"Studies in Ebla Graphemics, 1"
in *Studi eblaiti* 5 (1982) pp. 39-74

please see bookmarks

STUDIES IN EBLA GRAPHEMICS, 1

Giorgio Buccellati

1. The Theory of Graphemics.

1.1. Graphetics, Graphemics, Orthography.

The study of graphemics, if not the term, is well rooted in the Assyriological tradition. Its principles and presuppositions, however, are not usually made explicit and its application is somewhat on an ad hoc basis rather than systematic. The recognition of discrete values, and the consistency of the notation, goes back to the early days of the discipline, and it has remained with us ever since: this has facilitated both the practical task of the publication of cuneiform texts and, to some extent, our understanding of the cuneiform writing system. The recent discovery of a new adaptation of this system, namely the corpus of tablets found at Ebla, provides a welcome opportunity for assessing the theoretical and practical framing within which we operate - and it is to this topic that the present paper is dedicated.

The ending "-eme", with its derivatives "-emic" and "-emics" - as in "grapheme", "graphemics" - has come to be used in linguistics with a very specific meaning: it refers to an entity which (a) is conceived as a part of a self-contained whole, or a component member of a well defined set, and which (b) is tested by reference to a full parallel, closed system. A phoneme, for example, is a sound charged with special contrastive value within the framework of a given set of

sounds which alone are used to carry meaning. While phonemics deals then with a closed system of sound contrasts defined in terms of a parallel system of meanings, phonetics deals with sound contrast as an open ended system. Phonemically, the contrast between the sounds [i] and [ɪ] (as in English "sheep" and "ship") is relative to whether or not it can be *defined distributionally as cooccurring in two parallel, closed systems* - so that, in point of fact, this contrast *is* phonemic in English, while it is not, let us say, in Italian. Phonetically, on the other hand, the contrast between the two sounds is absolute and thus it obtains in Italian as well, even if there a phonemic contrast of the sort is missing.

The same distinction may be applied to writing. On the one hand, corresponding to phonetics, we have the study of individual signs in their graphic appearance - in the case of cuneiform, these are wedges which may be long or short, vertical or horizontal, single or in cluster, deep or shallow, etc. The question then arises as to how we must proceed in order to show the presence or absence of graphemic contrast. The nature of the concept implies, as we have just seen, the definition of a self-contained system distributionally correlated to another such system. Accordingly, a GRAPHEMIC SYSTEM may be defined as one which *includes all correlations between signs and the phonemic system*. In turn, a GRAPHEME is the *minimal contrastive unit* within such a graphemic system. In the cuneiform system, for example, the difference in length among wedges, while real in terms of a graphic or physical contrast, will not be found to correlate normally with any phonemic alternation. On the other hand, the difference between a vertical and a horizontal orientation, besides being real in terms of a graphic or physical contrast, is also found to correlate distributionally with a phonemic alternation. For instance, the different orientation of the two signs  and  , while real in terms of graphic or physical contrast, will not be found to correlate with any phonemic alternation, that is to say, every time occurs, the sign in which it occurs has the same phonemic value as , and viceversa: the value */diš/*, for instance, may be represented graphically as either or , and */ma/* as either  or . On the other hand, the different orientation of the two signs and , besides being real in terms of graphic or physical con

trast, is also found to correlate with a precise and specific phonemic alternation. That is to say, every time  occurs, it has a different value from , and viceversa: the value */diš/* is represented by and not by , the value */aš/* is represented by and not by . It was by defining such covariations that sign lists have traditionally been established. Thus it can be said that cuneiform writing, as an *open, graphic system*, includes all possible combinations of wedge-like signs, e.g. doodling and decorative patterns as well as graphemes; cuneiform writing as a *closed, graphemic system*, on the other hand, includes only those combinations of wedge-like signs which exhibit a covariation with phonemic values.

Next to the term grapheme, defined above, we will use the term GRAPH to refer to any combination of wedges or other signs impressed on clay, whether or not they correspond to a grapheme. Thus a graph may include a single wedge or cluster of wedges which correspond to a grapheme; it may include graphic marks such as punctuation (e.g. the colon or "Glossenkeil"); it may include a straight line used to divide lines or cases, whether impressed with the stylus or with a string; it may include doodling and decorative patterns in cuneiform shape, and so on. The study of graphs in their physical appearance may be called GRAPHETICS, after the pattern of the word "phonetics". This, in turn, may be divided into graphics and palaeography. In GRAPHICS, the graphs are sorted according to formal characteristics, such as the ductus or the concern for elegance ("calligraphy). In PALAEOGRAPHY, on the other hand, graphs are sorted according to chronological criteria.

The standardized graph which has graphemic value is generally known in Assyriology as a SIGN. Note that sign and grapheme are analogous but by no means identical concepts: "sign" refers to the physical configuration of a graph which has graphemic value, whereas "grapheme" refers specifically to the minimal unit which brackets the phonemic and the graphic registers. Thus "grapheme" is more comprehensive in its import than "sign".

