THE EBLA ELECTRONIC CORPUS: GRAPHEMIC ANALYSIS[1]

Giorgio Buccellati
University of California, Los Angeles

It seems especially fitting for this occasion, which brings us together to inaugurate the new Museum where the material remains of Ebla are stored, to propose another inauguration of sorts, that of the *Ebla Electronic Corpus*. As projected for the near future, this edition will contain four disks, one for each of the four volumes *ARET 1-4*, a disk of programs, and two *User's Manuals*. They are part of a series published by Undena under the name *Cybernetica Mesopotamica*.

The purpose of my paper today is to illustrate some of the salient points of this edition. What we are offering is in fact much more than a simple "electronic dump" of the printed edition of the texts. In other words, we have not simply entered the printed data on the computer, as originally published in the *editio princeps*. Rather, the data have been reworked thoroughly, in such a way that we can speak of a new edition of the texts not only because of the format in which they are distributed (i. e. the electronic medium), but also because of the philological re-elaboration. It is this aspect of the project that I would like especially to illustrate in my talk today. First, however, I wish to remember a few major stages in the history of the project, and to acknowledge thereby individuals and institutions that have contributed to making this event possible.

---

[1]This paper is the second in a series of "Studies in Ebla Graphemics," of which the first has appeared in *Studi Eblaiti* 5 (1982) 39-74 (also in Russian translation in *Drevnia Ebla*, Moscow 1985, 114-132). The paper reproduces the substance of the text as delivered at the meeting in Idlib, and deals primarily with matters of data structure and file format; for the treatment of specific substantive aspects one may refer to the article by J.H.Platt and J.M.Pagan, "Orthography and Onomastics: Computer Applications in Ebla Language Study," published elsewhere in this volume. – For a fuller presentation of the project, the reader is referred to the two *Manuals* which are forthcoming in *Cybernetica Mesopotamica*: G. Buccellati, *Cybernetica Mesopotamica: General Introduction and Graphemic Analysis of Cuneiform Texts*, and J.H.Platt, J.M.Platt et al., *The Ebla Electronic Corpus*. – Those interested in receiving, when available, the disks containing the electronic version of the data and the programs should write to Mr. James H. Platt, Assistant Editor, Cybernetica Mesopotamica, IIMAS, POB 787, Malibu, CA 90265, U.S.A.

1. Background.

It is fitting that I remember first the Directorate General of Antiquities and Museums, represented by all our friends who have also organized this Conference, and who have consistently and steadfastly supported the program of a quick and full publication of the texts of Ebla. Today's occasion is a fitting reminder of how uniquely supportive the Directorate has been in providing total availability for the study of the materials, to the point of establishing a new Museum to house this all important collection. Over the years, we have been spoiled by the unfailing Syrian openness towards scholarly enterprise, but should not be taking it for granted. I, personally, do not have to look far for some sober reminders: you may find it ironic that at the same time that the American press was fabricating stories of alleged lack of freedom for scholars in Syria, some Museums in California were forced, by local political intervention, to destroy the Museum labels on important historical collections and to rebury these collections in an amorphous dump. In this light, I view the announcemenet, here in Idlib, of the publication of the *Ebla Electronic Corpus* as one more piece of evidence to support our claim that Syria has continued to provide, throughout these years, an ideal setting for untrammeled scholarly enterprise.

The "Missione archeologica italiana in Siria" is also involved in a very special way in this event. This is so because the *Ebla Electronic Corpus* appears under the formal sponsorship of the "Missione" and as one of its official publications. The Director of the Expedition, Paolo Matthiae, has been supportive of the project since I first proposed it on the occasion of the initial meeting of the "International Committee for the Publication of the Texts of Ebla," which was held here in Syria just over ten years ago: one must remember that back then, and until recently, the notion of using computers in the humanities, let alone in Assyriology, was considered, benevolently at best, as a hobby to be condoned, rather than a serious method to be exploited. For my part, I had been working on electronic applications to cuneiform as far back as 1968, so I had already begun to tap the immense potential of this new tool, and when the International Committee began to plan for the publication of the texts of Ebla I assumed, with the willing cooperation of Matthiae and the other members of the Committee, the task of computerizing the epigraphic data on behalf of the Committee. The *Ebla Electronic Corpus* is now the fulfillment of that early commitment.

