

# LGPN

## MAKING THE BOOK

Beside the work of collecting and editing the material for this volume, there has been the separate task of creating the book itself. This has primarily involved integrating and giving a unified presentation to material which was the work of many minds and hands over a long period, and, in a decade when publishers became increasingly wary of handling large scholarly books, of accepting all the tasks of publication ourselves, up to and including typesetting. Here the role of the computer has been crucial – and tantalizing, holding out many new, undreamed of possibilities while at the same time withdrawing others which scholars have traditionally expected from their publishers. The ensuing relationship between scholarship and the machine, often uneasy, has not been static, having yielded, on one side, to our increasing awareness of, and wish to reflect, the complexities of our material, and, on the other, to changes of machine and technology. It has nonetheless endured, and the present book is the fruit of it.

Using computers to process information from the ancient world can induce a dislocation in the sense of time which makes the past history of the computing seem more ancient than the material that is being processed. Our early computing methods now seem – mercifully – remote: punching cards in a room reverberating with a metallic clatter, for submission to a machine only too prone to reject them because of some slight deviation from perfect straightness; everything punched in upper-case characters only, preceded by a multiplicity of codes, to mean 'Greek with the first letter upper case', 'Greek all lower case', 'English all lower case', and so on. Our present Codasyl-type IDMS network database will no doubt one day seem primitive too; certainly, the next Volume will be produced using a different system. It stands, for all that, as the means by which we came for the first time to master our material, and so to produce this book. The history of this progression merits some account here, partly because the decade in question has seen rapid developments in computing in the humanities which may be regarded as a piece of history, even of pre-history, in themselves, and other academic projects which began computing early may be helped to see that no-one need become fossilized in a particular system; but mainly because the form in which our data has been prepared and processed for computing at various stages has influenced many features of the presentation of the material in the book, and it will be helpful to the reader to have some understanding of the processes involved.

### 1. *The Early Computing System*

At the inception of the project the use of a computer was not yet envisaged, either for recording and storing the material, or for publishing it. The academic guidelines of the *Lexicon*, therefore, were laid down before the possibilities offered by the computer were realized. Computerization of material was introduced in 1975, as research of the first regions undertaken was being completed, and at a time when computing in the arts was becoming established at the Oxford University Computing Service. Under the guidance of Mrs. Susan Hockey, then the new Teaching Officer for Computing in the Arts, a member of the staff, Timothy Ashplant, began to punch the Peloponnesian material on to cards. The material was entered in single-line text-files containing four fields of information: Name (in translation), Place, Date, and Reference, this last including the information which appears in the final bracket [for a full statement of the contents of these fields, see LGPN I p.xii f, *Arrangement of Entries*]. Conceptually in 1975 we had not moved beyond regarding the computerized entry as an electronic version of the research-slip, more easily stored, searched, and sorted but, as we were later to realize, sharing many of the limitations of a card-file system. The four fields were marked off from one another by a symbol, but there were no structural divisions within any individual field, and no control at this stage over the content. That is to say, it was possible to make typographical errors which remained undetected by the system. It was intended that these files would ultimately be amalgamated by the Famulus package, and sorted into an order for publication based on name, then place, date, and reference. In anticipation of that stage, the single-line entries, once prepared, were

reformatted to six lines, and accents, places, and dates were given a code to enable the program to carry out the required sorting.<sup>1</sup>

The files were prepared on a regional basis, reflecting the character of much of our source-material, particularly the epigraphical material. This 'regionalism' is a mark of political and cultural forces at work in the modern no less than the ancient world: thus, Delphi and Delos have been the preserve of the French, under Greek licence, for the last one hundred years, and Cyrenaica and the Dodecanese that of the Italians between 1913 and 1945. Having characterized the discovery and publication of much of our material, and also the method of compilation of the *Lexicon*, it not unnaturally influenced the amalgamation of regions until publication became imminent. This was in a sense fortunate, since under our original computer system there was no possibility of our having a sufficient allocation of computer filestore to enable us to keep all of our files together; but it meant that we remained ignorant of the full dimensions of the material once integrated, which we have since found to be infinitely greater than the sum of its parts. The book as presented here made its first appearance, even to those who had worked long and hard on it, only in the first months of 1986.

