DOCUMENT RESUME

ED 025 001

EA 001 746

By- Thiemann, Francis C.

A Method of Synthesizing Large Bodies of Knowledge in the Social Sciences.

Pub Date Aug 68

Note- 15p., Paper prepared for presentation at the National Conference of Professors of Educ. Admin. (Albany, N.Y., Aug. 19-23, 1968).

EDRS Price MF-$0.25 HC-$0.85

Descriptors- *Computer Programs, Information Retrieval, *Information Utilization, Methodology, *Research Utilization, *Social Sciences

Employing concepts of formal symbolic logic, the philosophy of science, computer technology, and the work of Hans Zetterberg, a format is suggested for synthesizing and increasing use of the rapidly expanding knowledge of the social sciences. Steps in the process include formulating basic propositions, utilizing computers to establish sets, and determining deductive matrix relationships. (JK)

A METHOD OF SYNTHESIZING LARGE BODIES OF KNOWLEDGE IN THE SOCIAL SCIENCES

by

Francis C. Thiemann

University of Alberta

ED025001

Paper prepared for presentation at the National Conference of Professors of Educational Administration, Albany, New York, August 19-23, 1968.

By now an almost trite introduction to many papers in the social sciences is to recount the phenomenal growth and diffusion of knowledge in the past twenty years. Generally, these screeds develop the theme of one discipline's encounter with this problem. But, even without such accounts it is painfully obvious to both the practitioner and the academician that they are unable to keep up with the literature in their own area of specialization much less to find time to make the "fit" between the bits and pieces of related knowledge.

Suggestions have been made for a "publish control" to keep the literature explosion in check. One such suggestion would restrain young staff members from publishing 'until they had something to say.' Another suggests the need of a new format which would abbreviate the now more ponderous tomes flooding the market. Still another, the ERIC system, which has moved out of the suggestion stage, is providing abstracts in selected areas. While each of these suggestions and efforts has some merit, no one is foolish enough to believe that the tradition of "publish or perish" will soon be dropped, or that individual style will ever be replaced by some parsimonious rational scheme. Even the ERIC system, while providing excellent abstracts under scientific, rigorous controls, fails to provide the "fit", the interrelation between the parts, that requires so much of the scholar's time and effort. Realizing this and still confronted with the problem of time, volume, and relationship, a method needs to be developed to handle the ever increasing literary output from an ever increasing number of scholars.

It is the purpose of this paper, then, to suggest a method by which scholars with synthesizing interests may begin to alleviate the situation within their own disciplines and areas of specializations. The basic concepts expressed in this paper stem from the work of Hans Zetterberg[1], the philosophy of science[2] and computers[3], and the proponents of formal symbolic logic.[4] The method was initially developed in a study entitled, "A Partial Theory of Executive Succession,"[5] This present effort is an enlargement and extension of the methodological section of that work. The modifications are designed to increase the method's objectivity and to make more efficient utilization of the synthesizer's time and effort.

The format of this paper is: First, The Formulization of Propositions; Second, Utilization of Computers in Establishing Sets, Subsets, and Matrices; Third, The Deductive Relationships between Units Within a Matrix; Fourth, Other Logical Connectives to Increase Data Utilization; Fifth, The Extended argument: A Valid Synthesizing Devise; Sixth, Concluding Statement.

By the term proposition I mean any statement or assertion which is expressed by an indictive sentence and is either true or false. Since the same proposition may be expressed in a number of different languages each employing a sentence structure, the proposition is not to be confused with a sentence. Propositions are in the form of sentences but they are not identified by the form of the sentence. For example, if we were to say in three different languages, "It is very warm today." we would be using an indictive sentence in each case whose truth value could be established, but which takes a different form in each language.

Once this is understood, the major problem facing the synthesizer is that of abstracting from the literature the contained propositions.

Where the writers of the literature phrase their propositions in standard form, the propositions may be used as such. But, in most cases, the assertions must be rephrased. When this occurs the abstractor is faced with the difficulty of maintaining the original meaning while standardizing the form. In standard form each proposition is composed of two units joined together by a law-like connective. Initially each abstracted proposition employs the hypothetical ('If, then.') connective. Consider the following hypothetical proposition:

> If blue litmus paper is placed in a solution containing hydrogen ions, then the litmus paper will turn red.