During the first five years (from 1978 to 1983) actual work on a computerization of the Ebla texts was however slow, partly because we had to wait for the ini-

tial publication of the *editio princeps* in standard book format, and partly because the electronic medium itself was undergoing major transformations – remember, in particular, that in was only in the early 80'es that microcomputers came on the market. It was during this period that Alfonso Archi became the chief epigrapher of the Expedition, and he turned out to be a key factor in the implementation of our plans. The philological merit of the electronic edition rests in part on the fact that it includes a number of new readings based not only on a new understanding of the published texts, but also on collations from the tablets. This he has provided eagerly and systematically, so that our Corpus contains a number of significant, substantial improvements on the printed version which reflect his work in Aleppo before and, now, in Idlib. Since one advantage of the electronic edition is that updates are possible much more frequently and inexpensively than with the printed editions, we expect the *Electronic Corpus* to provide a signal service in this respect as well.

The second main phase of the project, which occupied the last five years, saw the transfer of the techniques and programs which we had developed on mainframe and mini-computers, to microcomputers, which were only then beginning to hold sway. As we were becoming weaned from the large logistic and administrative apparatus which surrounded the larger computer systems, we were gaining a measure of greater control on the computer as a tool which made its intellectual impact on the very conceptualization of the data more real and widespread. A fundamental need arose for the project, namely that of a more direct coordination between our staff in Los Angeles and the Expedition in Rome, pertaining not only to the philological, but also to the electronic dimension. For this task we were fortunate to be able to rely, just at that point in time, on Lucio Milano. During repeated stays in Los Angeles, where he also gave formal courses on Ebla, he contributed in a substantial way to the formalization of our project, in particular the philological and graphemic aspects. Back in Rome, he ensured the smooth functioning of logistic concerns, from the transfer of electronic data and programs to the coordination of the philological discussion.

During this same second phase of the project (again from about 1984 to the present) my former student John Hayes and three of my current students at UCLA took on a central role in the implementation of the project. In the early stages of the project, Hayes helped to supervise the work on the texts, and then James Platt came to serve as the main coordinator, in addition to concentrate on matters of graphemics; Joseph Pagan concentrated on both graphemics and onomastics, and Mark Arrington on onomastics. It was their joint task to establish

the text for the electronic edition, which is why the disks containing the first four volumes of *ARET* are published under their names as "authors." Archi and Milano followed very closely the philological aspect, while I retained the responsibility for the final codification of the graphemic and linguistic conventions, as well as for the inner logical structure of the overall system.

With the publication of the *Ebla Electronic Corpus* the second phase of our project is coming to a close, and we are on the threshold of the third phase. We will of course continue with the computerization of the Ebla corpus: the next volumes of *ARET* are in advanced state of elaboration, and in fact the disk of *ARET* 9, by Lucio Milano, will appear shortly *together* with the publication of the volume in printed format. But beyond issuing the data as such, we are working on the analysis portion of the project. Having taken such care to enter the data according to well defined and highly structured criteria, we can begin to reap the benefits of our efforts. We can, as it were, *talk* to the texts because we have lent them a special voice with which they may respond to us. Currently, we are working on both graphemics and onomastics, some aspects of which will be addressed in the paper by Platt and Pagan presented separately at this Conference.

This brief history of the project would not be complete without an acknowledgment of the sources which contributed to finance our research. IIMAS – The International Institute for Mesopotamian Area Studies, has been instrumental in raising the funds necessary. The two major funding agencies have been the Ambassador International Cultural Foundation, which has assisted especially with equipment and software, and the Packard Humanities Institute, which has contributed the major portion of the costs for data entry. Significant assistance for specific aspects of the research has also been provided by the Research Committee of the Academic Senate of the University of California, Los Angeles.

2. Aspects of the Electronic Edition – Data.

This sketch of the history of our work has already pointed at some of the fundamental substantive aspects which characterize the *Ebla Electronic Corpus*. Together, they lend the project a disctinctive physiognomy whose significance goes even beyond the realm of Ebla studies. The concept of the series within which the Corpus appears, i. e. the series *Cybernetica Mesopotamica*, is such that the texts acquire, as it were, a dual documentary value. Electronic data bases are not static. They are dynamic in the specific sense that data are not only stored – they are manipulated, conceptually, independently of our human, scholarly, brains. If

you think about it, what writing did to memory, computers are doing to writing. Some five thousand years ago, in this very land, the first scribes provided a whole new crystallization of mental processes, one which gave to internal, logical brain functions an outer existence, disembodied from the brain. Memory acquired an extrasomatic existence of its own, the mental world of man becamse accessible as a physical object apart from the mind − a clay tablet.