The system outlined above remained in operation from 1975 to 1983, surviving a change of computer (from an ICL 1906A to an ICL 2988) and the arrival of such benefits as terminals, the availability of lower as well as upper case, and more sophisticated editing facilities, all of which improved our methods of working, but made no impact on our conceptual thinking. Publication in book form had been a desideratum from the start, though it remained unclear exactly how this would be achieved. Once the computer was in use the precise form of publication became, if anything, less clear, as the words 'microfiche' and 'graph-plotter' passed into, and fortunately out of, our vocabulary. Two developments brought this aspect of our work into focus, The Oxford University Computing Service was chosen as the site of the Computer Board's Monotype Lasercomp photo-typesetter, which was installed in the summer of 1980, while in 1979 the Oxford University Press had offered to publish the *Lexicon* to the highest standards of traditional dictionary publication; but in discussions about typography and page-design, and about the computer program which would typeset our material, we were brought up against our ignorance of the overall patterns of our data, and the relatively unstructured nature of our fields. How, for example, was the typesetting program to distinguish between a book-title, volume, inscription, or line number, a status, or a relationship? What were the frequency patterns of names, without knowledge of which the page could hardly be designed? In trying to answer these questions we were frustrated by the limits of the computer systems we were using. It was suggested that we would overcome these difficulties if we converted to a structured database.

## II. *Conversion to a Database*

So far we had had a collection of computerized data, but not a database. In computing terms, a database proper is a structured information storage system, one important feature of which is that information which recurs in identical form in more than one context is, nonetheless, stored only once. This is economical, and ensures consistency of spelling and format. Identifying this so-called 'constant' information, and considering whether there were items in our data, previously submerged in one of the four fields, which could usefully be given an independent classification, were among the first tasks demanded of us. The purpose of this initial thinking process was to draw up a 'conceptual schema', usually recorded in diagram form as a series of interconnected boxes, each representing a category of information in the data (Records), and their relation to each other (Sets). [This was a network database]. The aim was that the new data structure worked out in the schema would be able not only to sustain existing requirements of the data, that is, output the material in proper sorted order for the book, but also to open up new possibilities of examining elements of the material, such as variant name-forms, statuses, or relationships, which in the book are buried in the body of the entry. That the conceptual schema is not tied, for its implementation, to any particular machine gives this

---

<sup>1</sup> For a full statement of this early computing system, see T. G. Ashplant, 'The Lexicon of Greek Personal Names', in *Computers in Literary and Linguistic Research*, Proceedings of the VII International Symposium of the Association for Literary and Linguistic Computing, edd. L. Cignoni and C. Peters (Pisa, 1982).

exercise a lasting value, and is particularly useful for academic projects subject to periodic changes of computer, over which they often have little control or influence.

Given the chance to think again, this time with the fruits of ten years' work to guide us, we examined all the entries due to go into Volume I, to make this exercise as effective as possible. In the resulting *Lexicon* database, the information previously stored in the four fields is now stored in eighteen Records. All trace of the four fields has disappeared, and with it the notion of a single-line entry for each person. Instead, the basic unit is the Person Record, where the unique identification number, the publication number, date, status or occupation, sex, and homonymity of each entry are recorded. Although each Person is a unique individual, the name he bears, the island or town he comes from, and the book containing the evidence for him are not necessarily unique to him, but may be shared by many others. Accordingly, personal names, place-names and book-titles ('constant' information) are stored once only, in the Records Name, Island, Town, and Book. Each Person is linked to the appropriate name, place-name and book-title in these Records, via the intermediate Records of Namelink, Residence, and Reference, where the exact nature of the link, such as the fact that a name contains an orthographic variant, or that a town is an additional residence because of multiple citizenship, is recorded.<sup>2</sup>

This restructuring along more precise lines carried many benefits, some practical, some of more academic application. In the new integrated storage system, all the material of Volume I (and ultimately all the volumes) can be held on-line together. Volume I occupies 29 out of the 120 megabytes allocated to the database. Detailed editing of individual entries can take place without the whole region in which they occur having to be called up; regular updating has become a feasible operation. Searches of the entire material are speedy and efficient; if the program is asked to find all the inhabitants of the island of Kos, it does not have to look at each entry to see *whether* the residence is Kos, as under the old system; it goes straight to the Record island, finds Kos, and pulls out only those Persons attached to it. Statistics can be compiled with ease: for example, the 66,486 persons who make up this Volume share 16,254 names, 14,150 of them primary, and 81,383 references.

