## Metrical Analysis of latin Hexameter by Computer


#### Abstract

The mass of work to be done in metrical analysis includes the collection of facts such as (for the hexameter): the number of dactyls and spondees per line; the position of the caesuras; the frequency and position of elisions; the types of words at the end of the lines, etc. A considerable portion of this work has been done up to now by patient manual labour. The great amount of work connected with these inquiries, however, has restricted the analysis to the works of some of the preclassical and classical poets. For similar reasons, some metrical characteristics that are interesting for the study of comparative history of metrics have never been analyzed on a larger scale.

Meter, as is commonly known, has to do with numbers and quantities; "numerus" was the ancient equivalent for what we call meter. The computer, therefore, seems to be destined for metrical analysis, at least as far as metrical analysis involves collection, comparison and statistics of large quantities of single observations.

The Philologisches Seminar of the University of Tübingen (Germany), in collaboration with the Computing Center of the University, has therefore developed a program to carry out automatically the part of the job just outlined. The program aims at reducing the preparatory work to a minimum and at obtaining a maximum of information. Because the Latin hexameter is relatively easy to handle, we have started to develop a program for this meter. Programs for other meters are to follow (the next being a program for the pentameter). Here we wish to give only preliminary general information. A more detailed description of the program for the hexameter


is to be published in the following number of Revue.

## TECHNICAL DETAILS

The program is a chain job, written for the IBM 7094, a computer used by German Universities at the Deutsches Rechenzentrum at Darmstadt. This machine has a core storage of 32.768 36-bit-words, 1 disk-storage IBM 1301, 13 magnetic tape units 729 IV (plus four for off-line input and output). Among the off-1ine output devices, the plotter IBM 1627 has proved to be useful for showing at a glance the distribution of single characteristics in a poem, in single works and in an author's work. The basic computer cycle time of the IBM 7094 is two microseconds.

The program is written for the most part in FORTRAN II (not in FORTRAN IV, first because FORTRAN II permits better utilization of the core storage, and secondly, because part of the subprograms that are available at the Deutsches Rechenzentrum are available up to now only for FORTRAN II programs). Thanks to the large software which has been developed at the Deutsches Rechenzentrum for non-numerical problems, we need only a little more than 2.500 FORTRAN II statements, plus some hand-coded subroutines written in FAP.

For input, we use 80-column punched cards, containing in col. 1-64 the text including, for the present, the punctuation as found in the editions. For signs such as semicolon or exclamation marks we punch a combination of two signs available on the IBM 26 card punch, e.g. ".," for ";" or "!." for "!"

For the capital letters in proper names, we punch an asterisk before the first letter of the word. In col. 67-72 we punch the scansion of the line (see below), in col. 75-76
the number of the book, and in col. 77-80 the number of the line. Col. 65 serves to exclude certain checking operations, such as the checking of line-sequence, where the editions number the lines differently. Col. 66 is used for the main caesura, if not coded in the prepared text by an "=l". Col. 74 is for variae lectiones.

## PREPARATION OF THE TEXT

As said before, the aim in setting up the program was to reduce the preparation of the text to a minimum. The ideal would be to punch the text just as it is found in the editions, without any added scansion and other signs. But the other aim we pursue is to secure reliability in the portion of the work that needs human intervention, that is, to exclude, as far as possible, the influence of punching errors occurring in the preparation of the input cards. Up to now, we have therefore preferred to leave some redundancy for the sake of better checking possibilities. Thus, the program requires the following preparations of the text:

1) in col. 67-72 the scansion of the line, consisting of 6 di gits, one for each foot: "0" for a spondee (and for the 6th foot), "1" for a dactyl;
2) distinction between "U" (vowel) and "V" (consonant), "I" (vowel) and "J" (consonant);
3) the hiatus must be marked by a "/" (slash) between the words;
4) asterisk before the proper names;
5) for some rare special cases of synizeses, we include the contracted vowels between special signs "+ . . .)", e. g.,

Ovid, Met. 6,635 • . IN CONJUGE TER+EO) where the final vowels "EO" are treated as one syllable. For the more frequent cases of synizeses, like in PROUT, QUOAD, DEORSUM, DEINDE, FUISSE, etc., the program decides automatically if the vowels in question are to be contracted for this verse or not. The opposite case is a word like CHAUCO (Claudian, de Cons. Stilichonis I,225 ... NON INDIGNANTE CHA-UCO), where the vowels "AU" are not to be interpreted as a diphthong, but as two separate vowels. In cases like this, there must be a "-" between the two vowels. Since these cases are extremely rare we prefer to mark these words when punching them instead of burdening the program with such relatively rare details.