Graphs belonging to the same grapheme will be called ALLOGRAPHS. Allographs may be free variants (e.g.  and occuring in the same text for MU, see TM.75.G.1452 v. 3:4 and v. 4:1, in *SEb*

3 [1980], fig. 9b), or else they may be combinatorial variants, if they occur in complementary distribution (e.g.  for KI when occurring isolated, in contrast with  also for KI when occurring at the end of a case, see P. Fronzaroli: *SEb* 3 1980 , p. 36). A special type of combinatory allography obtains with HOMOPHONES: these are different signs with identical phonological values, such as  KI and  KI_2. Here then we have one grapheme, but two signs.

Next to graphetics and graphemics, there is ORTHOGRAPHY, which may be defined as a stylistic choice of graphemes at the lexical level. While graphemics deals with the correspondence between the graphic and the phonemic level, orthography deals with the correspondence between the grapho-phonemic, or graphemic, and the lexical level. Orthography choices are operative at various levels. (1) The conditioning factor for a graphemic choice may be the SPECIFIC MEANING of a given lexical item: thus, out of many ways in which the phonemic string */bābilim/* may be written, it will appear regularly as KA.DINGIR.RA^{KI} in the meaning of "Babylon", whereas in the meaning "gate of god" it may appear as *ba-ab* DINGIR-*lim* (*CADB* 19b). A subtype of this orthographic situation obtains when the reading of a given word requires a unique value for one or more signs as would be the case for DU-*lu*^{KI} at Ebla if the first sign were to be read GUB (see below, 2.4.1). (2) The conditioning factor may be the GENERIC MEANING of a given lexical item: this is typically the case with semantic indicators (determinatives), which identify a word as a member of a given lexical class, and can occasionally provide a clue to the phonemic value of a word - e.g. ^{D}En-lil *[Enlil]* vs. EN.LIL^{KI} */Nibru/*. (3) The conditioning factor may be a chronological, geographical or other NON-LINGUISTIC VARIABLE, which covaries with a given distribution of signs per word: thus the word *bāb* tends to occur as *ba-ab* in Old Assyrian and as *ba-ab* in Old Babylonian (*CADB* 19 f.). (4) The conditioning factor may be GRAPHO-TACTIC, in the sense that a graphemic determination is made on the basis of the arrangement (sequence and relative position) of the graphemes within the boundaries of the word. At Ebla, for instance, the sign NI does not have the value *li* in word initial position, so that the preposition NI-*na* should be read *i-na* instead of *li-na* (the latter would

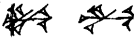
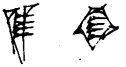
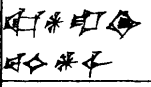
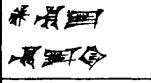
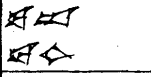
			graph	
			grapheme (sign(s) / value)	non-grapheme
GRAPHETICS	formal sorting (GRAPHICS)		 ki	decoration, doodling, etc. 
	chronological sorting (PALAEOGRAPHY)		 ki	
GRAPHEMICS	basic identification		 ki	—
	ALLOGRAPHY	free variant	 mu	
		combinatorial variant	 ki	
		homophones	 ki/kí	
ORTHOGRAPHY	the variable affecting the word is:	meaning specific	 KA.DINGIR.RA^{KI} ba-áb DINGIR-lim	—
		meaning generic	 DINGIR$_{En\text{-}líl}$ EN-LÍL^{KI}	
		non-linguistic	 ba-ab ba-áb	
		grapho-tactical	 ì-na	

Fig. VI - The Graphic, Graphemic and Orthographic Systems

obtain only if the reading could be demonstrated on the basis of independent criteria, in which case then the reading *li-na* would have to be understood as meaning specific, in accordance with the first orthographic condition given above; see below, 2.4.1).

On the basis of the definitions given above (which are summarized in tabular form in Fig. VI) certain terminological matters can be more properly appreciated. The term LOGOGRAM should be preferred over "ideogram" because it refers properly to the phonemic ("logo-") rather than the semantic level ("ideo-"). The term VALUE should be preferred over "reading" when referring to individual signs, because it refers properly to the relational function of the signs within a system rather than implying, as "reading" does, that the sign corresponds by itself to a definite phonemic configuration. For the same reason, the term PHONOLOGICAL (VALUE) should be preferred over "syllabic", because the latter implies a correlation between signs and syllables, and the term POLYVALENCE should be preferred over "polyphony", because the former implies more clearly the conditional value of the phonetic component of graphemes.

1.2. Graphemic Rules.

It is important to consider not only the graphemes as minimal units of a system, but also the rules which are operative in interrelating the graphemes. Here are a few well known examples. (1) The sequential order in which graphemes may cooccur is a rule, because it states a procedure of arrangement for the minimal units: in a typical cuneiform text, the spatial (graphemic) sequence from left to right within the same case or line corresponds to the temporal (phonological) sequence from start to end of an utterance - i.g. texts are read from left to right. (2) Graphemic boundaries in sign sequences, i.e. cases or lines, coincide with the phonemic boundaries at the beginning or the end of the words - i.e., a word is not generally split between lines. (3) Certain types of phonemic close juncture between words are reflected in the graphemic phenomenon of bound sequences - i.e. a preposition and the noun following it are not split between lines.

