AFFILIATION PROGRAMS

In the Classics Dept. of Göttingen University a series of colloquia has been held during the last terms on the subject. Generally led by the basic work of Father Froger (1) the theoretical background and the chances of the realisation of automatical procedures as essential part of the editorial task have been discussed. Tab. 1 gives the flowchart of the entire project.

There are two magnetic tapes, one with a standard text, the other one with the *variae lectiones* (for their genesis see below). The *variae lectiones* are sorted according to their occurrence in the passages of the text (see Tab. 2). The result of this sorting is the table of constellations (see Tab. 3). By 'constellation' I mean this : Given for instance 3 manuscripts M_c, M_1, M_2. M_1 and M_2 are compared with M_c. Then there are 4 possible constellations for a given passage, namely :

Does M_1 or M_2 resp. agree with M_c ?

M_1	yes	no	yes	no
M_2	yes	yes	no	no

If there are n manuscripts besides M_c, then there are 2^n possible constellations. By ordering and compressing this table of constellations a table of

types of constellations (see Tab. 4) is generated. It ist important to notice that the standard text with which the other manuscripts are to be compared has been chosen at random. If a manuscript differs from the standard text, it is theoretically uncertain whether the standard text is wrong and the manuscript right or vice versa. The ideal condition would be that M_c is identical with the autograph. In that case any deviation from M_c is a mistake. For the random chosen M_c the model-copy-relation is replaced by the one-step-distance relation. The more manuscripts are 'yes' in a constellation the greater is the distance of this class of manuscripts from M_c. Normally in the stream of tradition a mistake is not corrected and a distance is not reduced. The computer program AFFILI, which is to be applied next, can find a system of mutual relationships only but not the direction of the stream of tradition. A strict mathematical method to solve this problem has been described by Dr. Najock in his contribution; a less strict and more simple procedure is offered here.

If one of the assistant scholars (who did the collations) stated that at least one passage of the standard text is undoubtedly wrong and the *varia lectio* right, then normally all the passages of the standard text belonging to the same type of constellation can and must be replaced by the corresponding *variae lectiones* and vice versa (2). The standard text or collation manuscript will now resemble the autograph much more than before. The program starts again with setting up the list of constellations and that of the types of constellations. Now the computer program AFFILI converts the list of types of constellations into a stemma of manuscripts (see Tab. 5 and 6). This is now checked as to whether it yields information or not. As will be said in the explanation to Tab. 4 the resulting stemma can happen to be completely uninformative when many manuscripts appear more than once

because of corrections or contaminations made by the copyists by successful reasoning or by checking another - better - manuscript or by mere chance. Unluckily success is far less frequent than failure, i.e. corrections by chance are expected to be extremely rare and those by reasoning relatively seldom. Therefore it is promising to start the program AFFILI again and to consider those types of constellations only that have a certain minimal frequency. Further : As will be shown in the contribution of Mr. Ahnert, it is necessary to add to each entry of the table of *variae lectiones* a degree of certainty that the text was read correctly, since a second reason for an uninformative stemma may possibly be that uncertain readings have been considered the wrong way. The program has to be started again, this time considering only those *variae lectiones* that have the highest degree of certainty.

Additionally it may be desirable to know an estimated degree of probability for a deviation from M_c because a very likely deviation as e.g. itacism or accentuation is of little influence on the flow of tradition. We expect that the resulting stemma will give a clearer picture when only v.l.'s with a low degree of probability are considered. If the stemma is still uninformative, the ultima ratio will be to start the program once again, this time simply skipping that manuscript which appears most frequently in the stemma.

Explanation to Tab. 2. Given a set $\mathfrak{M}$ of Manuscripts M_1, M_2, ..., M_n of a certain ancient literary work the textual condition of which is such that in a great majority of cases the different copies have the same wording. One of the manuscripts is picked out theoretically at random and called collation manuscript (M_c). Actually an old and good one will be the most proper choice. It is copied in machine readable form with line and columns numbered.

The passage in M_c is then defined by p (a,b), where a and b are real numbers with their integer part denoting the line, their fractional part the character number.