In comparison to writing, electronic data processing may be described as the extra-somatic extension of *active* logical brain functions. Data are not only stored in dormant state until a human brain touches and awakens them (i. e., "reads" them). In a computer, data interact with each other and continuously realign themselves according to trajectories which have, as it were, a life and direction of their own. With computers, we have succeeded in projecting to the outside our logical inner functions in their active mode, and we can look at them crystallized in their new, awesome form.

This is to say that the *Ebla Electronic Corpus* is much more than a different version of the written word. It is difficult to show adequatley where the difference lies without being able to actually use the computer for a demonstration. It is for this reason that only the actual disks will truly speak for themselves when accessed on the machine. In our present context, I will refer to a few sample printouts, which are like frozen frames of a moving universe, and as such can offer only a limited idea of the full potential of the system.[2]

The first two printouts give the "directory," i. e. the table of contents of the distribution disks. The directory of the disk containing the texts (Fig. 1) begins with a long list of introductory "files," i. e. chapters, which refer both to the system as a whole and to the specific body of data contained in this disk. The data themselves are given in a single, long file at the end of the disk − in this sample, the data are the texts of *ARET* 1, given with the sequential number of disks of the system *Cybernetica Mesopotamica* within which the *Ebla Electronic Corpus* is being published. The data are stored in a transliteration system which is very close to traditional Assyriological usage, although there are a number of additions and variations which aim at rendering with greater precision the complexities of the graphemic aspects of the writing system. These are explained in the file labeled -E-TG, i. e. the encoding manual for the data rendered graphemically. − In addition to specific codes for the graphemic details, another important feature of an electronic edition is that pertaining to data structure, i. e. to the

---

[2]the data utilized in te printouts are derived from the elctronic edition prepared by J.H.Plat, J.M.Pagan and M.A.Arrington, for which see above, N. 1.

specific format in which the file appears; this is outlined in the file -F-TG. And so on for the various other files.

On the philological side, important characteristics of the electronic edition include the harmonizations among various readings (listed in the file -D-HARMO), which allow to cross-reference in a unitary fashion texts published at different times by different authors; or again the list of unknown signs (file -D-SIGNS), which lists in a systematic and uniform manner all the signs found across the various volumes of *ARET* for which a value is not yet known; or again the textual notes, which explain a number of deviation from the *editio princeps*, due to collations or the like.

The printout given in Fig. 2 summarizes the general configuration of the different kinds of data files, of which there are two major types. The first type includes a **single** text. Externally, it is characterized by the extension .G (with volume and text reference given only at the beginning of the file). The second type includes a **bundle** of individual texts. This type is further subdivided into two types, depending on whether the volume and text reference is given on each line (.GL) or only at the beginning of the text (.GT).

Each one of these types may in turn occur in three alternate configurations:

(1) with suffix ^, the diacritics appear as unresolved double AS-CII characters of the lower sequence (in practice, characters which appear on the keyboard, e. g. *s*^);

(2) without any suffix (simple type), the diacritics appear resolved as simple, high ASCII characters of the higher sequence (i.e., characters which do not appear on the keyboard, e. g. *σ*);

(3) with suffix ' the files are formatted with codes suitable for typesetting. Besides the proper rendering of the diacritics (e. g. *š*), this format includes commands that allow a word processor to recognize how to print italics, bold face, subscripts, small caps and the like.

3. Aspects of the Electronic Edition – Analysis.

Fig. 3 reproduces the directory of the disk that contains the programs applicable to the data. The first set of programs deals with the handling of files; they are, as it were, convenience programs which allow to pass from one to another of the formats illustrated in Fig. 2. For instance: the primary transliteration

system as given in the distribution disks uses standard keyboard characters, or combinations of characters, to represent the various diacritics (file type .ɢ^). Thus a shin is rendered by two characters, an *s* followed by a circumflex (s^). In this way the texts as distributed have truly universal application, and may be read immediately by any computer system. They are, as one says, fully "portable." But if one wishes to see these characters represented by a special single character on the screen (σ) or on a printer ($\check{s}$), one can run the files through the -ᴘ-ᴀꜱᴄɪɪ or the -ᴘ-ʟᴀꜱᴇʀ programs, respectively, and then the shin will appear as a single character.