While these examples are simple and relatively obvious, the rule

which they serve to illustrate has important ramifications. The most significant concerns the notion of polyvalence. Most cuneiform signs are polyvalent in the sense that one and the same sign may correspond to two or more phonemic configurations; for instance, the sign  UD has the values */ud/*, */tam/*, */ūmum/*, */ūmim/*, and so on. The choice among these values is made not on the basis of the graphemes themselves but rather of a graphemic rule. That is to say, the grapheme as a minimal unit is in fact a set which subsumes all the possible pertinent values - alternatively, it is "polyvalent". The distinction among the values is not a function of the minimal unit itself, but rather of the cooccurrence possibilities with other adjacent minimal untis. Given the two graphemic strings:

(1)  *an ni* UD

(2) *i na* UD

the two different values of  UD are determined by the two different types of graphemic clustering, which make */tam/* the only possible value in the first example, and */ūmim/* the only one in the second. The different values of a sign, therefore, are the result of the application of distributional graphemic rules - alternatively, they are contextually conditioned (on this and what follows see Reiner 1973).

This distinction between graphemes and graphemic rules has more important consequences than it may seem at first. To illustrate this point, let us look at another example. In the sequence

 *i pa ra ás*

one could argue that the sign  PA has the value */par/* since the phoneme configuration which corresponds to such a graphemic string is */iparras/*. Logically, the value PAR_x for PA is on the same level as TAM for UD: both values are of the same phonemic status, and both are derived from the relative position of the pertinent grapheme in a given graphemic sequence. It is, in other words, a graphemic *rule* concerning the sequential order which defines the existence of the two values.

Should we, then, introduce a value par_x for PA, and similarly for all analogous cases? Certainly not: a syllabary built on such premises would be, however logical in principle, extremely unwieldy.

The example proposed may thus serve as a *reductio ad absurdum* to show the nature and validity of the concept of graphemic rules. If par_x is not to be proposed as a grapheme, it is because of a rule which is normally left implicit, but which may be stated explicitly as follows: any sign value ending in vowel and followed by another sign within word boundaries, contains an optional consonantal length applicable to the first consonant of the following sign; the formula notation for the rule is: (C)V(ˀ)- . Conversely, a case like *tár-kab-bù* (see below, 2.4.1) contains a superficially deceiving indication of consonantal length: are we to propose a value ka_x for KAB in order to avoid this confusion? This, too, would be exaggerate. We may instead state a rule which says that a sign value with the configuration (CVC)CVC allows for the optional deletion of final consonant when followed within word boundaries by a sign value beginning with the same consonant; the formula notation of the rule is: $(CVC)CV\not{C}_1\text{-}C_1$.

What about cases where the set of values for the same sign depends on the alternation among phonemes belonging to the same phonemic class? For example, values beginning with a voiceless stop may alternate with values beginning with a voiced stop for the same point of articulation: here the Assyriological tradition, in contrast with the case of par_x(PA), provides distinct values, e.g. *ta* and *dá* for TA. Similarly with values containing the vowel *i* which also appear with the value *e*, e.g. *iš* and $eš_{15}$ for IŠ. These are cases where the application of graphemic rules is made explicit, in the Assyriological tradition, at the level of graphemic notation. Because of the growth of both textual data and linguistic analysis, the notation of graphemic rules by means of graphemic values has begun to pose some problems. With regard to Ebla, for instance, Fronzaroli has posed the question as to whether the set of signs ŠA, ŠI, ŠU, SUM should be translitterated *ṯa/ḏa*, *ṯi/ḏi*, etc. or should be left as *ša*, *ši*, *šu*, *sum* leaving the proper phonemic conversion to the operation of graphemic rules (*SEb* 1 [1979], pp. 87-89). In the first case we contribute to a proliferation of values, for which the work by Von Soden and Röllig has been especially called to task; in the other case we have the so-called simple values, which have been advocated especially by Gelb. It would appear that both ap

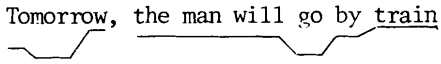
proaches are equally justified, depending on the point of view one chooses, and with the following reservations. The values must be clearly understood as values, i.e. there must be a clear appreciation of the conditioning factors, whether phonemic, lexical or contextual: the so-called proliferation, then, is not a misrepresentation of facts, but simply a type of notation which embeds graphemic rules into graphemic values. The so-called simple values, on the other hand, have no greater intrinsic simplicity than other values: they are symbols for sets of values, with the understanding that the appropriate choice has to be derived by means of correlative graphemic rules; if one opts for this choice of transliteration, it becomes imperative to articulate explicitly the graphemic rules which are in fact operative.