For instance the word *'perdas'* is denoted by p (4711.15, 4711.22) or shorter p (4711.15,.22). Both the blanks in front and behind the word have to be considered. The prefix *'per-'* would be denoted by p (4711.15, .18). In this case the fact that the passage starts with a blank and ends with a non-blank character tells that a prefix is denoted. All the manuscripts are compared with M_c. If for example M_7 differs from M_c by having *'operas'* for *'operam'* a note is taken that reads

$$M_7,\ 4711.14,.14,s$$

where M_7 stands for the manuscript being compared, the two real numbers as said before, s is the reading of M_7 at passage p (4711.14,.14). The system can only work when there are never two notes with the same passage and a different reading, i.e. : for any passage there are only one or two wordings, namely the text of M_c and sometimes one different reading.

If there is more than one different reading for a given passage, consistency of the system can easily be simulated by varying the limits of a passage and thus defining a new one. These notes are stored on a tape called *'variae lectiones'*.

Explanation to Tab. 3. The *variae lectiones* are ordered according to the

passage referred to, and the ordered v.l.'s are converted into a list of constellations. A constellation indicates for a certain passage and a certain wording which manuscripts have the wording of M_C and which differ from it by having the recorded *varia lectio*. Take the manuscripts A, B, ... H. For representation in the computer memory each manuscript is assigned a certain bit within a machine word or a string of machine words (or bytes). If for example the manuscript A is the collation manuscript M_C and if B is differing from A, the corresponding bit string looks like this :

Sigl.	A	B	C	D	E	F	G	H
bit nr.	0	1	2	3	4	5	6	7
bit val.	1	0	1	1	1	1	1	1

Then the manuscript having the respective *varia lectio* is represented in the machine memory by the above string of bits. The procedure for generating a constellation runs like this : each *varia lectio* entry is preceded by one (or more) all-one-bit-word(s). The bit with the number of the manuscript concerned is put to zero as shown. From this we get the constellation by ANDing all the manuscript bit strings of all the *varia-lectio-entries* for the given passage.

Explanation to Tab. 4. The table of types of constellations is generated by ordering the table of constellations according to the manuscript bit strings taken as binary numbers in descending order. The equal strings are counted and entered only once into the table of types of constellations together with their frequency.

Explanation to Tab. 5. The stemma matrix is deduced from the table of types of constellations. Each type of constellations belongs to a - let us call it - error level i.e. : suppose M_c is the all correct autograph. Then an error type of only one manuscript did not create successors and therefore belongs to the first level. An error- or constellation-type of two manuscripts belongs to the second level etc. A constellation type of say 4 manuscripts tells us that one of their copyists made a mistake and the others copied it. The first first-level type is taken and entered into the matrix as shown. Then the next entry into the table of types is checked as to whether its zero-bits include those of the preceding entry (a logical FORTRAN function made ad hoc and called SUBSEO is used. SUBSEO (n,m) is true, if the ϕ bits of n are a subset of those of m, SUBSEO (n,m) is false if that is not the case). If this is true, the zero-bits are counted and the constellation type is entered as shown (in the table the manuscript symbols stand for the zero-bits).

Philological interpretation : On the third level for instance there is a type CGH as predecessor of H and G. It can easily be seen that C made an error and that C was copied by H and G with at least one additional error each. On the second level there is a type DE, which signals that a lost manuscript introduced this or these error(s). Then it was copied by D and E. In general : If there is more than one entry in one column, these entries denote the common source of all the error types left of them in the same line. Looking for the father of H and G for example, it can be found in column 3 being that manuscript which does not occur on the left hand side in the same line. The FORTRAN procedure : The formula :

$$y = \text{COMPL (XOR (x,y))}$$

, where y is the entry to be reduced, x the one on the left hand side on the same line.

If by this procedure all bits of an entry have been turned to one, a lost common source, symbolized x in the graph, has been found. To avoid confusion in case of two adjacent lost sources two different symbols should be applied alternatively. The final stemma matrix is in itself explanatory and can easily be converted to a conventionel stemma. The simple stemma given here represents the ideal and extremely rare case that each copyist had only one manuscript at hand and did not try to correct an error of his text. If so-called contaminations occur, a manuscript will appear more than once in the stemma signalling that different parts of it have a different origin. What has to be done in this case has already been pointed out at pg. 2 and will be further dealt with by Mr. Ahnert. AFFILI was programmed in FORTRAN V for UNIVAC 1108 and tested under extreme conditions. Bit string operations were performed by using the FLD function supplied by that compiler. If it is not available FLD can be replaced by mask operations. Especially those who would like to join us in practical applications of the above system please write for further information and the programs.

