" if they contain the same phonemes and morphemes in the same order as previously defined by a linguistic analysis. "Same" or "different" for the non-linguistic responses refers to the alternatives present in the experimental situation.

The deduction of specific predictions from this general hypothesis is contained in a later section following a description of the experimental design. It should be stated in advance that the writer was thoroughly convinced that the hypothesis was sound and expected highly significant results.

#### EXPERIMENTAL DESIGN

This study focuses on the Athapaskan-speaking Navaho of New Mexico and Arizona. The structural statement which provides the basis for predicting experimental behavior is that of Hoijer who states:

The Athapaskan languages employ verb stems that refer not to a characteristic type of event, such as *stand* or *give* or *fall*, but to the class of object or objects conceived as participating in such an event, whether as actor or goal. . . .<sup>3</sup>

<sup>3</sup> Harry Hoijer, *Classificatory Verb Stems in the Apachean Languages* (International Journal of American Linguistics, vol. 11, pp. 13-23, 1945), p. 13.

Of the formally distinct categories described by Hoijer, three were selected for inclusion in the experiment: "long," "rope-like," and "fabric-like." These will henceforth be called "slender-rigid," "slender-flexible," and "flat-flexible," terms which make more obvious the physical characteristics of their referents. It should be noted that the English labels attached to these classes are a good, though not a perfect, approximation of their actual content in the case of the classes used experimentally. The practical consequences of this system are that one must use different stems in Navaho for saying "I pick up a cigarette," "I pick up a rope," and "I pick up a blanket." In his verbal behavior a Navaho is thus in the position of having made very frequent distinctions among objects on the basis of their form. He should, therefore, make these distinctions more readily in his non-verbal behavior than will a native speaker of a language not containing such obligatory distinctions.

A brief overview of the experimental design should make the detailed descriptions of procedure which follow more intelligible. Three groups of subjects were used: native speakers of Navaho, of English, and of a number of non-Athapascan American Indian languages. These were asked to divide sets of four objects into two groups of two objects each. Divisions could be made on the basis of the Navaho verb categories described above or on other grounds such as function or color. The primary expectation was that Navaho subjects would make significantly more divisions on the basis of form than would members of the other two groups. A central methodological requirement was that the procedure be as non-linguistic as possible. That is, an effort was made to prevent subjects from realizing that the experimental task had any connection with their language.

#### SUBJECTS

Sixty subjects were divided into three groups of twenty each. Sixteen high school students at the Albuquerque Indian School and four students at the University of New Mexico, all native speakers of Navaho, comprised the Navaho group. This group contained a range of English ability from none to almost perfect command. The group consisted of eleven females and nine males whose ages ranged from fourteen to thirty years and whose mean age was 17.9 years.

Sixteen students from the Albuquerque Indian School and four students of the University of New Mexico, all native speakers of non-Athapascan Indian languages, constituted the Pueblo group. This group contained a range of English ability from fairly good to almost perfect command and consisted of twelve females and eight males. Ages ranged from nine to twenty-eight years with a mean age of 16 years. The members of this group were all from Pueblos in New Mexico and

were distributed as follows: Isleta-3, Zuni-3, Laguna-3, Acoma-3, San Felipe-2, Santa Clara-2, San Ildefonso-2, Taos-1, and Sandia-1.

The English group consisted of twenty students at the University of New Mexico. All were native speakers of English and all were monolinguals with the exception of one subject who had spoken French for a number of years. The group contained six females and fourteen males. Ages ranged from seventeen to thirty-four years with a mean age of 21.75.

The types of subjects were chosen with several ends in mind. It had been demonstrated that the verb categories under consideration were obligatory in Navaho but not in any of the other languages represented. A minimum cross-cultural requirement could have involved a contrast between a Navaho group and a non-Navaho group. The English-Navaho comparisons would have satisfied this condition, but it seemed necessary also to have a non-Athapascan-speaking Indian group in order to insure that the results gotten would not merely reflect an Anglo versus Indian or majority-minority difference. The Pueblo group also acted as a check against the fact that the experimental materials were probably not equally familiar to Indian and non-Indian subjects.