With tens of thousands of entries already computerized under the old system, it was clear that the parsing of the old-style entries for loading to the new structure would have to be done by a computer program, which would have to be 'taught' to recognize, by text content and not by position in the entry, the correct classification of the various items of information. By the time we had refined the various formulae permitted in our data, and built their recognition into the program, we had in effect a copy-editing program which would report mis-spellings and aberrant formulae. This program, appropriately named 'Monster', is now run from the earliest stages of data preparation and editing; of mis-spellings of place-names and book-titles have not been eliminated by the time entries are loaded, the database will reject them. Thus, proof-reading in the ordinary sense, data-validation in more technical jargon, is introduced into our procedures at a much earlier stage than before.

A further benefit, not yet fully exploited, derives from using the structure of the database to gain access to classified items of information which in the book are always subsumed under the entry. This can be illustrated by looking in detail at what has happened to personal names. Under the old system, every entry had a primary name in the name field, which would determine its position in the book. Some entries also had further information about the name recorded in the final bracket, such as editorial marks, orthographic or dialectal variants, or all the components of a multiple name [see LGPN I p.xii, *Arrangements of Entries, I. Names*]. In the database, *all* names are stored in the Record Name (still using a transliteration system), whether they occurred in the old name field, in which case

---

<sup>2</sup> The account of the database given here is intended to be, as far as possible, a non-technical description; as a classicist the author apologizes for her inability to resist the use of 'data' as a singular noun. For a fuller description of the *Lexicon* database see E. Matthews and S. Rahtz, 'Designing and Using a Database of Greek Personal Names', in *Computers in Literary and Linguistic Research*, Proceedings of the VIII International Symposium of the Association for Literary and Linguistic Computing (Nice, 1985); L.D. Burnard, 'Knowledge Base or Database? Computer Applications in Ethnology', in *Towards a Computer Ethnology*, edd. J. Raben, S. Sugita, and M. Kubo, *Senri Ethnological Studies* 20, National Museum of Ethnology (Osaka, Japan, 1987).

they are classified as 'primary', or in the old final bracket, in which case they are 'secondary'. This 'secondary' category, however, is further subdivided and classified to reflect the various reasons why names are recorded in the final bracket: 'editorial', 'orthographic', 'dialectal', 'carrying a comment' ('patr. adj.', '—lap', etc.), 'multiple name', with further divisions into *tria nomina*, *ὁ καί*, *ὁ ἐπικαλ.*, etc.. Each of these classifications is easily accessible in its own right, and can be output and examined. In short, the computer is no longer simply a means of data manipulation; it has become an analytical tool which enhances the value of *Lexicon* material for research purposes.

### III. *Creating the Book*

In the database the data is no longer in any normal sense 'readable', but this 'illegibility' is concealed from the user by the output programs, which reconstitute the data to familiar forms. The data can of course be output in many formats, but that most commonly used is almost identical to the research-slip and original computer text-file, and is the basis of the presentation of material in book form; for this reason, the old language of 'fields' has remained in our vocabulary. Output can be 'normal', that is, normal computer print-out, or with embedded typesetting commands, ready for printing by the Lasercomp. When the program is outputting the book, however, it does more than simply reconstitute the data it finds; it makes judgements which enable it, in effect, to 'write' the book.

The entries are sorted first in alphabetical order of name; the name will be output once only as a heading, with all the instances subsumed under it, each of which is assigned a publication number printed in bold before the entry. (This number is quite different from the identification number of the entry in the database, and, unlike that number, which is permanent, is revised whenever the book is output, to take account of addenda.) Within all the instances of a particular name, sorting then takes place according to alphabetical order of place – an island or region, a town and deme where appropriate. On the page, the island or region will be printed first, the town on the next line with an indentation, and the deme on a further line, with a double indentation. Within instances of the same name from the same place, the entries are organized on a chronological basis, and if an entry has the same date as its predecessor (within island, town and deme groupings, but not across them), a swung-dash is substituted for the date.

The next basis of sorting is by reference. Greek titles are given in transliteration to allow all books to fit into one sorted alphabetical order. Beyond reference, nothing is taken into account for sorting except homonymy. If a father and son, I and II, have identical references, then the father will appear first; if not, then each will appear in his proper reference-sorted position. Having made a final adjustment to take account of reference, the program then operates an 'ibidem' system, again within but not across regional divisions (except, illogically, that it operates across demes). That is to say, as much of the reference as is identical with the last reference of the preceding entry (where appropriate, all of it) is replaced by 'ib.'. If the only difference between the two is the line number, then, for example, 'ib.1. 9' is output. [In later volumes, 'ib' operates across name blocks.]