In this example of a causal relationship[6] the first unit is 'blue litmus paper is placed in a solution containing hydrogen ions' and the second unit is, 'the litmus paper will turn red.' It is the connective "If, then.' that provides the relationship between the two units. These units, it is noted, meet the requirements of our definition of a proposition in that each unit is an indictive sentence in and of itself, and the truth value of each unit can be ascertained. By convention we symbolize the first unit of a statement proposition, called the determinant, by a capital letter. In this case we use the capital letter 'A'. The second unit, called the resultant, is assigned the latter 'B'. The 'If, then.' is symbolized by the horseshoe ($\supset$). By using such symbols the proposition if blue litmus paper is placed in solution containing hydrogen ion, then the litmus paper will turn red, is now shown as:

$$A \supset B$$

In many cases, however, researcher's do not see the world in such a clear two unit relationship so the abstractor is required to break the longer, more involved propositional sentences into several two unit propositions where either the determinant or the resultant is held constant. In other cases proposition are either not clearly expressed or are only implied; abstraction, then, must be by inference. In either case the apparent problems of value judgement are inherent. While this is to be considered a serious limitation, its degree of seriousness is in relation to how the propositions are to be used. If for instance, one is attempting to define the known limits of a body of knowledge, as we in this paper are suggesting, then it is not as serious a problem, as if we were attempting to deductively extend our knowledge while maintaining a high level of verifibility. One may doubt that the latter may ever be the case and yet some social scientists, Zetterber for one, would contend it is possible. It would seem more prudent to maintain that any proposition so derived needs to be empirically tested.

Once the propositions have been abstracted it becomes imperative to order them in some logical way. The method of ordering may be visual or intuitive when there are only a few propositions, but once they exceed 3 or 4 some mechanical means needs to be devised to conserve the synthesizer's energy and to eliminate the fictionalizing so prevalent among social scientists when they are forced to analyse large amounts of data. This fictionalizing has been more commonly referred to as selective analysis. But, with the aid of computors this can be eliminated.

Computer programs are available which will handle string data. Two such programs are PL1 and Snobol.[7] "A string is a sequence of characters,"[8] whose length may vary from one string to another. Each unit of a propositions could be a string or each proposition may be used as a string. Since both programs provide the same output for the needs of this study, no preference is given.

To see how one might employ string data programs let us assume that we have 1500 propositions regarding executive succession and we want to establish sets and subsets. Where a set is defined as "any collection of definite, distinguishable units ordered by the intellect and treated as a whole,"[9] we could, from a perusal of the literature intuitively establish the set of Successor and the subsets of Inside Successor and Outside Successor. We would then read each of the 1500 propositions to determine in which set or subset they best belong. If a proposition did not fit, i'. could either be discarded or a new set created. One could continue doing this until each of the propositions was placed in one or more of the sets and subsets. But, by using one of the suggested string data programs, the computer could print out by rank order a frequency list of each word used in the 1500 propositions. Once this frequency dictionary has been established, selected combinations of words could also be run. And by use of a thesaurus, such as the one used by ERIC, words or phrase could be brought together under a single term for reduction by definition. For example, using executive succession as our major category, we may find in our dictionary that the term "anticipatory" was used x number of times and successor was used y number of times and now we would like to know how many times anticipatory was used in conjunction with successor. Or, again, how many times anticipatory is found in combination with Inside Successor and Outside Successor. Or, to employ definition reductive words such as Executive, Chief Administrator, Leader, Ruler, President may be brought under one term such as the Old Man. Determining frequency relationship and the relationship between words provides one with a more objective means of establishing sets and subsets then the aforementioned visual and intuitive methods. Following the establishment of the sets the computer can provide a listing of all the propositional units within each designated set. Set refinement, of course, is a matter of individual choice and to a certain extent is dictated by the number and the nature of the propositions involved and the developmental level of the discipline.

With the establishment of the sets another machine program can be written which would list each unit of each proposition within a set with every other unit. Thus, if units A, B, C, and D, were used in set #1, the computer would print out in hypothetical form:

$$A \supset B; A \supset C; A \supset D; B \supset C; B \supset D; \text{etc.}$$

Each of these relationships could then be read and a judgement made as to its acceptance or rejection. This would provide one side of a symmetrical matrix with the other side being determined by reading each of the above hypothetical propositions in reverse. These could then be plotted in a matrix to provide a visual check or a row and column frequency count of filled and voided cells. A simplified matrix where all the units in combination are reversible may appear as:

	A	B	C	D
A		$A \supset B$	$A \supset C$	$A \supset D$
B	$B \supset A$		$B \supset C$	$B \supset D$
C	$C \supset A$	$C \supset B$		$C \supset D$
D	$D \supset A$	$D \supset B$	$D \supset C$	

If one, however, accepted $A \supset C$ but rejected $C \supset A$, he might show the $C \supset A$ cell as not reversible by means of a slash (/) in that cell area. Or, if neither $B \supset C$ or $C \supset B$ was accepted then an X would be placed in each of the cells.

