

# **AN INDEX TO CIL VI TWO ASPECTS**

## **I. preparation of the text**

Scholars, ancient historians in particular, have long awaited an index to volume VI of the Corpus Inscriptionum Latinarum. This volume contains almost 40.000 inscriptions originating in Rome and ranging in length from a single fragment of a single word to lengthy sacred records of more than 12.000 words. The first fascicle of this collection was published in 1876 and nearly one hundred years later manual methods of indexing, slow, laborious and in this instance subject to the hazards of war, have produced only an Index Nominum, published in 1926, to aid the student of the material.

Modern science in the form of the computer can now provide scholars with the full benefit of this monumental collection.\*

The scope of the present project must be defined at the outset. Its object is merely to provide investigators with information as to where in volume VI they can find answers to their queries. In short it is an Index to volume VI and not a replacement of it. Since this is so, no attempt has been made to bring the readings of the inscriptions up to date nor have inconsistencies in the editing of the inscriptions been "corrected".\*\* Where later fascicles of volume VI have given new readings these have been noted and internal con-

cordances provided.\*\*\* However no new readings from later collections such as ILS have been incorporated nor have concordances to such collections been added.

It was originally planned to produce an Index which would approximate to the classified indices of the remaining volumes of the Corpus. However it quickly became apparent that this, although practicable, was of limited value and did not fully exploit the possibilities opened up by the use of the computer available. Planning was then oriented towards the primary goal of a complete word-index. Preliminary results have shown the feasibility of such an index and it is hoped that it will be completed within two years. While the basic requirement is the word index and it is this which will be published in the first instance, the opportunity has been taken of including a great deal of ancillary information which will be of value to future historians, epigraphists and even palaeographers.

Firstly, as much information as possible has been recorded about the physical characteristics of the inscription. An indication is given of its position on the monument, left, centre or right, on the obverse, sides or reverse; whether there is any decoration such as palm leaves, ivy leaves, crowns etc; whether any words are written irregularly, backwards, perpendicular, upside down, in the form of a circle or even cursively; whether it contains elements of verse; whether it is written in a language other than Latin or Greek and whether it forms part of the unclassified collection of fragments. All this information follows the number of the inscription.\*\*\*\*

In recording the text of the inscription care has been taken to reproduce every peculiarity shown in CIL.\*\*\*\*\* A bare list of some of the main points may be informative. Distinguished are Greek letters, numerals, cognomina, words written "in litura", words still legible after deliberate erasu-

re, letters added by later hands, letters reported on first discovery but no longer extant, restorations, uncertain restorations, letters erased or missing, uncertain readings, variety in the forms of letters or numbers, Claudian letters, feminine indications for Gaia, puella and femina, tall letters, short letters, marks over letters or through them, abbreviations, ligatures, punctuation by point, ivy leaf or other marks and breaks in the inscription.

The reproduction of this information in machine readable form presented several problems. Those associated with computer limitation are discussed below but the difficulty of recording all the required information with a limited character set (63 characters) was the key problem in the development of the code. It was solved by extensive use of delimiters. For example, an asterisk is used to delimit cognomina and in order to distinguish Roman numerals from letters the former are bracketed by an equals sign. A dollar sign has been similarly used to show that the enclosed letters are Greek and not Latin.

In the application of the code the fragmentary nature of the material itself presented by far the most difficult problems. This can best be demonstrated by an example.

**EXAMPLE.** Procedure to be followed when a new line of an inscription begins unbroken with a word lacking an unknown number of letters at the beginning:

a) If there is a break at the end of the preceding line and the last word is not complete but is continued on the new line, type:

EXE+/M, . . .

b) If there is a break at the end of the preceding line and the last word is not complete but not part of the first word on the new line, type:

EXE+, @/WORD, . . .

c) If there is a break at the end of the preceding line although the last word is complete, type:

EXEMPLUM,+ @/WORD,.. .

The + signifies a break in the inscription, the / the end of a line, the , the end of a word and the @ an unknown number of missing letters.

Countless problems of a similar nature have presented themselves and while the solution of each has individually presented little difficulty the mass of detailed instruction and the need for absolute consistency in typing fragmentary and at times unintelligible text has been the determining factor in assessing how long the project will take.\*\*\*\*\*

Other problems unrelated to the coding have added to the intricacy of the work. For example many inscriptions are supplemented by Additamenta in later fascicles of the volume. How were these Additamenta to be incorporated? It seemed both unnecessary and uneconomic to type the inscriptions in full a second time. Automatic adoption of the later text was rendered impossible by the fact that while in the majority of cases the second edition is fuller or more correct, in some cases there is doubt as to whether the first or second edition is more accurate and occasionally the later readings are clearly inferior or even forged. This led to the adoption of the following complicated procedure for what at first sight appears a simple problem.

## **ADDITAMENTA**

Before typing the inscription check the no. in the relevant section of the Additamenta. If the no. occurs in the Additamenta the following procedure is to be adopted:

- A. When the two inscriptions are identical (under this include those nos. where no new readings are thought worthy of re-printing).
- 1) Type the later inscription with its no. and concordances e. g. :  
-30848=840-: abcde. . .
  - 2) For the earlier inscription merely type the concordance e. g. :  
-840=30848-
- B. When the later inscription gives a new and better reading or when it adds to the previous inscription
- 1) Type the later inscription with its no. and concordances e. g. :  
-32515=2375=2404-: abcde. . .
  - 2) For the earlier inscription merely type the concordance nos. e. g. :  
-2375=32515-  
-2404=32515-
- C. When the later reading is clearly false or less accurate
- 1) Type the earlier inscription in its usual place but include the later no. in the concordance e. g. :  
-214=30716-: abcde. . .
  - 2) For the later inscription merely type the concordance no. e. g. :  
-30716=214-

D. When there is doubt as to which is the better reading

Type both in their proper order but give concordance nos. in both cases e. g.:

-405=30757-: abcde. . .  
-30757=405-: abcde. . .