#### *Guidelines for the Reader*

Some features of the presentation of the material in book form require further explanation: they result from an inevitable mixture of policy and pragmatism.

##### 1. *Names*

Names containing orthographic variants, besides being stored in their normalized 'primary' form, are also stored in the database in full in their variant form, with the variant syllable recorded separately. On the printed page, the syllable only is output, except in the following circumstances:

- (a) if the name is part of a multiple name, e.g. *Οὐλπ. Νοβία Βασιλόκληρα* [*Βασιλόκλεια* (3)],
- (b) if it contains more than one variant syllable, e.g. *Καλλείνικο* [Καλλίνικο] (10),
- (c) if the syllable contains more than the seven letters which we (imprudently) set as the limit for this category of information when we were designing the database.

If a name appears in the final bracket abbreviated to one letter, the reader can assume that it is identical with the main name-form. This occurs most commonly with *tria nomina* and other multiple names. We must apologise for not having taught our program that names beginning with a diphthong should be abbreviated to the diphthong: so *Εὐξένο* for example, is abbreviated to 'E.', and not to *Εὐ*.

## 2. Dates

Contributors were allowed a fairly generous degree of freedom in dating-formulae. These could be exact dates, span-dates, century-dates, period-dates, or a totally non-committal '?'. In order to be able to sort this varied range of dates, the database loading program 'interprets' the date, as well as storing it in the formula it finds in the file, and the output program converts it into yet another version for the printed page. This interpreted date, which is linked into a notional chronological scale running from 999 BC to 999 AD, is the basis of the sorting of entries by chronology for the book. The starting date is the decisive one, but where two starting dates are identical the finishing date is also taken into account; perhaps rather illogically, *the later finishing date has a higher priority in chronological sorting than the earlier*. thus, 'iii BC' comes out before 'f. iii BC', '?ii BC' before '195 BC'. 'imp(erial)' before 'i AD', and 'inc.' before everything. The following list gives some idea of the range of dating-formulae, and their meanings:

| <i>Date in input file</i> | <i>Interpreted date</i> | <i>Date on printed page</i> |
|---------------------------|-------------------------|-----------------------------|
| 150BC                     | 150-150BC               | 150 BC                      |
| 6B                        | 599-500BC               | vi BC                       |
| F4B                       | 399-350BC               | f. iii BC                   |
| M2B                       | 175-125BC               | m. ii BC                    |
| S1A                       | 50-99AD                 | s. i AD                     |
| 4/3B                      | 325-275BC               | iv/iii BC                   |
| 4-3B                      | 399-200BC               | iv-iii BC                   |
| B(efore)212BC             | 262-212BC               | b. 212 BC                   |
| A(fter)100AD              | 100-150AD               | a. 100 AD                   |
| A(rchaic)                 | 999-480BC               | arch.                       |
| H(ellenistic)             | 323-31BC                | hell.                       |
| I(mperial)                | 31BC-310AD              | imp.                        |
| B(yzantine)               | 311-700AD               | byz.                        |
| ?                         | 999BC-700AD             | inc.                        |

## 3. References

When references are loaded into the database, they are all stored in full, parsed into book-title, volume where appropriate, text of reference, and line number; for example, *AJPh/34(1913)/p.455 no. 7(a)/45* and *AJPh/34(1913)/p.455 no. 7(b)/22*. In matching the reference of one entry against its predecessor to decide whether an 'ib.' is needed, the output program matches block against corresponding block, but is not able to differentiate among the contents of a block. This can create apparently unnecessary repetition of items: to take the above example, '*AJPh/34(1913)/p.455 no. 7(a), 45*' followed by 'ib. p.455 no. 7(b), 22'. The program is always logical and consistent about this, to a fault, but we have been tempted to suppress the repetitions in extreme cases for the sake of the appearance of the book, and will, therefore, have introduced some inconsistency.

In operating the 'ib.' system, the program looks back only to the immediately preceding reference, so that two or three references recurring in identical sequence in several succeeding entries produce no 'ib.' at all. Conspicuous examples of this can be found s.v. *σο* *σί*, *λιτο*

#### 4. Homonymity

Our editorial policy is to use I, II, III, etc. only for parent-to-child homonymity. The program, however, insists on outputting a person's homonymity number, even where the homonymity is not in relation to the current entry: thus, [Ἰόργη]ν (7) is son of 'Ἀριστέϊδα] II. Since this might aid the reader trying to identify individuals, we decided to let this departure from our intended practice survive; but we emphasize that the program was operating on its own in making these judgements.