With the completion of the matrices for the desired sets and subsets basic logical arguments can be established. Where the statements we have been discussing have been called hypothetical propositions, the argument called, the hypothetical syllogism or the Chain argument, may be formed by selecting two or more hypothetical propositions which can be joined together in such a way that the resultant of the first premise (the first assertion in the argument) becomes the determinant of the second premise. If more than two premises are used in the argument, then the resultant of the second premise becomes the determinant of the third premise and so forth. Given any number of premises so connected, a conclusion may be formally constructed by combining the determinant of the first premise with the resultant of the last premise used in the argument. Symbolically, we may represent these as:

$$\begin{array}{rl} A \supset B & \text{first premise} \\ B \supset C & \text{second premise} \\ C \supset D & \text{third premise} \\ \hline \therefore A \supset D & \text{conclusion} \end{array}$$

The B and the C are the linkages that join the determinant A in the first premise with the resultant D in the third premise. Two premises are a sufficient condition for the hypothetical syllogism form, while the argument's form is a necessary condition for its validity.

To have a valid argument form, it is necessary that in no case the premises be true and the conclusion false. Since truth and falsity are related to statements, a premise may be said to be true or false, but an argument is only described as valid or invalid. The truth value of a connective is determined by means of a truth table. To determine the truth value of A $\supset$ B, one must follow the rule which states, if the determinant is true and the resultant is false, then statement A $\supset$ B must be false. All other relationships are consigned as true. The following illustration shows the truth table value for A $\supset$ B:

A	B	A $\supset$ B
T	T	T
T	F	F
F	T	T
F	F	T

It is obvious, then, that each proposition in each cell of an all reversible matrix could begin a hypothetical syllogism as its first premise. If all units were reversible, 24 hypothetical syllogisms are possible from the above 4 x 4 matrix, but if any of the cells are empty or if an assertion is not reversible, then the possible chains are lessened. In the above matrix example, if two cells A $\supset$ D and D $\supset$ A were empty so that no relationship existed between the two, then 16 of the 24 possible chains would be eliminated. On the other hand, if the statement D $\supset$ A were not reversible and a slash (/) appeared in cell A $\supset$ D, 8 of the syllogisms could not be formed. It is apparent that when units are not reversible a considerable amount of data is lost to the hypothetical syllogism.

Now up to this point we have discussed the theory building concepts of Zetterberg, but since he utilizes only one of the four connectives of symbolic logic and makes no reference to negation, the logical meanings of these other connectives needs to be briefly discussed before we can incorporate all the data of a partially filled matrix into an argument. The connectives will be briefly discussed in the following order: disjunction; conjunction; material eqivalence; and negation.

The disjunctive connective is symbolized by a 'v' which is known as the 'vel'. It is a logical connective that joins two units called 'disjuncts' and is interpretated in its inclusive sense. With disjunction it is true in every case except where the disjuncts are both false. Symbolically, it appears as A v B and is read A or B. Its truth table is:

A	B	A v B
T	T	T
T	F	T
F	T	T
F	F	F

Conjunction is represented by the dot (.) and connects two units called conjuncts. In symbolic form the statement would appear as A . B and are read as A and B. The truth value of this connective is true only when both conjuncts are true. This is illustrated by its truth table:

A	B	A . B
T	T	T
T	F	F
F	T	F
F	F	F

Material equivalence is symbolized by three lines (≡) and is read as 'if and only if' or 'is materially equivalent to.' The latter the most commonly used meaning. Material equivalence is true only when the units are both true or are both false.

A	B	A ≡ B
T	T	T
T	F	F
F	T	F
F	F	T

Finally, negation, symbolized by the dash (-), of either a single unit or of a compound statement, is the opposite value of that unit or statement. When A is given as true, the negative of it is false. If the negative A is again negated, then it is materially equivalent to A.

With the inclusion of these connectives and negation any statement or unit within a matrix which could not be joined before by the hypothetical can now be brought into the argument. For example, in one matrix developed in the aforementioned study of Executive Succession two units from the work of O'Donovan were used. He stated that Successor's see their promotion was due to their own individual merit and experience. He also notes that these same executives perceive the advancement of their peers to have been due to seniority. If we were to attempt to combine these into a hypothetical proposition, it would read as follows:

> If a successor sees his own promotion was due to merit and experience, then he sees other promotions are due to seniority.

By using conjunction, however, this may be stated as:

> A successor sees his own promotion was due to merit and experience, and other like promotions were due to seniority.