## **prospects for the future**

Once the index to CIL VI has been completed the possibility is envisaged of using the same code to produce similar indices to the remaining volumes. Nor need the work end there. Each year numerous inscriptions are being published, many in periodicals without wide circulation and difficult to procure. A comprehensive annual index to the inscriptions in such periodicals would be invaluable to all who study the ancient world. Such an index could be produced within three months of a set yearly terminal point, for example Dec. 31st. It only needs money.

## **II. machine indexing and information retrieval systems**

The steps taken in computer processing of indices such as those considered for CIL VI are coding (including manual annotation), data preparation and editing for machine input, processing and output in some suitable format. The absence (as yet) of devices which can "read" printed text makes the problem of data preparation non trivial. Even if such devices were available, the format of the CIL VI layout would

necessitate manual preparation as the programming of a computer to sort out the eccentricities of the inscriptions would be a frightening task.

Punch paper tape was preferred to cards because of the cost factor. Editing of paper tape can be a time-consuming task but the nature of the computer available (a DEC PDP-6) with remote console time-sharing facilities which allow its simultaneous (and hence economic) use by many users through teletypes has been of considerable help. After uncorrected paper tape has been transferred onto magnetic tape, corrections of punching errors are made directly onto the magnetic tape through the teletype. This is by means of a programme called the EDITOR which allows automatic access to any particular line of text. The use of paper tape will be bypassed altogether when the current computing system is expanded in 1967, as it will be then economic to type directly through the computer onto magnetic disc or tape via the consoles for long periods as well as for the relatively short periods needed for data correction.

The processing of the data can be considered as two separate problems. The first goal is to produce pre-assigned types of indices such as a word index, a cognomina index or a numeral index. The programming of routines to handle these problems is relatively straightforward and although difficulties have been encountered (particularly due to the bulk of data in question) these are technical and will not be discussed here. However, at all times attention has been given to generality in the preparation of computer programmes. For instance, although the code designed for CIL VI is not suitable for the text of a novel, essentially the same programmes are required for index preparation in each case. In both instances a preliminary table which indicates what symbols are to be considered as delimiters etc. may be optionally read in. Also considerable flexibility has been retained in the searching, sorting, merging and output routines developed and

under development for the present project, so that they may be easily linked together in many different ways.

This leads into a problem which is more interesting from the information processing viewpoint. Having so much information stored in a readily accessible form the research worker should be allowed to interrogate this data with a minimum of effort and without the specialist knowledge of the computer expert. For instance, if the research worker wanted a list of all inscriptions which contained any of a particular group of words or any other criteria, then he should be able to give simple commands in the language of his own field to allow such a search to be carried out. This problem is essentially no different from that which would occur in many other fields, for example, the medical research worker requesting information from suitably stored hospital records about patients with a particular type of medical history.

An area of research which is of interest to the indexer is that of the possibility of machine recognition of features of text that are in the current study marked manually. For example, is it possible to develop algorithms for the recognition of cognomina and if so what will be the probability of error from over and under recognition. The availability of a large text such as CIL VI with (presumably 100%)\*\*\*\*\* correct marking will allow tests of such algorithms to be carried out.

Of considerable influence on the format of output of the indices will be the cost of their printing. In all likelihood this printing will be by a photographic offset process direct from the computer print-out. The volume of material may influence the final decision as to whether a simple word index is to be produced or a KWIC type index\*\*\*\*\* which would include perhaps five words before and after the key word to provide a context.



The indexing of CIL VI is then part of a more general study in information storage and retrieval techniques, appropriate interrogation languages and their applications. In addition to index preparation the medical and library worlds appear to be particularly promising areas for these studies.

E. J. Jory & D. W. Moore,  
University of Western Australia.

## notes

- \* This research is being carried out at the University of Western Australia, is financed by grants from the Australian Research Grants Committee and the Australian Universities Commission and has the cooperation of the Deutsche Akademie der Wissenschaften zu Berlin.
- \*\* These inconsistencies may at first confuse the user of the Index. For example most inscriptions have been printed in the Corpus with restorations incorporated. In such cases account has been taken of the restorations in the Index. However, in some cases the inscription has been printed with no restorations and is followed by a fully restored example. In the latter cases the procedure has been to incorporate in the Index only the unsupplemented inscription. This procedure was adopted because many of the fully restored examples contain not only restorations of the original text but also interpretations of it, e. g. VI 2001, lines 4, 10, 16, 19, 20 etc. This practice has been extended to all the small number of inscriptions in which the supplements contain interpretations e. g. 32370, a later edition of 2070.

- \*\*\* No concordance has been given to other volumes of the Corpus in cases where the same inscription is printed twice.
- \*\*\*\* Because of frequent lack of information in vol. VI no record has been made of the material on which the text is found, cf. J. S. & A. E. Gordon, Contributions to the Palaeography of Latin Inscriptions, p. 65. Provision has been made for this to be incorporated in future Indices dealing with recently discovered inscriptions.
- \*\*\*\*\* Care will have to be exercised when analysing the information gained from the Index. The Corpus is the result of collections by many hands not all equally reliable. A warning has been given by J. S. & A. E. Gordon, Contributions to the Palaeography of Latin Inscriptions, p. 186 et passim.
- \*\*\*\*\* The expected early date of completion is almost entirely due to the enthusiasm, energy and diligence of the research assistant Mrs. H. Zaliki.
- \*\*\*\*\* This is an idealistic assumption and not a statement of God-like infallibility.
- \*\*\*\*\* Key Word in Context.

\* \* \*