The major program, however, is the one that produces various types of indices which sort lexical items in graphemic format in a number of different ways. The data are drawn from full corpora or sub-corpora, defined according to specific criteria. For instance, one may have individual signs or sign triads, or proper names from texts of a certain archive or a certain date. Indices are produced with items in context or in isolation, in alphabetical order or by frequency, by text occurrence or by item. A simple amount of statistical computation is provided in the form of frequency counts. A complete listing of the indices produced by the program -ᴘᴛ-ɢ1 is shown on Fig. 4. The following figures illustrate the fortmat detail of some of these indices, the data being drawn mostly from the first four volumes of *ARET*. The printouts are exactly as derived by the program, after ·the output has been processed by the program -ᴘ-ʟᴀꜱᴇʀ; the only editorial change introduced is that certain portions of the index have been expunged (and are marked by dots within brackets) in order to combine on one page some of the various features of the indices.

Fig. 5 gives a listing of signs subdivided into graphemic categories (such as logograms and phonograms), and within these they are sorted by alphabetical sequence. The program distinguishes between certain graphic peculiarities, for which special notations have been introduced (these are not explained here, but a full description is provided in the *User's Manuals* and in the file -ᴇ-ᴛɢ).

Fig. 6 gives a listing of personal names with occurrences, always in graphemic format, sorted alphabetically. When a name occurs in some variation from the form given as template (e.g. *zu2-[ba]* for *zu2-ba*), it is so indicated on the right. The prefix *p* stands for cases where the sex of the name-bearer is not known (otherwise one finds *m* or *f*); the totals are given separately under the heading. For entries which total more than four occurrences, the total is given explicitly at the end of the entry.

Fig. 7 lists the divine names as items, in alphabetical sequence, with total for each individual item (and without occurrences).

Fig. 8 gives a listing by decreasing order of frequency of the words other than digits and names. The total number of occurrences is given (17316), the total number of lexical items (1961), and then the individual entries with total number of occurrences, percentages of the whole corpus, and relative histograms.

The same format is found in Fig. 9 for geographical names, and a comparison of the two pages shows that there is more repetition in the index of words than in the listing of geographical names, the frequent words being, predictably, the ones which refer to textiles.

*   *   *

A differentiated study of outputs like these allows to reach significant conclusions about cooccurrence, covariation, non-occurrence of given phenomena, with a speed and accuracy that remains practically impossible without the computer. And it is precisely on such observations of facts that we can base a distributional type of analysis which is the safest means of attributing meaning to the data.

```
            -C-FILES
            ------------------------------------------
            FILES ON CM DIRECTORIES
            (April 12, 1990)

            CM directories may contain any one of the following files
            (but note that only a few will be contained on any one disk)


!.EXE       gives first orientation and reads introductory files

--READ.ME   gives first orientation

-C-         (prefix for general files about Cybernetica Mesopotamica)
-C--PREF    brief preface to system as a whole
-C-DISKS    system of disks and volumes
-C-FILES    describes introductory files (= this file)
-C-INTR     describes Cybernetica Mesopotamica as a system
-C-SIGNS    cuneiform signs with unknown reading, not in standard lists
-C-TITLS    catalog of disk titles in Cybernetica Mesopotamica
-C-UTIL     utility programs available for further use of data bases

-D-         (prefix for files dealing with data on a given disk)
-D--PREF    brief preface to data base on disk
-D--VERS    synopsis of characteristics of version on disk
-D-ACKNO    acknowledgments
-D-AUTHR    authorship and copyright data
-D-BIBL     list of references sorted in bibliographical order
-D-CATEG    identification of data items by provenience, date, type, etc.
-D-EPIGR    documents as artifacts (field numbers, archaeological settings)
-D-HARMO    harmonization principles and changes from established edition
-D-INTRO    general introduction to data base included in disk
-D-REF      list of references sorted in the order of the data
-D-REFCO    concordance between data files and references

-E-         (prefix for files with information about encoding)
-ET-G       encoding rules for texts, graphemic format
-ET-M       encoding rules for texts, morphological format

-F-         (prefix for files which explain file format for data files)
-FT-G        data entry format for texts, connected graphemic version
-FT-GL      same as -F-GT, but with volume label on each text line
-FT-GT      same as -F-G, but bundled in single file
-FT-GI      output format for texts, alphabetical list of items
-FT-GO      output format for texts, alphabetical list of occurrences

-G-         (prefix for general information outside current scope)
-G-COMPS    compositional analysis of basic data unit (e.g., text outline)
-G-TAB      tabulation of numeric data (e.g. summaries of entries by type)
```

