COMPUTER ANALYSIS OF FOLK-BELIEFS

ILDIKÓ FEHÉR

József Attila University Szeged

Computers are widely used today in different fields of the humanities, including folkloristics as well. The material one has at hand is immense. Such masses of data are, as one can guess, impossible to file manually. Computers, however, are ready to help, not merely with the files but also to discover new relationships. Although one should not expect wonders from applications such as these, we may have the promise of significant achievements, if we can formulate a problem in accordance with the characteristics of the machine. In the present paper I am going to describe two approaches which I have already put to the test, and which I have used in an attempt to ground a computer analysis of folkbeliefs. I ran the necessary programs as an experiment in PASCAL, on the R40 computer of the József Attila University Kalmár László Cybernetics Laboratory.

I. Quantitative (statistical) applications

Ethnographers describe their collected data together with several additional data such as the place of collection, the age, sex, religion, etc. of the informants (additional data will henceforth be referred to as *properties*). Quantitative

summarizing of these properties can yield important results, for instance on the range of certain kinds of data in a given area (or in age groups, sexes or religious groups); we can further examine the existence of data occuring together, or occuring only in certain areas, and whether the occurence of the data is subject to subordination. Ethnographers have always done this sort of summarizing but hardly ever without manual labour. Machines, on the other hand, provide exact results within seconds after the data-input, and can even display them in a graphic way with the help of mathematical-statistical methods (like frequency and correlation analysis). The utility of this kind of work is hardly debatable, as computers process only those properties of the data that are independent of feelings and can be described objectively.

The most difficult and most highly responsible part of the work is coding, i.e. formulating the data in a way fit for computer processing. It is the ethnographer who decides what pieces of information he considers relevant describing data, and where to draw the line separating the different properties (i.e. what it is that he still includes in a property-group, and what it is that he does not). In the most frequently used method of coding each property type corresponds with a column, and the possible property occurences (within a column) with a sign (numbers, or number-letter combinations)

Eg: 1st column: Place of collection

(within this) 01: Bátya

02: Kalocsa

03: Dusnok

.
.
.

2nd column: age of informant

I (old) above 60

K (middle aged) 30-60

F (young) below 30

3rd column: sex of informant

F: male

N: female

If this cannot be done (because for example a piece of data has several properties of one property type), we can list the possible properties and put 0 against those not belonging to and 1 against those belonging to by the given piece of data. (This way of coding is not very space-economical, but is sometimes unavoidable.)

The coding done, the collector transfers - if possible - his data to some data-carrier (punched cards, magnetic tapes), so that he can have the machine sort out for him the data with given properties. Besides statistical processing the significance of this method in terms of speed should also be appreciated, as a lot of time can thus be saved in comparison with reading the files one by one.

II. Non-statistical applications

The significance of non-statistical treatment may be more controversial than that of the former. Results - if any - can only be achieved by teamwork among experts.

We have attempted such a novel application with a collection available to us, containing beliefs in connection with death. We examined omens of death from three areas far from each other. We drafted the data in a special form inspired by Minsky's frame-theory (2). Minsky regards as frames data-structures that describe stereotype situations. A frame consists of several parts and among other things, contains information as to the relationship of the frame with other frames. Below, we depart from Minsky's theory and will use *frame* in the following sense:
(1) There are two types of frames: action type to describe actions, and state change type to describe changes of states.
(2) A frame consists of the following five fields defined by the type: in the case of action type: 1 frame type
2 action
3 actor
4 object
5 possible result

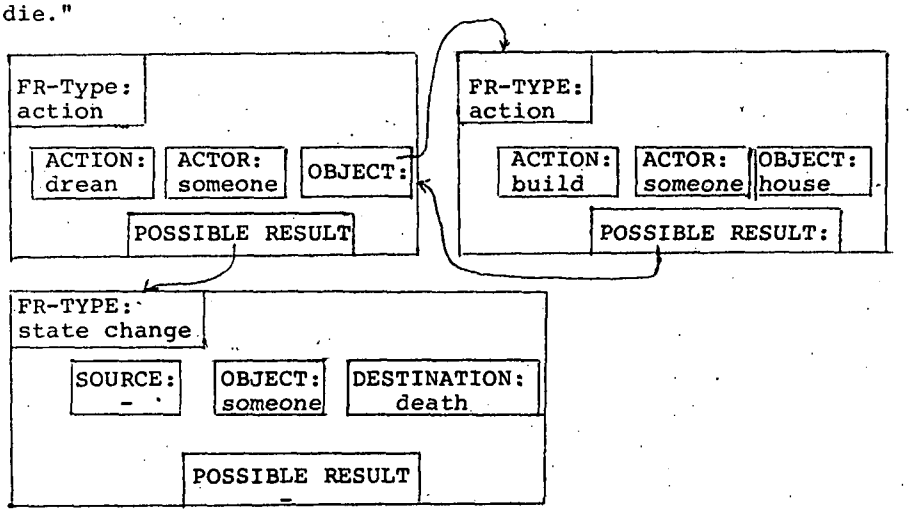
in the case of state change type:
1 frame type
2 source
3 object
4 destination
5 possible result

(3) Within either type fields 2, 3, and 4 can be further frames, and any of the fields can stay blank.