To summarize, the question of current types of notation for graphemic rules may be analyzed under three headings.
(1) The most obvious, and most frequently applied, rules have a standard notation distinct from the graphemes themselves - as in the case of the rules stated at the beginning of this section: "the sequential order of graphemes is expressed by the convention of transliteration signs from left to right"; or: "the graphemic boundaries at the beginning and the end of the line of a text are represented by graphic devices such as extra space".
(2) The less obvious, and less frequently applied, rules are expressed by alternate values given as part of the graphemic inventory: thus *ta* and *dá* for TA, *íš* and *éš*$_{15}$ for IŠ.
(3) The rules which lie in between these two extremes are usually not expressed at all in transliteration, as in the case of *i-pa-ra-áš* or *tár-kab-bù*.

As a result, transliteration may be described as a hybrid notational system, because it provides a perfect identification of graphemes on the one hand, but, on the other hand, an imperfect identification of graphemic rules (presented, for the most part, as graphemes or else lacking a notation of their own). This situation may become downright harmful when the transliteration is used at face value, without the expertise necessary to make the necessary phonemic compensation by applying the appropriate graphemic rules. Once the

notion of graphemic rule is properly understood, as well as the corresponding imperfection of the notation system, the function and limitations of the standard system of transliteration can be more clearly appreciated.

One may compare this with a similar situation in English. One could say that the phoneme /f/ can be represented by the following graphemes or graphemic clusters: F as in "if", PH as in "telephone", and GH as in "rough". None of these, however, are univocal, in the sense of a one-to-one correlation between graphemic and phonemic configurations. The grapheme F can stand for the phoneme /v/ as in "of"; PH can stand for /ph/ as in "loophole"; and GH can stand for Ø as in "bough". The determination of the correct "reading" will depend on graphemic rules, which may be, for instance, positional (e.g. double FF will always be /ff/) or lexical ("rough" /r f/ vs. "bough" /bou/). Another helpful analogy is the one which compares graphemes to lexical items, and graphemic rules to grammar: just as we cannot load all grammatical rules onto the lexicon, thus we cannot load graphemic rules onto the graphemes. For instance, an (English) word may be susceptible of a variety of intonations, but these are not given lexically, because most words are totally polyvalent in this respect and the specific value is conditioned in each case by the context. Thus there is no reason to list lexically "train↓", "train", and "train↑" as three different items; rather, the pertinent intonation may best be rendered by using a different notational register which refers to intonation exclusively, e.g.



Similarly, the best way to express fully the graphemes and the graphemic rules is to use a double notational register, the transliteration and the transcription. The transliteration provides a complete rendering of graphemes and an incomplete rendering of graphemic rules (e.g. *i-pa-ra-ás*): because of the inadequacy in rule notation, transliteration is only a partial graphemic notation, hence it cannot render fully the underlying phonemic reality. Instead of striving for a superposition of notational systems (i.e. rules above graphemes), it is simple and more effective to use transcription next to transliteration: transcription is not graphemic, but purely phonem

ic, and is based on the graphemic system as a whole, although it does not reproduce its component parts (e.g. *iparras*). This system, advocated especially by I.J. Gelb, has been regularly and exemplarily applied to the Ebla materials by P. Fronzaroli.

1.3. Distributional and Structural Analysis.

While the graphemic rendering of a living language such as English may be made to adjust easily to the intended phonemic configuration, because the latter is known independently of the writing medium, the same is not true in the case of a dead language, where the phonemic system is not only known through its graphemic counterpart. As a result, graphemic transliteration remains the best way, for dead languages, to bracket into one and the same notational system both the identity of the original graphic notation and the presumption of the parallel phonemic register. The degree to which such a phonemic presumption may approximate reality will vary from case to case and from time to time, but at least the graphic register, which is a basis for that presumption, will remain expressed unequivocally.

Given the fact that graphemic transliteration, as a system, can render unequivocally the graphic register, what are the criteria for bracketing the phonemic register onto it? What are the discovery procedures for going from graphic to graphemic recognition? The answer lies clearly in the application of distributional anlysis, whereby patterns of cooccurrence and covariation can be identified. What results is both a syntagmatic and a paradigmatic study of the "signs": a SYNTAGMA will be a recurrent sequence of signs as found in the texts, and a PARADIGM the set of alternants which covary in parallel fashion with the environment. Thus a GRAPHEMIC PARADIGM will be, for instance, a set of different graphemic sequences, conditioned by positional and lexical factors, used to express the same phonemic sequence (see below, 4.1, for an example). For a dead language, a total control of cooccurrences is necessary in order to determine paradigms which may be applicable especially to nuances of the graphemic system.