16.

```
-N-         (prefix for notes)
-N-EDIT       editorial (i. e., relating to the published edition)
-N-GRAPH      graphemic
-N-HIST       historical
-N-LING       linguistic
-N-LITR       literary
-N-MISC       general or miscellaneous
-N-TEXT       textual (including graphemic)


Xxxa(--z).--- data files labeled as follows:
             X     CME (Cybernetica Mesopotamica, Electronic files)
                   A - archaeological materials
                   S - secondary literature
                   T - texts
             xx    sequential number of CMX disk (from -1 to 99)
                     A1 to T99
              a    generation of CMX disk (from A to Z)
                     A1a to T99z
               (--z) sequential number of files for given disk (optional)
                     A1a-01 to T99z999
                   -(--) extension identifying type of data
                         G - graphemic version
                         GO - graphemic index of occurrences


         T1A--1.G
         T           CME (Cybernetica Mesopotamica, Electronic Files)
         1           disk number
         A           generation
           -01       text number
             .G      extension specifying format
```

Fig. 1. Sample Disk Directory: Texts

	single text with vol/text at beginning of text	bundled texts with vol/text at beginning of each text	bundled texts with vol/ref at beginning of each line
diacritics rendered by double low ASCII	.G^	.GT^	.GL^
diacritics rendered by single high ASCII	.G	.GT	.GL
text formatted for typesetting (not for input)	.G'	.GT'	.GL'

SYNOPSIS OF RELEVANT ENTRIES:

	.G and .GT	.GL
suffix ^	$$ T4A-03 $ ARET2.3 r.1,2 20 gu2-[^li]^-lum S^U-BA-[TI]	$$ T4A-03 $ ARET2.3 ARET2.3.r.1,2 20 gu2-[^li]^-lum S^U-BA-[TI]
no suffix	$$ T4A-03 $ ARET2.3 r.1,2 20 gu2-⌐li¬-lum ΣU-BA-[TI]	$$ T4A-03 $ ARET2.3 ARET2.3.r.1,2 20 gu2-⌐li¬-lum ΣU-BA-[TI]
suffix '	*T4A* 03 *ARET* 2 3 **r.1,2** 20 *gu₂-⌐li¬-lum* ŠU-BA-[TI]	*T4A* 03 *ARET* 2 3 **ARET2.3.r.1,2** 20 *gu₂-⌐li¬-lum* ŠU-BA-[TI]

Fig. 2. Types of data files (input)

!.EXE   gives first orientation and reads .XPL files

EXPLANATORY FILES

-P-xxxxx.XPL   explains goals and procedures for program "xxxxx" (e.g. -P-BUNDL)

FILE-HANDLING PROGRAMS

A. Changes between file formats

-P-BUNDL.EXE   bundles several .G format files into a single .GL or .GT file
-P-SPLIT.EXE   splits a bundled .GL or .GT file into individual .G files
-P-SWTCH.EXE   switches between .GL and .GT formats

B. Disk operations

-P-SGMNT.EXE   segments long files into arbitrary length files (generally used to copy long index
                     file from hard to floppy disk)

C. Screen and printer operations

-P-ASCII.EXE   rewrites text or output files by converting single ASCII into double ASCII diacritics
                     (e.g. the two ASCII characters s^ to single ASCII 229 = Greek sigma)
-P-LASER.EXE   rewrites text or output files by introducing proper codes for laser jet printing of
                     sub- and superscripts, italics, bold face, etc.

INDEXING PROGRAMS

-PT-G1.EXE   this is the major indexing program, which generates up to 19 indices from text files
                     in any .G format. Besides an index of note occurrences, there are indices by oc-
                     currences, by items in alphabetical order, and by items in order of frequency,
                     for each of the following six categories: words, numerals, personal names,
                     divine names, geographical names, other proper names.