#### MATERIALS

The materials consisted largely of objects common to American culture. There were forty-eight objects grouped in twelve Items of four objects each. The Items were further classified in four Types of three Items each. Objects are symbolized by small letters, Items by Arabic numerals, and Types by Roman numerals in this presentation. The Navaho verb categories are represented by SR (slender-rigid), SF (slender-flexible), and FF (flat-flexible); these are the categories reflected in the "form" classification indicated below. The category contrast present in each Item is described within the parentheses following the Item number. The "x," "y," and "z" notations refer to the three sorts possible for each Item and are defined as follows: "x" equals an ab/cd division, "y" an ac/bd division, and "z" an ad/bc division. The "y" division correlates with the Navaho verb categories in Types I, II, and III.

*Type I: Three bases for classification* [(x): Function or Material; (y): Form;  
(z): Color.]

Item 1: (SR vs. SF): a. blue ruler; b. green tape measure; c. green candle;  
d. blue electric cord.

Item 2: (SF vs. FF): a. blue rubber band; b. green rubber coaster; c. green  
cloth shoelace; d. blue handkerchief.

Item 3: (SR vs. FF): a. blue pencil; b. green sheet of paper; c. green shoelace; d. blue washcloth.

*Type II: Two bases for classification* [(x): Function or Material; (y): Form; (z): None.]

Item 4: (SR vs. SF): a. metal bolt; b. flexible metal wire; c. cigar; d. string of artificial pearls.

Item 5: (SF vs. FF): a. metal spring; b. aluminum foil; c. strand of yarn; d. irregular piece of cloth.

Item 6: (SR vs. FF): a. metal needle; b. handkerchief; c. paper straw; d. sheet of paper.

*Type III: One basis for classification* [(y): Form; (x) and (z): None.]

Item 7: (SR vs. SF): a. red ruler; b. gold colored chain; c. wooden paint brush; d. length of rope.

Item 8: (SF vs. FF): a. red rubber mat; b. metal pull-chain; c. sheet of paper; d. length of ribbon.

Item 9: (SF vs. FF): a. metal nail; b. sheet of cellophane; c. wooden match; d. irregular piece of cloth.

*Type IV: No basis for classification* [(x), (y) and (z): None.]

Item 10: a. metal padlock; b. yellow crayon; c. plastic sphere; d. cork.

Item 11: a. leather wallet; b. plastic disc; c. rubber band; d. metal hook.

Item 12: a. sheet of plastic; b. rubber eraser; c. small stone; d. bells tied on a ribbon.

The Item Types were intended to contrast the Navaho verb categories with other ways of classifying the materials. They varied in the number of alternative possibilities placed in the materials by the experimenter.

#### PROCEDURE

The basic procedure required the subject to divide the four objects in an Item into two groups of two objects each. Subjects were seated at a table facing the experimenter and given the following instructions:

I'm going to put four objects along this line in the center of the table. I want you to look them over and as soon as you become familiar with them, place two over here on the right and two over here on the left. You may find that they divide easily in two groups and, if you don't notice a natural pairing, and there may be none, please make a division anyway. There is no right division; I'm interested only in your own personal opinion. Now we'll have a couple of practice trials to give an idea of how this is going to work.

Each subject was given three pre-experimental trials to insure his understanding of the instructions. The writer instructed all subjects except eight monolinguals in the Navaho group. In these cases the procedure was explained by regular Navaho interpreters employed at the Albuquerque Indian School. Each subject was given all twelve Items with the order of Items and the placing of objects from top to bottom on the table top being randomly varied. Subjects were not given a speed orientation.

Two scores were taken for each subject on each Item: latency of response in seconds, and the division of objects (x, y, or z). In addition to the scores describing experimental performance the linguistic background of each subject in the Navaho and Pueblo groups was obtained in an interview immediately following the experiment. The subject was asked to estimate for each period in his life whether he had spoken his native language all of the time, most of the time, half of the time, less than half of the time, or none of the time. This resulted in a five-cell table for each subject which contained the absolute number of years he had spent speaking only his native language, mostly his native language, etc. Weights of 1.00, .67, .50, .33, and 0 were assigned to the cells and the sum of the cells was then divided by the subject's chronological age to obtain his Language Experience Index or L. E. I.<sup>4</sup> The scores on this index can range from .000 (never had spoken language) to 1.000 (had spoken language in question exclusively). Although this is a very rough way of rapidly estimating the linguistic experience of an individual, it should correlate fairly well with degree of bilingualism as measured by other methods. It may also be regarded as a measure of acculturation, as experience with language inevitably implies contact with the culture that uses it. In this context an L. E. I. of 1.000 would describe a relatively, but not necessarily completely, unacculturated person and degree of acculturation would increase as L. E. I. decreased.