Ad hoc suggestions can be imaginative and heuristically useful, but

a test of their validity depends ultimately on how they fit within a structural system. This means that we must attempt to view the part in function of the whole, and not as an isolated fragment. To obtain this, we must operate not atomistically with bits of information, but structurally in terms of distributional classes: this procedure allows a higher degree of formal verification and objective testing. One of the advantages of graphemic analysis in this respect is the size of pertinent data available, from which distributional patterns are seen to emerge. We have literally millions of pieces of information which occur in diverse environments. An accurate quantification of such a wealth of data is made possible by the application of electronic data to the texts, as we shall see: we can then control a vast amount of information with procedures which are extremely diversified and flexible, we can perform a truly microscopic analysis on a macro-universe. But, and this is most important, such a procedure is not just a matter of the availability of a *mechanical* tool. It also requires new *technical* tools, which might allow us to properly categorize the data. It is this degree of formalization, even more than the technical awesomeness (or bother!) of the machine, which we find forbidding in applying the computer to our data. And yet through such an approach we can arrive at a more adequate structural control, and thus ultimately at a better understanding of the data, in terms of the *same* concerns we have in a traditional approach. I will try to show in what follows how this delicate interaction between abstract formalization and actual understanding of the data might in fact be implemented in actual practice particularly in regard to the texts of Ebla. I should first outline certain aspects of my research strategy which condition this presentation.

1.4. Research Strategy.

One advantage of theoretical explicitness is that we can more easily keep levels of analysis properly distinct. This alone would have significant returns even on the practical level, and justify empirically the abstract exercise about theoretical presuppositions. As is happens, the theoretical reflections offered above are not an end in itself, but rather an effort at making explicit the conceptual

underpinnings of a research which is, in fact, text oriented and ultimately very practical in its applications.

This paper is presented within the context of a broader research on the texts of Ebla, and is meant to serve ultimately a philological purpose. Conceived in part as a working report for a meeting of the International Committee for the Publication of the Texts of Ebla, the paper describes (in Part 2) a set of research tools which I propose be adopted by the Committee itself as part of its normal operations: this pertains in particular to the concordances and indices described below under 2.1, which it is hoped may become the standard operative files of the Committee. Future installments of these Studies on Ebla Graphemics will then serve at the same time to inform colleagues on the development of this more practical aspect of the project.

The substantive considerations presented in Part 3 below are illustrative of the specific analytical results which may be obtained with the procedures and tools outlined here. Further studies on Ebla graphemics, for which this paper serves as the Prolegomena, will deal more and more with individual substantive issues and provide a fuller documentary basis.

Finally, it must be stressed that, while this paper is presented within the context of the research on the texts of Ebla, its presuppositions and the accompanying technical support stem from a long-standing research project on electronic data processing of Mesopotamian materials. (For literature on the goals and results of the project see below, Part 4: *Bibliographical Note*). This means that my concern with Ebla graphemics is related to problems which go beyond the specifics of the Ebla corpus: the larger perspective which ensues would seem to assist in the choice of pertinent methods, just as the technical support already developed can provide the benefit of more powerful tools of analysis.

An important aspect of the approach followed here to data processing has been the double concern for multi-purpose encoding and multi-tiered formatting. The former has been in response to the need for a tight structural economy of the system as a whole, which aims at a broad utilization of the data and avoids fragmented and ad hoc

solutions. By multi-tiered formatting, on the other hand, I refer to the concern for obtaining a published presentation of the results which is so differentiated as to emulate the interactive work with the machine itself.

2. *The Practice of Graphemic Analysis.*

2.1. Introductory.

The needs of distributional and structural analysis are best served by various types of outputs which have been designed for the study of cuneiform texts in general, and have been adapted to the Ebla corpus in particular. The technical support provided be these tools is new in its configuration, and needs some explanation: this will be found below, together with illustrations of some of the outputs currently available. The data are derived from a preliminary corpus of some 156 texts, excerpted fully as far they have been published or made available in manuscript form; they can be subdivided as follows:

135 administrative texts
5 letters
4 lexical texts
2 lists
1 mathematical text
9 texts of unknown or unclear type.

In addition there is a group of Ebla personal names (EPNS) which have been quoted without context and have been entered as such in the data base. In the illustrations given below, texts are referred to by field number: references to the pertinent publication are provided by means of concordances which are included in the complete version of the outputs.

The total number of sign occurrences found in these 156 texts is 22,321, including signs for number figures. By way of comparison, we may note that this amounts to a little less than half the total number of sign occurrences for the Akkadian letters from Amarna (57,502, for which see below, 3.2).

The total number of inventory items, i.e. of discrete signs is 283 -

which is rather similar to other corpora from different areas and periods (on this too, see below, 3.2).

This is obviously a small sample, somewhat heterogeneous in its composition, and without any particular selection criteria other than availability at a given point in time. It is however sufficient to demonstrate the intended use of tools and applicability of procedures.