Fig. 3. Sample disk directory: Programs

Output files – signs

    1. -I-ARETS.GV sign index by values in alphabetical sequence

    2. -I-ARETS.GS sign index by sign shape

    3. -I-ARETS.GF sign index by frequency

    4. -I-ARETS.GA sign index by number sign

    5. -I-ARETT.GA triad index by sign shape

Output files – words in graphemic format

    6. -I-ARET!.NO notes

    7. -I-ARET!.SLC list of texts that have been selected for indexing

    8. -I-ARETW.GO index of words (other than digits and names), with occurrences

    9. -I-ARET#.GO index of digital notations, with occurrences

    10. -I-ARETP.GO index of personal names, with occurrences

    11. -I-ARETD.GO index of divine names, with occurrences

    12. -I-ARETG.GO index of geographical names, with occurrences

    13. -I-ARETN.GO index of other proper names, with occurrences

    14. -I-ARETW.GC index of words (other than digits and names), with occurrences and context (and so on for digits, etc.)

    15. -I-ARETW.GA index of words, by items, in alphabetical sequence

    16. -I-ARET#.GA index of digits, by items, in alphabetical sequence

    17. -I-ARETP.GA index of personal names, by items, in alphabetical sequence

    18. -I-ARETD.GA index of divine names, by items, in alphabetical sequence

    19. -I-ARETG.GA index of geographical names, by items, in alphabetical sequence

    20. -I-ARETN.GA index of other proper names, by items, in alphabetical sequence

    21. -I-ARETW.GF index of words, by items, in order of frequency

    22. -I-ARET#.GF index of digits, by items, in order of frequency

    23. -I-ARETP.GF index of personal names, by items, in order of frequency

    24. -I-ARETD.GF index of divine names, by items, in order of frequency

    25. -I-ARETG.GF index of geographical names, by items, in order of frequency

    26. -I-ARETN.GF index of other proper names, by items, in order of frequency

Fig. 4. List of Output Files

SIGN INDEX BY VALUES IN ALPHABETICAL SEQUENCE
(.GV FORMAT - NO CORRELATIONS BETWEEN VALUES AND SIGNS)

Processed on 01-18-1990

Data derived from:
- file 1: ARET2x8.gl

Value	Occurr.	Percent of Corpus (Totals: items, 8765 occurrences)

LOGOGRAMS, DETERMINATIVES AND UNKNOWN VALUES
(Totals: 3600/524/966 = 5090 occurrences)

A	30	.34 %
A.KAS$_4$	1	.01 %
A.MU	1	.01 %
AB.SI	2	.02 %
AB$_2$	4	.05 %
AB$_2$.RU	1	.01 %
AGA$_3$	2	.02 %
[...........]		

PHONOLOGICAL VALUES   (Total: 3455 occurrences)

'a$_3$	57	.64 %
a	124	1.41 %
a$_2$	2	.02 %
ab	2	.02 %
ab$_2$	4	.05 %
abba$_2$	1	.01 %
ad	10	.11 %
ag	2	.02 %
[...........]		

NUMERALS

(10+)10	1	.01 %
(GIN2)	3	.03 %
(N)	2	.02 %
...N+10	1	.01 %
1	495	5.65 %
1(wv)	3	.03 %
[...........]		

(Continued on next page)

21

UNASSIGNED VALUES

GEŠTUG$_0$(GIŠ.PI)	1	.01 %
GUG$_0$(GUL.ZA)	1	.01 %
HA$_0$(ŠAG	4	.05 %
PA$_4$.SADA$_0$	1	.01 %
SA$_0$(NINDA$_2$	16	.18 %
[...........]		

SPECIAL SIGN NOTATIONS

C700	13	.15 %
C701	1	.01 %
[...........]		
LAK20	1	.01 %
PA.C702	1	.01 %
SADA$_0$(C717)	1	.01 %
TAG$_0$(LAK492)	33	.38 %
[...........]		

BREAK NOTATIONS

...	51	.57 %
X	64	.72 %
(X)	4	.05 %

SIGN CONFIGURATION   (Total: 85 occurrences)

@< (sign reduction)	54
@/ (oblique sign)	1
[...........]	

LIGATURES   (Total: 88 occurrences)

*E$_2$x*PAP	9
*GA$_2$x*LA$_2$	4
[...........]	

GRAPHIC PECULIARITIES   (Total: 17 occurrences)

;	(wrap-around)	1
cb2	(col. break)	1
[...........]		