Or, in a disjunctive relationship as:

> A successor sees his own promotion was due to merit and experience, or that other promotions were due to seniority.

Lastly, by material equivalence:

To say that a successor sees his own promotion as being due to merit and experience, is materially equivalent to saying that he sees other promotions were due to seniority.

Now, in each of these statements there is a degree of truth which the individual synthesizer must value out for himself. One proposition is much stronger (logically) than the others. Some have implications so weakened as to be meaningless (in a semantical sense), but by one or the other uses of a connective the units can be brought together into a new form called the extended argument. As the hypothetical syllogism was a deductive method showing the causal relationships between units and linking units in two or more statements so that the reduction of the total units involved are expressed by the units in the conclusion. So, the extended argument does the same thing by using three additional connectives and negation.

With the matrices still providing the bases from which the primary units, that is, the reversible units, are chained together, the other relevant non-reversible secondary units can be included in the argument as additions, exceptions, equivalences, or negations.

For example, if we had nine units in a matrix, three of which could not be hypothetically joined, all nine could be used as in an argument by using symbols:

$$[(A \vee B) \supset C] \supset D$$

$$D \supset (E \cdot F)$$

$$[(E \cdot F) \supset G] \supset (H \vee I)$$

$$\therefore B \supset I$$

In this case ($A \supset B$), ($E \supset F$), and ($H \supset I$) were too strong a relationship for the variables involved. Thus by placing them with weaker connectives they could be included in the argument. It is also noted that the conclusion is a reduction of the argument in which only true statements are used and where the argument form is valid.

As units can be related by the four connectives and negation, the length of arguments can be increased to the point that validity by truth table method is no longer practical. This is obvious when one considers that an argument with eleven units would need a truth table with 2,048 rows. It is necessary then that another method be employed to determine validity. Two methods are available: one, is a modified truth table which hypothetically attempts to establish a case where the conclusion is false and the premises are true. The second method, employs inferences from simple arguments and logical equivalences as developed by symbolic logic. While it is possible that both of these methods can be programmed for the computers, a number of problems, mostly human, have plagued our attempts. Simple arguments have been validated thus far and there are encouraging signs that the problems of the Complex Arguments will soon be solved.

By constructing within and between set extended arguments the synthesizer and researcher of the future will be able to reduce large bodies of knowledge to manageable size from which general propositions can be constructed and testable hypothesis can be formed. This is a process, not unlike the methods mathematician and physicists use to reduce large numbers of variables. An excellent example from theoretical physics is Einstien's $E = mc^2$. A formula which represents a number of expressed and unexpressed variables and relationships. Since the number of extended arguments is limited to the number of possible combinations of units involved, the empirical data gathered from a research study can be played against the derived arguments to ascertain which has the best fit. This of course implies that a number of theories can be tested at one time with one collection of data.

The implications for such a method is obvious when one considers again the great bodies of knowledge found within each discipline which are in one way or another related. It is possible that by having synthesized data available the advancement of the social sciences will leap forward with the greatest thrust in their history. More specifically, in the area of school administration not only would the synthesization of relevant knowledge be of value for more precise and intensive research but the developmental aspects could lead to a more objective rational selection of people, materials, and procedures in plant and systems operation.

FOOTNOTES

[1]Hans L. Zetterberg. On Theory and Verification Sociology. New Jersey: The Bedminster Press, 1965.

[2]Richard B. Braithwaite. Scientific Explanation. Cambridge. Cambridge University Press: 1964, p. 22.

Karl R. Popper: The Logic of Scientific Discovery. New York: Basic Books, Inc., 1961

[3]Allen Forte. Snobol 3 Primer. Cambridge, Mass: The MIT Press, 1968.

[4]Alfred North Whitehead and Bertrand Russell. Principia Mathemalica (1910), 2nd ed. Cambridge: Cambridge University Press, 1950.

Susanne K. Langer. Symbolic Logic. New York: Dover Publications, Inc., 1967.

Robert R. Stall. Sets. Logic and Axiomatic Theories. San Francisco: W.H. Freeman & Co., 1961.

[5]Francis C. Thiemann. A Partial Theory of Executive Succession. Final Report. Eugene, Oregon: Center for the Advanced Study of Educational Administration, 1968.

[6]The causal relationship while the most commonly held and used is but one of the four possible meanings contained within the hypothetical connective. This connective may also represent a decisional relationship, a logical relationship and a definitional relationship.

[7]The Programming Language. LISP: Its Operation and Applications. Cambridge, Mass.: The MIT Press, 1966.

[8]Forte. op. cit., p. 4.

[9]Thiemann. op. cit., p. 10.