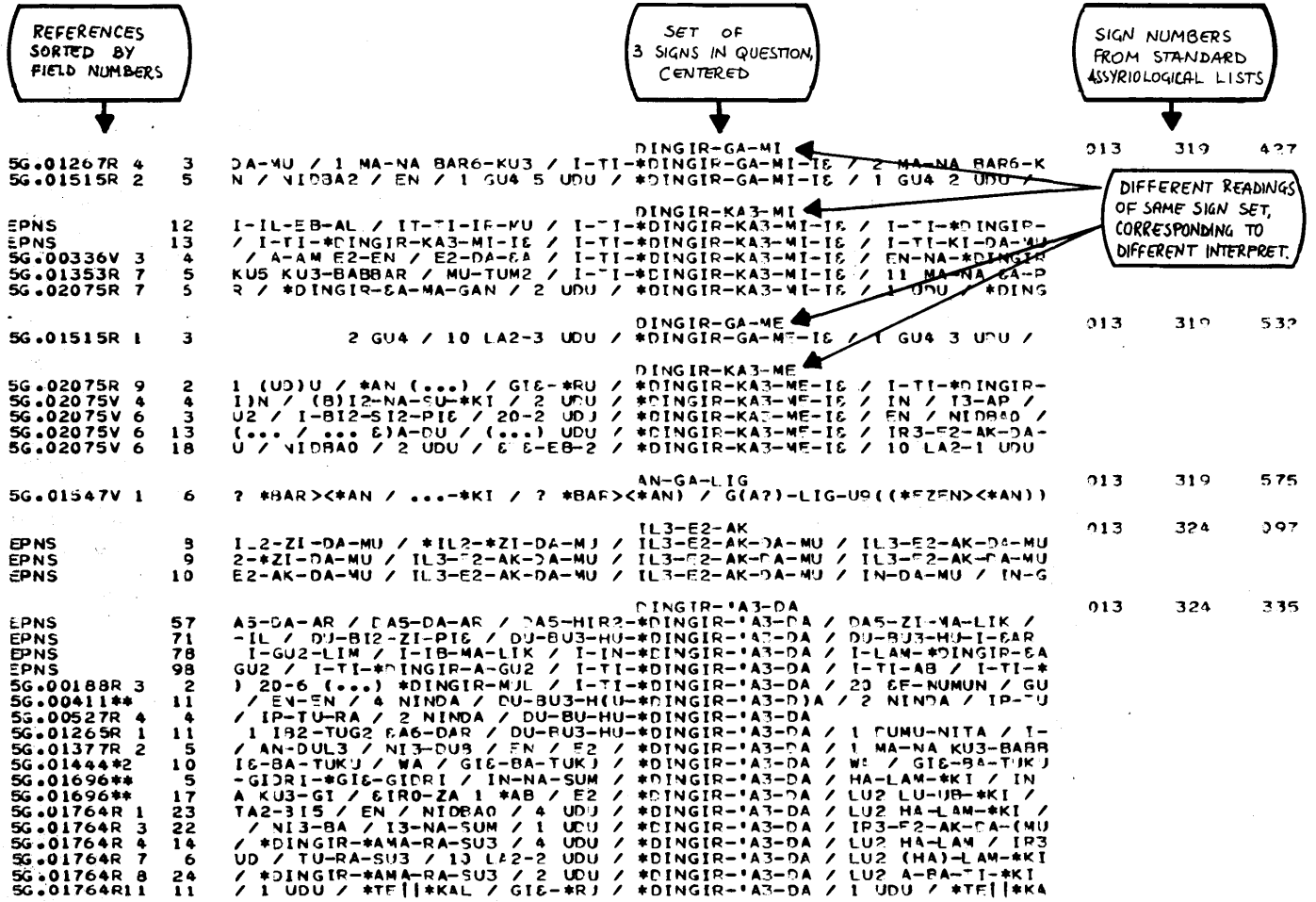
2.2. Categorization.

The primary tool for graphemic analysis consists of a categorization system, which lists each sign in its immediate environment. Such categorization system is produced in three different configurations: while the data and the basic categorization principles remain the same, the sorting cryteria and the degree of documentary comprehensiveness vary in each case so as to provide complementary outputs, which may be used for different types of searches and some of which may be produced more conveniently for a conventional type of publication.

The major categorization tool is the SIGN CONCORDANCE. This is based on a sorting by units of three signs; all passages which exhibit a pertinent 3-sign unit are quoted in extenso, with arbitrary cuts at left and right according to the conventional KWIC (Key Words In Context) format. The sorting by units is determined by the shape of the cuneiform signs; the sorting within units is determined first by the transliteration of the same 3-sign unit, and then by the references in ascending order by field number. The combination of the three signs in each unit is given a sequence which is based on the progressive graphic arrangement of cuneiform signs as standard in Assyriology. The main feature of interest for graphemic analysis is the fact that each individual sign is listed according to its graphic shape, with a subdivision of values as represented in normal transliteration and with a further subdivision based on the specific graphemic context, defined in terms of the two signs occurring to the right of the sign in question. Such listings provide for a rapid and exhaustive control of all possibilities of cooccurrence within a given corpus - i.e. allow a check on graphemic ruled as

1. SIGN CONCORDANCE

SAMPLE OF TWO ADJACENT PAGES, FROM WORKING COPY PRINTOUT.