Fig. 5. Sample output file: -I-ARETS.GV

(The complete index consists of 10 Kb, for a total of 14 printed pages)

INDEX OF OCCURRENCES
- PERSONAL NAMES FROM:
- file 1: *ARET*1x8.gl
- file 2: *ARET*2x8.gl
- file 3: *ARET*3x8.gl
- file 4: *ARET*4x8.gl
- Total occurrences - males: 293 / females: 133 / other: 8613 // all: 9039
- Total items: 3330
Processed on 02-01-1990

[...........]
*AN-du
    p 1. 5:*ARET*1.5.v.12, 14-1

*BAD-*E2
    p*ARET*1.5.v.1, 7-1
    p*ARET*1.8.v.7, 7-1
    p*ARET*1.17.v.10, 4-2
    p*ARET*3.239.v.1, 2'-1
    p*ARET*3.320.r.2, 3'-1
    p*ARET*3.371.r.2, 9'-1
    p*ARET*3.457.r.9, 4'-1
    p*ARET*3.458.v.6, 1-1
    p*ARET*3.468.r.3, 17-1
    p*ARET*3.930.r.1, 5'-1
    p*ARET*4.14.v.2, 18-1
 -- *BAD-*E2 -- 11

*BAD-*HI
    p*ARET*3.468.r.3, 19-1

*BAN-*NI-za-dum
    p*ARET*3.968.v.2, 2'-1

   [...........]
zu2-ba
    p*ARET*1.10.r.2, 2-1
    p*ARET*1.10.v.6, 22-1
    p*ARET*3.527.v.4,12'-1

    p*ARET*4.18.r.7, 3-1   zu2-[^ba]^
zu2-ba-*LUM
    p*ARET*3.100.v.1, 6'-1
    p*ARET*3.177.r.4, 6'-1
    p*ARET*3.533.r.4, 1'-1
    p*ARET*4.13.r.10, 13-1
zu2-ba4-:2
    p*ARET*3.497.r.2, 1'-1

Fig. 6. Sample output file: -I-ARETP.GO

(The complete index consists of 342 Kb, for a total of 298 printed pages)

INDEX OF ITEMS, SORTED ALPHABETICALLY
- DIVINE NAMES FROM:
- file 1: *ARET*1x8.gl
- file 2: *ARET*2x8.gl
- file 3: *ARET*3x8.gl
- file 4: *ARET*4x8.gl
- Total items: 98
- Total occurrences - male: 1687 / female: 69 / other: 193 // all DN's: 1949
Processed on 02-01-1990

[............]	
DINGIR- = *PI-da-'a3-nu- = KI	1
DINGIR = *NI-da-*BAL	1
DINGIR = -*AMA-is^	4
DINGIR = -*AMA-ra	6
DINGIR = -*EN	1
DINGIR = -*HAR-si-in	1
DINGIR = -*KU-*RA	105
DINGIR = -*KU-*RA-ma-i-da	1
DINGIR = -*KU-*RA-i-da-ma	1
DINGIR = -*KU-*RA-ma-i-da	1
DINGIR = -*KU-*RA	1
[............]	
DINGIR = -'a3	1
DINGIR = -'a3-da	48
DINGIR = -'a3-da- = KI	1
DINGIR = -'a3-ma-ri2-ig	1
DINGIR = -a-da-ma	6
DINGIR = -a-dam-ma	2
[............]	
'a3-da	1
a-gu2	1
da-gan	4
da-mu	500
ga-mi-is^	7
ga-mis^	1
li-im	86
ma-lik	758
ma-lik-:2	4
ra-sa-ap	1
[............]	

Fig. 7. Sample output file: -I-ARETD.GA
(The complete index consists of 2 Kb, for a total of 3 printed pages)

INDEX OF ITEMS, BY FREQUENCY
-WORDS OTHER THAN NAMES AND NUMERALS FROM:
- >   file 1: *ARET*1x8.gl
- >   file 2: *ARET*2x8.gl
- >   file 3: *ARET*3x8.gl
- >   file 4: *ARET*4x8.gl
--Total occurrences: 17316
---Total items: 1961
----Processed on 02-01-1990
-----(Percentages are computed on the basis of total number of occurrences.)
-----(Bar histograms correspond to percentage values: ■ = 1/2 percentile.)