The belief-collection we used contains the beliefs in the same natural language code as they had been recorded on tapes. As we had to intention of examining the grammatical regularities, the frames thus defined were suitable for describing our data.

Here is an example to illustrate the meaning of the individual fields:

"If someone dreamed that he was building a house, he was to die."



(The first frame here is action type and its object field is another frame.)

The coding looked like the following:

The beginning of the frame is marked with a left parenthesis.

Then the type follows, then the fields separated by commas. If a field is blank, the comma is there, nevertheless. If a field is a further frame, it is parenthesized. End of frame is marked with right parenthesis.

Our example coded:

(action, dream, someone, (action, build, someone, house,), (state change,,someone,death,))

This formulation is apparently simple enough for coding belief statements quickly, and also contains sufficient information necessary for certain examinations. From the data so coded we can look up a belief e.g. on the basis of one of its words, or its frame-type; but also, the function of a word within the given belief can also be determined (e.g.: we know that in action type the word after the second comma after the initial parenthesis denotes the actor(subject)).

The study we have done does not utilize all these possibilities. We could have done well with simpler coding but our purpose was not to aim at results but only to illustrate possibilities. The way I formulated the data for storing in suitable for later (and possibly better) treatment.

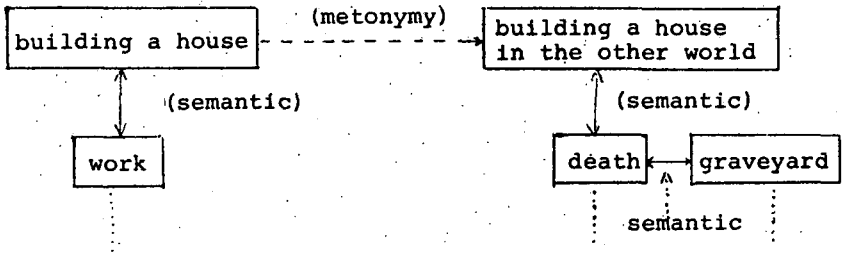
Scrutinizing the data of the collection it stood out that some belief-statements contained "explanations" too, like building a house in one's dream meant death because it is but building a house in the other world; or if one weaves in a dream, she/he will die because the loom means a coffin. It

is obvious that the members of the "house - otherwordly-house" and "loom - coffin" pairs are in some way related. This relation is called *metonymy*. Apart from this we have found two more sorts of relations among the words used for filling in the frames. One is a *consequence*-relation, with word-pairs, in which the second is the result of the first word, e.g. (tree) falls → dies.

Semantic relation exists between words which are related by their meaning, e.g.: all the words {coffin, tombstone, priest, candle, grave, graveyard, other-world} belong to the same semantic field, and any two of them are semantically related.

Following from this we put down all the word-pairs whose members were in some way related and fed them into the machine. A program made especially for this purpose constructed a graph of these connections, where angular points were the words used for the coding, and two vertices had edges between them if the two words represented by the vertices were in some way related. In such cases we labeled the edge with the relation type.

e.g.: (part of the graph)



From the point of view of application those words were considered *relevant* which did not belong to any of the following: action, state change, see, hear, dream, something, someone. A process, which fed in a belief description and picked two relevant words out of it, found a way between these two words in the graph previously constructed. First it found the first word, then setting out from it, started going around the graph in all possible ways, and displayed all routes which somehow led to the angular point symbolizing the other word. Thus, the essence of the application: with the help of the connections within the data we tried to find relations between the omens and their consequences. The result depends to a great extent on the quantity and the quality of the material fed into the computer for the construction of the graph. Obviously, the addition of a few new connections to the graph may increase considerably the number of possible connections. In the case of a mass of data, the machine might give us clues to discover new relations which might have been difficult to trace by the human mind alone (not only because they are so complicated, but also because one has to consider to several things at the same time, to keep hundreds of relations in one's head and to apply them in the right place). The computer remembers all the relations and uses them at the appropriate places, too.

The connections printed on paper should not be considered to be full explanations, the less so as there are several different routes between two words. The ethnographer still

must do the lion's share of the work, examining the results he has obtained. These may just give him ideas how to go on. He can try out his own hypotheses too, seeing to what extent the number, length, etc, of the exploratory routes change after adding a new connection to or taking an old one out of the graph.

A great number of possibilities exist besides those described above.

Experts in different fields should work together so that these possibilities may be realized.

Notes

1. For a detailed account of the programs and the results see my diploma-work. I. Fehér, 1982.
2. Minsky, M. 1975. quoted by P.H. Winston: Artificial Intelligence Addison-Wesley Publishing Co., Reading-Menlo Park-London-Amsterdam-Don Hills-Ontario-Sydney

Literature

Ildikó, Fehér: Computer Analysis of Folk-Beliefs, Szeged, 1982.

Minsky, Marvin: A Framework of Representing Knowledge in: Patrick Henry Wilson(ed): The Psychology of Computer Vision, McGraw-Hill Book Co., New York, 1975.