GÖTTINGEN University

Prof. Dr. Jürgen MAU.

NOTES

(1) Jacques FROGER, *La critique des textes et son automatisation,* Paris, 1968. Cf. Anežká VIDMANOVÁ, *Les textes contaminés et l'ordinateur,* dans *Revue de l'Organisation internationale pour l'Etude des Langues anciennes par Ordinateur,* 1972, fascicule 1, pp. 5-22.

(2) Corrections occur rather seldom; therefore with a high probability the observed better *varia lectio* is no correction but the original reading. Then the other *variae lectiones* of this type of constellations - if we can exclude contamination - are original too. Otherwise i.e. if they are mistakes, the observed better reading (which has been produced by the same scribe) would be a correction; this however has been assumed to be very unlikely.

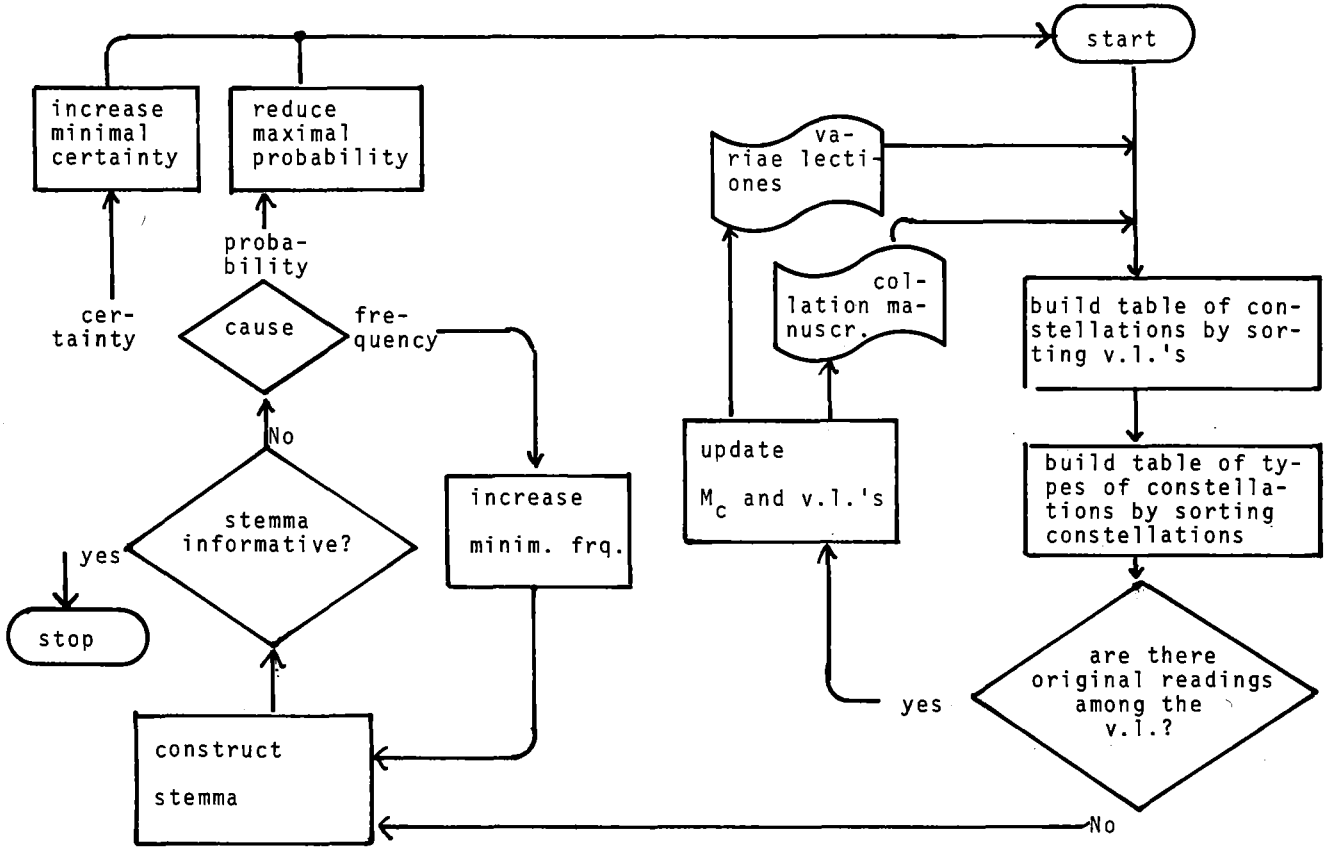