#### 5. Relationships

In the exceptional circumstances of a parent belonging to a different location from his child, the place is given in brackets after the relationship in the child's entry, e.g. 's. [Ἰ]ν (Samos)'. In book format, a '?=' relationship is recorded by a cross-reference to the entry concerned by means of the publication number, which operates only within individual names. If two people attested as bearing different names are thought possibly to be identical, we have been obliged to subsume both entries under one of the names, and have recorded the divergent name in the final bracket of the entry.

IV. For the production of the book, all the data, sorted as indicated above, is output complete with embedded typesetting commands. The structure of the database, which enforced consistency of presentation because it would not accept non-conforming data, is also the basis of typesetting with eleven different font-calls, since the appropriateness of each font is determined by the position of the data in the database. What in input is a simple hyphen will emerge as em-rule, en-rule, or hyphen, according to context. A Greek name will necessarily carry a Greek font-call because of its classification in the database as a Greek name.

The text of the book as published is, in one sense, ephemeral. It exists only when it has been generated by the output program. The data which underlies it may be edited, added to, or subtracted from, but, these minor adjustments aside, it remains constant. Any one of several possible books could be programmed out of it – or indeed no book at all. We continue, however, to believe in the value of the printed book; one does not browse through a database; the database comes into its own *after* the user has formulated the questions he wishes to ask. These points are more than abstract musings. The recent history of our computing shows that flexibility of data editing, manipulation, and searching, and the needs of a relatively static end-product, a printed book, are not incompatible. Equally, for the book itself options remain open and flexible. Typography has not become frozen by decisions taken years ago: 's.' could become 'son of', 'iii BC' could become '3 BC' by the simple adjustment of a program.

Without the Computer Board's Lasercomp, publication of this book to these standards would not have been possible. At the risk of seeming ungrateful, however, we must draw attention to one important feature not available on it, the ability to produce perfectly justified pages. For this facility we have to express our appreciation of Microset, software developed by M.I. Clarke of Bradley Computing, which enables the user to intervene between formatted text and the Lasercomp, and set up pagination. This software also enabled us to hyphenate in all our languages, including Greek, and so to avoid many (but not all) of the widely spaced lines which can be such a conspicuous feature of computer typesetting. In fairness to all concerned it should be said that Microset became available for three-columned output only shortly before we were due to go to press; there followed a rapid modification of both our programs and the software, and an intensive learning of new techniques by staff, and the whole book was paginated in five weeks. It would be a fair statement that, though we think the end-product very pleasing, neither the potential of the software nor the skills of the staff were fully exploited. Hyphenation, in particular, was a skill learnt on the job; our principle was to follow *Hart's Rules* – except where special circumstances dictated that we should not.

As in the compilation of the book, so in its production, this has been a team effort. Every member of staff who, over the years, has declined to continue to use a particular machine or a particular program if another was available which would do the job more efficiently, has contributed to the group attitude which says that the computer must serve the needs of scholarship, and not vice-versa. Nonetheless, individuals must be picked out. At Oxford University Computing Service, Susan Hockey and

Catherine Griffin made important contributions in the earlier stages, and have continued to show an interest in the project. Lou Burnard and Paul Salotti set up the database, and the latter has continued to play a consultative role. Jonathon Moffett, the computer programmer for the British Academy and the Ashmolean Museum, has modified our programs in the run-up to publication. On the staff, Dinah Verman typeset the Preface, and took the lion's share of paginating the book, assisted by Jessamy Richards who also worked gallantly on an intractable list of Abbreviations. Above all, however, we have to thank Sebastian Rahtz, now in the Department of Electronics and Computer Science at Southampton University (now Information Manager at OUCS), who joined the staff in 1982 to key in data, and conveniently grew into a programmer under our very noses. His inventive intelligence led him to prod at the sleeping Monster of *Lexicon* material until he got it on the move. He worked on the database design, wrote programs for loading, interrogating, and outputting from the database, and continued to respond to calls for help after he had moved to full-time employment elsewhere. That Volume I of the *Lexicon* is a lively beast, responding to the goads of new discoveries and ideas all the way to the typesetter, and beyond, is the legacy of his work.

E.M.