## GHECKING THE INPUT CARDS

Before being analyzed and statistically evaluated, the prepared text is treated by the program for the purpose of revision. The first control tests whether or not the scansion given in col. 67-72 fits the number of syllables found in the text; this means: if there is a vowel missing or in excess, if the "/" for the hiatus is missing, or if a "V" is punched instead of a " $U$ " or vice versa, etc., or if the hyphen in a word like CHA-UCO is missing or if the signs " + 。 - .)" are missing in a word like TER+EO), this line is marked by asterisks. A number printed at the end of that line indicates how many syllables are missing or in excess. The same diagnosis is given if, though there is no mistake in the text, the scansion given on the cards presupposes that there is a number of syllables other than that actually found in the text.
These checks are also performed during the analysis phase of the program. For checking purposes, there is another operation. Depressing Sense Switch 3 on the operator's console causes the program to skip all the statistical evaluation
and interpretation operations of the analysis and to check the scansion given on the input cards. Based on the length of diphthongs and of vowels followed by two consonants other than muta cum liquida, and on the length or shortness of some frequent words and word-ending syllables, and on a few rules*, a second scansion independent of the scansion given on the input cards is made and compared with the original scansion. If no correspondence is found, the line is marked as wrong. For the small percentage of lines that cannot be completely scanned in this way without using a dictionary, only the shortness or length of the syllables scanned automatically is compared to the shortness or length of the corresponding syllables resulting from the scansion given on the input cards. For easier checking, both the scansion given by the punched card and the scansion independently computed by the program are transformed into dashes ("-") for long syllables and full stops (".") for short syllables and are printed in two lines above the corresponding vowels and diphthongs. Another check concerns the order of the cards, based on the sequence of line-numbers, if there is no punch in col. 65 (see above).

These checking operations are made at a rate of about 100 lines per minute, including output of both right and wrong lines together with their scansions and all metrical details of the single lines on magnetic tape. If one can be sure that there are no mistakes in the punched cards, this checking phase can be skipped.

## THE METRICAL ANALYSIS

As for the information to be obtained by the program, we took as a guide the famous Appendix of $E$. Norden's commentary on Aeneis VI and some of the literature quoted there (e. g., the tables published in 1911 by A. Siedow, De elisionis aphaeresis hiatus usu in hexametris latinis ab Ennii usque
ad Ovidii tempora, Diss. Greifswald 1911) and some other works, e. g., for the caesuras, the article by H. Fränkel, Der kallimacheische und der homerische Hexameter, Neu. Gött. Gel. Anz. 1926, 197ff., etc. In addition, since the machine is working at high speed and without shrinking from doing most tedious work, we easily obtain information about several further details such as the absolute and average number of letters, syllables and words in a single line or in a single word; the number of different types of word-positions in the lines**, etc.

First phase: Analysis of a single line.
As a first phase, the machine analyzes the text, card by card, and writes the text read with its analysis on the output tape and stores it at the same time in the disk storage for subsequent statistical evaluation. In appendix $I$, we present a sample analysis of a short poem, "Vir bonus", contained in the Appendix Vergiliana with, above each line, the scansion as described above. In the text itself, all "U"'s and "V"'s are printed as "V", all "I"'s and "J"'s as "I". The second half of the line gives the octal print of three 36-bit-words per line, each of which contains information about two characteristics of the line:

Left half of the first word: information about the caesuras. The first digit, an octal 5 -binary 101, indicates a caesura after the first and the third "unit" of the first foot. Each foot is interpreted as containing three "units", the first unit being the first long syllable of the foot, the second and third units being either the two following short syllables or the second long syllable of the foot. The second digit, an octal $4=$ binary 100 , indicates a caesura after the first syllable of the second foot (only the first of the three binary digits is a "1"), etc.

Right half of the first word: 6 digits, $a^{" 1 "}$ for a dactyl, a "0" for a spondee. The "4" occurring in lines 3 and 9 contains information about the monosyllabic words in elision (not input information, such as the scansion contained in this half of the word, but the result of the analysis), octal 4-binary 100 indicating a monosyllabic word in a spondaic foot (spondaic, because the last binary digit is a "0"; similar account for a dactylic foot is octal 5=binary 101) elided after the first syllable of this foot. For elided monosyllables coming after the second or the third syllable of a foot, the second binary digit is a "1". For further details, see the description of the program in the following number of Revue.