There was a detailed interview with one Navaho subject on his reaction to the experiment and his interpretation of the whole situation.

#### EXPECTED RESULTS

A number of hypotheses were made in advance of the experiment. These are of three types: design validation, sorting, and latency.

##### *Design Validation*

1. The three groups should not differ significantly among themselves nor should any group depart significantly from randomness in the sorts made on Item

<sup>4</sup> The quantification of this index was suggested by Dr Charles Solley of the Menninger Foundation.

Type IV. This Type was designed to be a random array of objects and all groups should sort it randomly.

2. All three groups should have longer latencies for Item Type IV than for Types I, II, and III. The random Items, with no obvious classifications, should require a longer response time.

### *Sorting*

3. The Navaho group should make significantly more "y" sorts than the other groups for Types I and II.

4. The number of "y" sorts made within the Navaho group should correlate positively with the L. E. I., but this correlation should either not exist or exist to a lesser degree in the Pueblo group. This suggests that the more experience a Navaho subject has had with his language, the more likely he is to sort in terms of its categories. The Pueblo, on the other hand, should not show this correlation, since it is presumed that their native language experience has no systematic connection with the experimental materials.

### *Latency*

5. The Navaho group should not differ in latency among Types I, II, and III, while the English and Pueblo groups should have significantly longer latency for Type III than for Types I and II. The Navaho should have an equally available solution for all three types whereas the other groups should find Type III, with only a form solution, more difficult.

6. Navaho latencies for "y" sorts will be less than "x" or "z" sorts for Types I, II, and III, and this difference will be greater than similar differences, if such exist, within the English and Pueblo groups. This, and the other hypotheses involving latency, are based on the common psychophysical assumption that a difficult task will be associated with a longer response time.

## RESULTS

The results are organized in terms of their bearing on the hypotheses put forth in the previous section. Additional results of interest are also presented. Table 1 summarizes the sorting and latency results.

1. Chi-square tests on Type IV revealed no significant differences among the experimental groups and no significant variations from randomness in the sorting of any one group. This hypothesis is confirmed.

2. All groups had longer latencies for the random Items of Type IV than for the Items with built-in solutions, Types I, II, and III. An Analysis of Variance

compared the Types with regard to latency for each group.<sup>5</sup> The English and Pueblo groups had F's significant at the .01 level of confidence while the Navaho F was significant at the .02 level. Thus the trend of a latency increase through the sequence of Types is validated. However, the Navaho latency for Type IV does not

TABLE 1  
Sorting

<i>Type</i>	<i>Group</i>	<i>x</i>	<i>y</i>	<i>z</i>	<i>Latency (in seconds)</i>
I	English	34	16	10	190
	Navaho	33	17	10	340
	Pueblo	41	12	7	337
II	English	38	21	1	251
	Navaho	39	19	2	370
	Pueblo	44	13	3	401
III	English	3	51	6	303
	Navaho	12	46	2	626
	Pueblo	17	34	9	484
(I, II, III) Total	English	74	88	17	744
	Navaho	84	82	14	1336
	Pueblo	99	62	19	1222
IV	English	15	21	24	461
	Navaho	11	29	20	738
	Pueblo	13	26	21	694

differ significantly from the latency for Type III, while this difference is significant for the other two groups. This hypothesis is confirmed.

3. Chi-square tests were run on the sorting results for Type I, Type II, and the combined total of Types I and II, and in no case did significant differences exist among the three groups although the Pueblos consistently made fewer "y" sorts than the other groups. An unexpected result was the almost identical Navaho and English performance on these two Types. This hypothesis is not confirmed.

4. Product-moment correlations ( $r$ ) were run between L. E. I. and number

<sup>5</sup> The Analysis of Variance permits one to evaluate the differences among a group of means. The resulting value ( $F$ ) can be tested for significance and the probability of the differences having arisen by chance can be stated. See Allen L. Edwards, *Experimental Design in Psychological Research* (New York, 1950) for a full expansion of this procedure.

of "y" sorts for the Navaho and Pueblo groups. The Navaho correlation (+.48) was significant at the .05 level while the Pueblo correlation (+.19) was non-significant. The difference between the  $r$ 's was significant at the .05 level using a one-tailed test. This hypothesis is confirmed.

