NOTES ON THE INPUT FOR AN AUTOMATIC SCANSION PROGRAM

Ott in his admirable article on Metrical Analysis of latin Hexameter by Computer (*Revue* no. 4, 1966) said "the ideal would be to punch the text just as it is found in the editions without any added scansion ..." He considered this impracticable, however, and added the scansion pattern in the form of a six digit binary number after each line in the input. Challenged by the problem of automatic scansion Greenberg (*Revue* no. 1 & 3, 1967) and Dyer (*Revue* no. 4, 1967) devised ingenious programs which will successfully scan a very high percentage of lines in Vergil and Homer. For certain lines, however, both programs fail. Some of the failures the computer itself detects and marks but there are others which it overlooks.[1] Considering the type of error which the computer makes I do not believe it is feasible to devise a completely successful scansion program without some prior attention to the text. To reduce this attention to a minimum and at the same time to avoid the possibility of computer error, I should like to propose a different kind of program based on a different method of scanning hexameters.

The programs of Ott, Greenberg and Dyer follow the conventional method in which the length of each syllable is first determined and the line then divided into feet. As I have pointed out elsewhere ("On Reading

Extrait de la Revue (R.E.L.O.)
IV, 1 à 4, 1968. C.I.P.L. - Université de Liège - Tous droits réservés.

Hexameters Aloud", *Classical Outlook*, December, 1968), there is a great deal of redundancy in this method. In order to scan a line correctly, it is necessary to know only where the ictus is going to fall. As soon as this is known the metrical quantities of all syllables are automatically determined. And since the position of the first and last ictus is always the same, only four decisions need to be made. These decisions are based on the length of the syllable immediately following the previous ictus. If this "key syllable" is long, another ictus follows immediately; if short, it falls on the next but one.

For a program based on key syllables instead of complete scansion the computer would need only four bits of information in addition to what the text itself supplies. To prepare the text for cards or tape, long key syllables would be marked (e.g. with a slash "/") and short key syllables left unmarked. In addition, any vowel that combined or failed to combine with another vowel contrary to rule would be marked with a "+" or a "-".[2] With very simple programming, tapes prepared in this way could generate full scansion patterns, identify metrical word types, and provide the data for statistical treatment of various kinds.

For an error-free program all that the computer need know is what constitutes a syllable and how (in Latin) final vowels and *m*'s are treated before another word beginning with a vowel. It would not need to know whether a vowel were long or short, or whether a stop and a liquid did or did not give position. For scanning, the computer would be instructed to place a dash (-) under the vowel (the first vowel if in a diphtong) of the first syllable and to read the second syllable. The instructions would then be : if there is a slash, "/", before the vowel in the second syllable, place a dash "-" underneath it and underneath the vowel in the third syllable and read the next syllable; if there is not a

Extrait de la Revue (R.E.L.O.)
IV, 1 à 4, 1968. C.I.P.L. - Université de Liège - Tous droits réservés.

slash before the vowel in the second syllable place a full stop, ".", underneath it and under the vowel in the third syllable, place a dash under the vowel in the fourth, and read the next syllable; proceed in the same way until four decisions have been made. At this point there will be either three or four syllables left whose metrical length will be determined by their number, sc. if there are three, the pattern must be - - -; if four, . . - - . (If five syllables remain, either the line is hypermetric or an error has been made in the preparation of the text.)

The first lines of the *Iliad* and *Aeneid* would be written as follows :

```
MHNIN AEIDE JEA P/HLH-IADE+W ACILHOS3
  -  .   . -   -  . -   -   - . .      -  .  .  - -

ARMA VIRUMQUE CANO TR/OIAE QU/I PRIMUS AB ORIS
 -    .   .  -     .   .  -     - -       -    -   .   .   -  -
```

Except as a means of detecting errors in the input, the scansion pattern is of little interest, since it cannot be used for statistical operations. The scansion program, however, can be readily adapted to generate manipulable symbols. Elsewhere (TAPA 1966, 97, 275-280) I have suggested using octal numbers to identify the 32 scansion patterns of the Homeric hexameter. The octal number is obtained by representing each spondee by a one and each dactyl by a zero and converting the resulting six-digit binary into the corresponding octal. The code would be programmed by instructing the computer to print a one under each key syllable preceded by a slash, and a zero under all other key syllables; the last two digits would be either 1 1 or 0 1 depending on the number of syllables remaining :

Extrait de la Revue (R.E.L.O.)
IV, 1 à 4, 1968. C.I.P.L. - Université de Liège - Tous droits réservés.

```
MHNIN AEIDE JEA P/HLH-IADE+W ACILHOS
    0     0    1   0       0    1              = 11

ARMA VIRUMQUE CANO TR/OIAE QU/I PRIMUS AB ORIS
    0       0           1       1       0          1 = 15
```