Left half of the second word: 6 digits, one per foot, containing information about the correspondence of word accent and ictus, "0" indicating that there is no correspondence in this foot, "1", that there is correspondence of ictus and word-accent of a monosyllabic word, "2" ("3", "4", . . .) indicating correspondence of ictus and word-accent of a word of $2(3,4$. . .) syllables.

Right half of the second word: information about elisions, "0" meaning "no elision in this foot", "1" (=binary 001) meaning "elision in the last syllable of this foot" (in order to see whether this last syllable is the second long syllable or the second short syllable of the foot, one examines the corresponding scansion digit in the right half of the first word); "2" (=binary 010) meaning "elision in the. first short syllable of the foot", etc. The same information about the hiatus is given in the right half of the third word.

Left half of the third word: information about the ends of phrases, computed on the basis of the punctuation marks found on the cards. On a card which is to be read before the cards containing the text, one can indicate which of the
punctuation marks are to be interpreted by the program as concluding a sentence; if this card is absent, as it is the case in the example given in appendix $I$, only a full stop ("。") is interpreted in this way.

The seven subsequent numbers, at the end of the line, represent the number of words, syllables and letters contained in the text, the number of elisions, of elided monosyllabic words, of correspondence between the end of a foot and a word, and the number of feet in which correspondence between ictus and word accent is found.

As said before, this information is simultaneously written on the output tape and stored in the disk storage. Only the three 36-bit-words per line of poetry just mentioned remain in the core storage, plus one machine word per line containing the current number of the line concerned (indicating the address of the processed text in the disk storage). In addition to the program, the core storage of the IBM 7094 can thus store the characteristics of 2.000 lines of poetry at the same time, enough for the statistical evaluation of the analysis of single books containing up to 2.000 1ines.

During this analysis, the text is segmented into words. These words are provided with all metrical details and are written in binary form on a magnetic tape for further treatment. If sense switch 5 on the operator ${ }^{\prime}$ s console is pressed, the machine also writes these words in BCD-form on an output tape which is used to punch off-line cards containing one word per card and the corresponding metrical information. As an example, a list of some of these cards is given in appendix II. The cards contain in col. 1 a conventional code for the rhythmical type of the corresponding line***. The next two digits indicate the beginning of the word in the line according to the following conventions:

```
\(\left.\begin{array}{cccccccccccccccccc}01 & 02 & 03 & 04 & 05 & 06 & 07 & 08 & 09 & 10 & 11 & 12 & 13 & 14 & 15 & 16 & 17 & (18) \\ (- & - & \cdot & - & - & \cdot & - & - & \cdot & - & - & \cdot & - & - & \cdot & - & .\end{array}\right)\)
```

(the numbers 3, 6, 9, 12, 15 occur only in dactylic feet).
The subsequent digit, coi. 4, indicates the length of the word in such "units". An addition of col, 2-3 and col. 4 therefore indicates the beginning of the following word. For the last word of a line, the result of this addition is 19; this indicates that the following word is the first word of the next line, since one line contains 18 "units".

Col. 5 contains a digit indicating the number of syllables in the word.

Col. 6 contains a hyphen ("-"), if there is an elision of the last (or only) syllable of this word; an ictus ("'"), if there is correspondence between word-accent and verse-ictus in this word; an asterisk, if both elision and correspondence of accent occur in the same word.

Col. 7-26 contain the form of the word as found in the text, col. 27-38 contain information about the length of single syllables and the position of the word in the line, plus information about the elision indicated in col. 6 by a hyphen or an asterisk. The letter "A" (for "arsis") indicates that the last syllable of a word is elided in the following first long syllable of the next foot; "T" (for "thesis unica"): elision in the second long syllable of the same foot; "T1" (for "thesis brevis prima"): elision in the first short syllable of a foot; "T2" (for "thesis brevis secunda"): elision in the second short syllable of a foot.

Co1. 39 contains an " $N$ ", if the word is a proper name (-as-
terisk in the text).
Col. 40 contains an "0", if the word is the last word in a line.

Col. 41 contains information about the main caesura of the line (left blank in our example, because the main caesura was not indicated on the input cards).

Col. 43-48 contain conventional codes, indicating author, work and edition (two digits each, either alphameric or numerical. In the example given, no indications of this kind were punched in the input card, so that six "0"'s appear in these columns).