DINGIR-*A3-DA 013 324 335

5G.01764V 1 2 UDU / IN UD / *DINGIR-*AMA-RA / *DINGIR-*A3-DA / EN-EN / NIDBAD /
5G.01764V 2 3 MI-AT 40 UDU / KU2 / DU-BU3-HU-*DINGIR-*A3-DA / 15 UDU / KU2 / A
5G.01764V 3 16 / EN / LUM-NA-NU-*KI / DU-BU-HU-*DINGIR-*A3-DA / I3-NA-SUM / IN D
5G.01764V11 26 -SI2-PIŠ / (...)-10 LA2-2 UDU / *DINGIR-*A3-DA / LU2 HA-LAM-*KI /
5G.01764V12 2 / *DINGIR-*AMA-RA-SU3 / 3 UDU / *DINGIR-*A3-DA / LU2 / HA-LAM-*KI
5G.01841** 14 6 MA-NA KU3-BABBAR / DU-BU3-HU-*DINGIR-*A3-DA / AN-ŠE3-GU2 2 LI
5G.02075R 7 8 U / *DINGIR-KA3-MI-IŠ / 1 UDU / *DINGIR-*A3-DA LU2 / LU-UB-*KI /
5G.02075R10 29 -DUB / 10 UDU / KU2 / DU-BU3-HU-*DINGIR-*A3-DA / 14 UDU / KU2 / G
5G.02075R11 5 / GURUŠ-GURUŠ / E2 / DU-BU3-HU-*DINGIR-*A3-DA / 60 LA2-1 UDU / K
5G.02238R 1 11 / IN UD / *A3-SI2-PU3 / 2 UDU / *DINGIR-*A3-DA / EN / NIDBAD / E2
5G.02238R 6 12 / E2-NUN / MA-LIK-TUM / 2 UDU / *DINGIR-*A3-DA / HA-LAM-*KI / I-D
5G.02238R11 18 UD / *DINGIR-*AMA-RA / 2 UDU / *DINGIR-*A3-DA / *SAG-DA-MU / NID
5G.02238V 2 7 A) / 1 UDU / E2-LUM / GIŠ-*RU / *DINGIR-*A3-DA / 1 UDU / E2-LUM /
5G.02238V 4 8 MI-AT 20 UDU / KU2 / DU-BU3-HU-*DINGIR-*A3-DA / 15 UDU / KU2 / I
5G.02238V 5 2 / GURUŠ-GURUŠ / E2 / DU-BU3-HU-*DINGIR-*A3-DA / 15 UDU / ŠU-RA /
5G.02238V 7 25 Š / LU2 A-DA-GAR3 / (D)U-BU3-HU-*DINGIR-*A3-DA / I3-NA-SUM / 1 UD
5G.02238V10 10 / EN-ZI-MA-LIK / 40 LA2-2 UDU / *DINGIR-*A3-DA / HA-LAM-*KI / EN
5G.02238V10 17 IN / *DINGIR-*AMA-RA / 12 UDU / *DINGIR-*A3-DA / EN / NIDBAD / IN
5G.02306R 2 5 A / 1 MI 20 UDU KU2 / DU-BU3-HU-*DINGIR-*A3-DA / 30-4 UDU GURUŠ /
5G.02306R 5 2 1-GA / 80-2-(1) UDU / DU-BU3-HU-*DINGIR-*A3-DA / 1 MI 20 UDU GURU
5G.02306V 2 8 GURUŠ / E2 IN / WA / DU-BU3-HU-*DINGIR-*A3-DA / 1 MI 50 LA2-3 UD
5G.11010R 2 27 NIDBAD / IN UD / KUR6 / 2 UDU / *DINGIR-*A3-DA / 2 UDU / *DINGIR-
5G.11010R 2 29 UDU / *DINGIR-*A3-DA / 2 UDU / *DINGIR-*A3-DA / EN / (E2 / *DING
5G.11010V 2 1* NITA-DUMU-NITA / EN / DU-BU3-HU-*DINGIR-*A3-DA / I3-NA-SUM / 3 UD
5G.11010V 2 9* ... / ...) / 20 LA2-(... UDU) / *DINGIR-*A3-(DA) / *A3-(TA2)-NI-(
6G.00089V 1 5 INGIR-KU-RA / ŠI2-IN / DU-BU-HU-*DINGIR-*A3-DA / IGI-DUB / WA / D
6G.00523V 4 17 I2-*KI 1 AKTUM-TUG2 / DU-BU3-HU-*DINGIR-*A3-DA / 1 GU-SUR2-TUG2 /
6G.00523V 9 5 3 / NI3-*DINGIR-MUL / DU-BU3-HU-*DINGIR-*A3-DA / LU2-TUŠ / MA-NU-