The scansion program could equally well be adapted to generate McDonough's 9 digit code ("Homer, the Humanities, and IBM", Literary Data Processing Convention Proceedings, September, 1964, Modern Language Association, New York, pp. 25-36) and the "metrical word types". To represent a syllable at the beginning, middle, or end of a word, or in a monosyllabic word, McDonough used a 1, 2, 3 or 4 if the syllable was long, and a 5, 6, 7 or 8 if it was short. A 9 was used "wherever there could have been another syllable" i.e., after a spondee; so that each coded line would contain 18 digits. To code a line of hexameters with long key syllables marked, the computer would be instructed to count the syllables in each word and instead of a dash or full stop (as in the scansion program) to print the appropriate digit underneath each syllable, with a 9 intercalated after every pair of longs. For example :

```
MHNIN AEIDE JEA P/HLH-IADE+W ACILHOS
  1 7  52  7 53   19266    3 5 6 23   9
```

By means of a table McDonough's code can be translated into Porter's metrical word type numbers (MWT) to represent the "inner metric" of the hexameter. Each number identifies a particular combination of long and/or short syllables within a single word and the position in the line where it can be found. Dyer (op. cit.) lists 145 of these MWT's. The notation for *Iliad* 1.1 would be 7-79-49-106-13.

Extrait de la Revue (R.E.L.O.)
IV, 1 à 4, 1968. C.I.P.L. - Université de Liège - Tous droits réservés.

These metrical word type numbers are arbitrary and cannot be interpreted without reference to a table. I would prefer a more descriptive code with numbers which could be derived directly from the line itself. One way to achieve this would be to number the syllables in the line, list the numbers of the final syllables in each word but the last, and add the octal code number in parentheses, thus defining the inner as well as the outer metric. In this system *Iliad* 1.1 would be written 2-5-7-12 (11). (The last word is not included in the numbering since the information is already contained in the octal code number). Such a system would be easier to program than the other. It would serve for many statistical purposes, and could generate the MWT's if they were wanted.

This paper, however, is concerned with input rather than output. I have raised these questions only to illustrate the variety of operations that can be performed with an hexameter line after the long "key syllables" have been marked.[4] It is in the simplicity and economy of the method for printing and proofreading that the principal merit of the method lies.

Frank P. JONES
Tufts University

Extrait de la Revue (R.E.L.O.)
IV, 1 à 4, 1968. C.I.P.L. - Université de Liège - Tous droits réservés.

FOOTNOTES

1. Greenberg's program failed to scan 37 lines in *Aeneid I* and gave alternate scansions for 21 more. Of the 37 failures, however, all but three were detected by the computer. (Greenberg, 1967, 3, pp. 13-15). In Dyer's program (Dyer, 1967, 4, pp. 33-37) as implemented by David Packard the proportion of failures is higher, although most of them are due to synizesis or unresolved diphthongs. It might be possible to elaborate the programs so as to reduce the number even further but I do not think that failures and alternate scansions can be completely avoided without giving some prior attention to the text.

2. A "+" indicates synizesis, a "-" a resolved diphthong. This is the notation used by Ott (*Revue*, 1966, 4, pp. 9-10).

3. The transliteration follows G. P. Zarri's Table 1 (*Revue* no. 4, 1967, p. 70), where eta = H, theta = J, chi = C and omega = W.

4. It would be possible, of course, to write a program along these lines without marking the text in advance. This would involve giving the computer a set of instructions for distinguishing among syllables as definitely short, definitely long, and doubtful. The instructions then would be to examine the two syllables after the first ictus (both of which must be long or short); if either of the two was definitely long, to print a "I" underneath; if either was definitely short, to print a "0"; if both were doubtful to treat them as short. The decision would determine which pair of syllables to examine next. If in the process of examining the line two syllables were read, one of which was definitely long and

Extrait de la Revue (R.E.L.O.)
IV, 1 à 4, 1968. C.I.P.L. - Université de Liège - Tous droits réservés.

the other definitely short; or if the line itself proved to be a syllable too short, the computer would be instructed to return to the doubtful pair and treat them as long, then proceed as before. After the scansion was completed the six digits would be converted into a two-digit octal. Such a program would correctly scan a very high proportion of the hexameters in the *Iliad* and *Odyssey*. (It would scan all of *Iliad* 21 and *Odyssey* 9, for example). There are some lines, however, at which it would fail. Unless the possibility of computer error could be eliminated completely I should prefer to mark the input in some such way as I have described.

Extrait de la Revue (R.E.L.O.)
IV, 1 à 4, 1968. C.I.P.L. - Université de Liège - Tous droits réservés.