Col. 49-50 contain the number of the book (blanks or 01, if the poem contains only one book), col. 51-54 contain the line number, col. 55-56 the number of the word within the line, col. 57-58 the number of the word within the sentence, col. 59-63 the number of the word in the poem, col. 64-65 the number of the letters in the word. The rest (col. 42 , 66-80) is free for further information.

As said before, identical information about each word is also stored in binary form on a magnetic tape (and, in a further step, in disk storage) for further inquiries and evaluations.

In addition to the cards just described, a second card is punched for each word, according to the conventions described in nr. 1/1966 of Revue; col. 2-18, 54-56 and 76-80 are left free by our program; by means of the program developed at the LASLA, the lemma, the grammatical analysis and the number of the word in the index can be added in these columns, as described in $n r$. 1/1966 and 2/1966.

The part of the program just described performs the follo-
wing operations: reading the text from magnetic input tape (prepared off-line on an IBM 1401 by means of the punched cards described above, at a rate of 800 cards per minute); analyzing the lines; output of the results of this analysis; storing the text and results in disk storage; decomposition of the text into words accompanied by the information described above, and storage in binary form on magnetic tape. 250 to 270 lines are processed in one minute. If the information necessary for punching the two cards per word is written in BCD-form on an output tape while the analysis is performed****, the rate is about $160-180$ lines a minute.

As mentioned above, the core storage can hold the results of the analysis of 2,000 lines at the same time. The disk storage can hold simultaneously the text and metrical details of 21.000 lines, plus $175.000 * * * * *$ words with the metrical details described above.

Second phase: evaluation of the analysis
As our preliminary description suggests, this analysis can be used to obtain various data, average values, statistics and tables. A more detailed description of what we have done up to now, and some examples, will be given in the following number of Revue. Here is only a short list:

- total and average values obtained from the metrical analysis. described above;
- a classification and complete synopsis of the different rhythmical types of lines (cf. Norden, op. cit., pp. 418 ff.) ;
- complete statistics on the caesuras (cf. the statistics in H. Fränkel, op. cit.), followed by the complete reference

```
    (i. e., line numbers);
```

- statistics on the combination of certain caesuras, with re-
ference (cf. Norden, pp. 425 ff.);
- a list of spondaic words in the first foot, with reference
(cf. Norden, pp. 435 ff.);
- a classification and complete synopsis of all occurring ty-
pes of line endings (cf. Norden 437 ff.);
- statistics and synopsis of the different types of lines ac-
cording to the various rhythms and caesuras ("word-position-
types");
- complete documentation about the occurrence of equal lines, half-lines and shorter epical formulas (if not found in identical metrical position in the line, of a length of at least two words; identical words in identical positions are listed completely);
- references about the letters occurring together at the end of one and at the beginning of the subsequent word;
- references about alliteration, homoioteleuta and rhyme。

Other, more special investigations can easily be made by using a short program on the basis of the punched word cards or of the same information recorded on the magnetic tape which we preserve for such purposes for each poem we analyze.

In addition, the machine produces graphs with the following information: number of words per line; number of correspondences between word endings and foot endings; frequency of elisions; length of words and syllables in individual lines;
correspondence between ictus and word-accent. In Appendix III an example is given of one of the two graphs of the correspondence between ictus and word-accent for "Vir bonus" and "Est et non", another short poem of the Appendix Vergi11ana******。

Up to now this program has enabled us to analyze some parts of the works of Vergil, Horatius, and Claudian. The first author to be analyzed completely will perhaps be Lucretius, De rerum natura. The text of this poem has been completely scanned and punched on cards at the LASLA, and we hope to analyze it soon. When we have finished our analysis of Vergil an Horatius, we will start to analyze the works of Statius. The program has been designed in such a way that the progress of our work only depends on the time needed for the scansion of the lines and for the punching of text and scansion on cards.*******
Universite de Tübingen
W. OTT

NOTES

* E。 go, if the possibility that the line is a hypermeter is* to be excluded, then the penultimate syllable of the line must be long.
** E. g., among the 385 lines of Claudian's first poem on Stia licho's consulship, 95 lines correspond to one or more other lines as to the position of the caesuras and as to the number of dactyls and spondees, whereas among the 476 lines of Horatius' Ars Poetica, there are only 30 correspondences.


## appendix I

 -




## 

## appendix IIIa




## appendix HIb

ESI ET NON CLNETI MONOSYIAFGA NOTA FREQUENTANT,

-     - REALE WETE
 - . VEPGFISS j+