	Total	Item	Percentage	
	4106	DAR	23.71%	■■■
	3001	...	17.33%	■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■
	2248	AKTUM-=TUG2	12.98%	■■■■■■■■■■■■■■■■■■■■■■■■■■
	2090	TUG2	12.06%	■■■■■■■■■■■■■■■■■■■■■■■■
	1990	SA6	11.49%	■■■■■■■■■■■■■■■■■■■■■■
	1740	IB2-:3	10.04%	■■■■■■■■■■■■■■■■■■■■
	1414	MI2-=TUG2	8.16%	■■■■■■■■■■■■■■■■
	1177	'a3-da-um-=TUG2-:2	6.79%	■■■■■■■■■■■■■
	1147	LU2	6.62%	■■■■■■■■■■■■■
	1137	IB2-:3-=TUG2	6.56%	■■■■■■■■■■■■■
	778	EN	4.49%	■■■■■■■■■
	720	SIKI	4.15%	■■■■■■■■■
	717	KU3-GI	4.14%	■■■■■■■■■
	697	BABBAR-KU3	4.02%	■■■■■■■■■
	640	in	3.69%	■■■■■■■■
	631	GU-DUL3-=TUG2	3.64%	■■■■■■■■
	549	X	3.17%	■■■■■■■
	539	'a3-da-um-=TUG2-:1	3.11%	■■■■■■■
	503	mi-at	2.90%	■■■■■■
	474	IB2-:2-=TUG2	2.73%	■■■■■■
	456	S^U-BA4-TI	2.63%	■■■■■■
	441	wa	2.54%	■■■■■■
	434	C700	2.50%	■■■■
	425	GU-MUG-=TUG2	2.45%	■■■■■
	418	MA-NA	2.41%	■■■■■
	416	MAS^KIM	2.40%	■■■■■
	351	GIN2	2.02%	■■■■■
	338	DILMUN	1.95%	■■■■
	333	S^U-MU-TAG0(LAK492)	1.92%	■■■■
	325	S^U-DU8	1.87%	■■■■
	[..........]			

Fig. 8. Sample output file: -I-ARETW.GF

(The complete index consists of 57 Kb, for a total of 37 printed pages)

INDEX OF ITEMS, BY FREQUENCY
-GEOGRAPHICAL NAMES FROM:
->   file 1: *ARET*1x8.gl
->   file 2: *ARET*2x8.gl
->   file 3: *ARET*3x8.gl
->   file 4: *ARET*4x8.gl
--Total occurrences: 3900
---Total items: 1201
----Processed on 02-01-1990
-----(Percentages are computed on the basis of total number of occurrences.)
-----(Bar histograms correspond to percentage values: ■ = 1/2 percentile.)

Total	Item	Percentage	
160	kak-mi-um-=KI	4.10%	■■■■■■■■■
147	ma-ri2-=KI	3.76%	■■■■■■■■
98	ma-nu-wa-ad-=KI	2.51%	■■■■■■
89	ar-mi-=KI	2.28%	■■■■■
80	*NI-ra-ar-=KI	2.05%	■■■■■
76	du-ub-=KI	1.94%	■■■■
67	du-lu-=KI	1.71%	■■■■
59	ib-al6-=KI	1.51%	■■■■
58	i3-mar-=KI	1.48%	■■■
58	ra-'a3-ag-=KI	1.48%	■■■
57	bur-ma-an-=KI	1.46%	■■■
51	lu-ba-an-=KI	1.30%	■■■
48	a-ru12-ga-du-=KI	1.23%	■■■
42	gar3-mu-=KI	1.07%	■■■
41	i-bu0(NI)-bu-=KI	1.05%	■■■
41	ir-i-dum-=KI	1.05%	■■■
41	ur-sa2-um-=KI	1.05%	■■■
38	a-da-bi2-ig-=KI	.97%	■■
38	ar-ha-du-=KI	.97%	■■
33	'a3-za-an-=KI	.84%	■■
31	hu-ti-mu-=KI	.79%	■■
30	'a3-da-ni-=KI	.76%	■■
27	da-ra-um-=KI	.69%	■■
27	gu2-da-da-num2-=KI	.69%	■■
26	ha-ra-an-=KI	.66%	■■
25	ab2-s^u-=KI	.64%	■■
21	'a3-ma-du-=KI	.53%	■■
21	kab-lu5-ul-=KI	.53%	■■
20	lum-na-an-=KI	.51%	■■
19	na-gar3-=KI	.48%	■
19	sa-nab-zu-gum2-=KI	.48%	■

[...........]

Fig. 9. Sample output file: -I-ARETG.GF
(The complete index consists of 40 Kb, for a total of 23 printed pages)

LES ANNALES ARCHEOLOGIQUES
ARABES SYRIENNES

REVUE D'ARCHÉOLOGIE ET D'HISTOIRE

CONTENT