5. The Analysis of Variance described under Hypothesis 2 showed a Navaho latency increase for Type III as against Types I and II significant at the .01 level. The same difference for the other groups is significant at the .05 level. This hypothesis is not confirmed.

6. Mean times for "y" sorts as against "x" or "z" sorts were obtained for each group. Both the Navajo and English groups took less time for "y" sorts than for "x" or "z" sorts but these groups could not be separated by differences in relative latency. The Pueblo groups tended to have longer response times for "y" sorts than for "x" and "z" sorts. The results here are much like those under Hypothesis 3, with the Navaho and English groups falling together and the Pueblo group standing apart.

#### DISCUSSION

The confirmation of the design validation hypothesis (1, 2) on the presumed random sorting and increased latency for the objects in Type IV indicates that the subjects reacted to the Items in terms of the possible solutions placed there by the experimenter. This is further supported by the sorting results for Types I, II, and III where the built-in solutions were chosen far more often than the alternate possibilities where no obvious classification existed.

The central hypothesis of the study has been stated earlier:

If Language A unites referents  $x$  and  $y$  and Language B separates them, then speakers of Language A should be more likely than speakers of Language B to exhibit the same non-linguistic response to them.

Of the predictions of experimental results only the correlation between L. E. I. and number of "y" sorts in the Navaho and Pueblo groups occurred as expected and this single affirmative result is weakened by absence of the English group in the comparison. The direct comparison among the three groups with respect to both sorting and latency failed in every case to produce the expected differences (3, 5, 6). Clearly, the hypothesis must be rejected. This amounts to saying that, given a known linguistic category, it cannot be reliably predicted that non-linguistic behavior will correlate with it. The difficulty lies in specifying what "united" and "separated" mean in a concrete case. This arises from the fact that it is almost impossible to produce two objects that can be linguistically classified in only one way. The post-experimental interview with a Navaho informant who was asked

to name the objects in the experiment and also to use them in the frame, "I pick up a —," produced information which indicated that the verb categories were not the only bases for classifying the objects in terms compatible with prior linguistic experience with Navaho. It was sometimes the case that, in addition to the built-in form solution, two of the four objects in an Item had the same name in Navaho and could thus be said to be classified together linguistically on a lexical basis.

If every object and event participates in a number of intersecting linguistic categories one can never make the absolute statement, "Navaho does it this way," as if the language systematically cut the universe up into neat and mutually exclusive segments. It is rather the case that any language has a variety of alternative ways of classifying stimuli. If languages, then, can do anything, how can one possibly make meaningful predictions? A way out of this dilemma may lie in the distinction between Language and Speech or between language structure and language behavior. The structure is the abstract statement of the potentialities of the system made by a linguist. Language behavior is the observed verbal activity which such a statement describes. It is evident that a correlation will exist between the two if the structural statement is adequate. This will not be a perfect correspondence, since a process of abstraction can never take account of every variation in the data which form its basis. A structural description, furthermore, consciously ignores certain aspects of language behavior such as frequency of occurrence. Newman has stressed the importance of actual selection from structural alternatives in observational studies of relations between language and non-linguistic patterns of behavior.<sup>6</sup> In the case of experimental investigations relative frequency will probably be an even more important variable; perhaps the single most significant factor. This suggests that experimentation should be preceded by a thorough investigation of actual usage on the language behavior level. Several ways of grouping particular stimuli may be present in the structure; the experimenter must know in what proportion these are actually used.

It is probable that different linguistic structures have different potentials for indexing non-linguistic behavior. The results of this study indicate that direct predictions from structure, while they may correlate with large scale cultural patterns, are not likely to predict more concrete behaviors accurately unless supplemented by information on frequency.

UNIVERSITY OF ILLINOIS  
URBANA, ILLINOIS

---

<sup>6</sup> Stanley Newman, "Semantic Problems in Grammatical Systems and Lexemes: a Search for Method" (in Harry Hoijer, ed., *Language in Culture*, pp. 82-91, Chicago, 1954).