Table 1

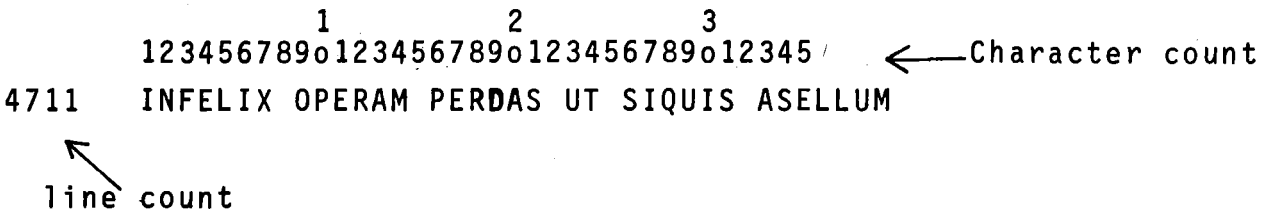

Table 2

Generation of a

Constellation table entry

M_c	A	B	C	D	E	F	G	H
1	o	1	1	1	1	1	1	1
1	1	o	1	1	1	1	1	1
1	1	1	1	o	1	1	1	1
1	1	1	1	1	o	1	1	1
1	1	1	1	1	1	o	1	1
1	1	1	1	1	1	o	1	1
1	o	o	1	o	o	o	1	1

Table 3

Sample stemma and corresponding Table of

Types of Constellations

M_c	A	B	C	D	E	F	G	H	frq.
1	1	1	1	1	1	1	1	o	5
1	1	1	1	1	1	1	o	1	7
1	1	1	1	1	1	o	1	1	1
1	1	1	1	1	o	1	1	1	1o
1	1	1	1	o	1	1	1	1	3
1	1	1	1	o	o	1	1	1	4
1	1	1	o	1	1	1	o	o	3o
1	1	o	1	o	o	o	1	1	1
1	o	o	1	o	o	o	1	1	1
1	o	o	o	o	o	o	o	o	6

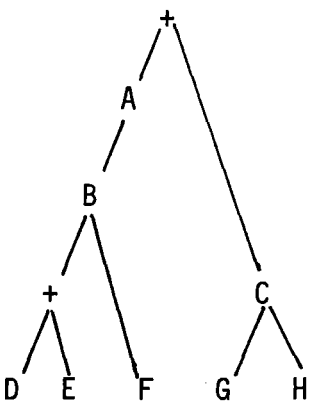

Table 4

Error levels

1	2	3	4	5	6	7	8
H		CGH					ABCDEFGH
G		CGH					ABCDEFGH
E			BDEF	ABDEF			ABCDEFGH
F	DE		BDEF	ABDEF			ABCDEFGH
D	DE		BDEF	ABDEF			ABCDEFGH

Table 5

Stemma

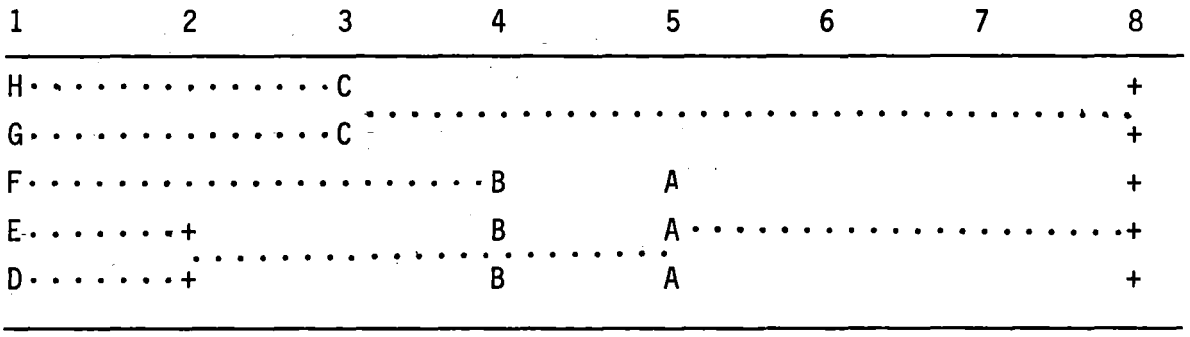

Table 6