AN-RA-SA 013 328 104

5G.01345R20 5 DA / 1 ? SIKI / ŠU KEŠDA GUN3 / *AN-RA-SA-AP-*KI / E2-DA-*NI-*KI

DINGIR-RA-SA

EPNS 56 -NI-IL / EN-NA-RA-SA-AP / EN-NA-*DINGIR-RA-SA-AP / EN-NA-ŠUM / EN
5G.00336R 7 7 U_A 5 / XU3-BI2-SI2-PIŠ / EN-NA-*DINGIR-RA-SA-AP / NI-E2-I3-LU /
5G.00336V 3 5 I-TI-*DINGIR-KA3-MI-IŠ / EN-NA-*DINGIR-RA-SA-AP / A-A-BU3-IR3-KU
5G.00336V 5 8 IP-KUR-IA3 / TI-LA-IA3 / EŠ-DU-*DINGIR-RA-SA-AP / IP-HUR-*DINGIR
5G.01764R 5 18 MA-LIK-TUM / NIDBAD / 12 UDU / *DINGIR-RA-SA-AP / *A3-TA2-NI-TU3
5G.01764R 6 19 IR-*AMA-RA-SU3 / 10 LA2-2 UDU / *DINGIR-RA-SA-AP / *A3-TA2-NI-*KI
5G.01764R 7 22 UD / *DINGIR-*AMA-RA / 2 UDU / *DINGIR-RA-SA-AP / *A3-TA2-NI-*KI
5G.01764R 7 27 I / ZA-A-ŠE3 / NIDBAD / 2 UDU / *DINGIR-RA-SA-AP / *A3-TA2-NI-*KI
5G.01764R 8 15 UD / *DINGIR-*AMA-RA / 2 UDU / *DINGIR-RA-SA-AP / *A3-TA2-NI-*KI
5G.01764R 9 3 U-DUB / 2 UDU / *DINGIR-(...) / *DINGIR-RA-SA-AP / *A3-TA2-NI-*KI
5G.01764R 9 6 A-AP / *A3-TA2-NI-*KI / 2 UDU / *DINGIR-RA-SA-AP / *A3-TA2-NI-*KI
5G.01764R 9 13 / 2 UDU / GIR13-ŠUM / WA KUR6 / *DINGIR-RA-SA-AP / *A3-TA2-NI-*KI

DIFFERENT READINGS OF SAME SIGN SET, CORRESPONDING TO DIFFERENT WORDS

NOTES. EPNS in the Reference column corresponds to a separate repertory of Ebla Personal Names.

Printout codes which differ from the standard Assyriological codes:

& = š̌ C = ṣ X = ṭ * denotes upper case || = + >< = × RUO = ru_x

Fig. VII - Sign Concordance

operative in terms of sequential order.

In line with this characteristic of the Concordance (sorting by sign shape rather than by alphabetical value), normalization of sign readings is not necessary in order to retrieve together all occurrences of the same sign. In the illustrations given here, the transliteration follows the one proposed in the original version, whether published or in manuscript form: final formalization will be a function, precisely, of further graphemic analysis. It must be remembered that one of the advantages of a computerized data base is the ease of update; hence it is that the concordance in this format can serve as a current working instrument, constantly updated in terms of both new data and new interpretations of existing data. This will be the main practical use for the Committee and all scholars working on the publication of the texts: the availability of the concordance as a comprehensive file of all occurrences of all signs, sorted according to the graphic configuration of the signs quoted in each case and with their immediate environment. It will be apparent at a glance that the Sign Concordance subsumes in itself a lexical concordance, and one which is not conditioned by diverging interpretations in the choice of values. It should also be noted that, from a practical point of view, there is the considerable advantage of great ease of duplication, either on tape or on paper, unlike any other type of card file.

Another important philological use of the Concordance is the assistance it can provide in restoring broken passages or reading obscure passages. Where parallels are available (either exact parallels in the case of formulaic texts of a similar type) the sequence of a few signs without context can readily be matched against sequences for which a context is known.

The program which produces the Concordance from which the two pages of Fig. VII are derived is currently operational, and a volume of 781 pages has been produced for the Ebla texts listed above. The sheer bulk of this output makes it unfeasible to publish the results in conventional paper format - especially when one considers that the 781 pages Ebla volume corresponds to a limited input, inadequate in itself for any meaningful graphemic or other type of analysis.

Fig. VII

EBLA SIGN INDEX, 1

Sample, May 1980, based on a total sign count of 283 inventory items and 22,321 text occurrences.

Figures on the left refer to standard sign numbers: the correlative sign readings are given in **bold** face; a bullet (•) denotes variant readings of identical signs. Figures in parenthesis followed by a multiplication sign refer to the number of times a given sign sequence occurs in the texts; if no figure is given, then the pertinent sign sequence occurs only once, e.g. **DINGIR-ni-da**-kul occurs 56 times, **DINGIR-ni-da**-la only once. A slash (/) denotes line or case boundary. The transliteration follows the one given in the original text edition, without normalization.

13	1	114		**i-ti-DINGIR-aš-tár** (3×)
				eb-du-DINGIR-aš-tár
				kum-DINGIR-aš-tár
13	1	335		**DINGIR-aš-tá-pi$_5$** (2×)
				DINGIR-aš-tá-bi$_5$ (2×)
13	5	59		**DINGIR-ba-li-ha**
13	5	328		**DINGIR-ba-ra-du**
13	13			**dingir-dingir**
13	13	13		**dingir-dingir-dingir** (3×)
13	13	38		**dingir-dingir / uru-bar**
13	13	148		**dingir-dingir / in**
13	13	471		**dingir-dingir / 20**
13	13	480	•	**AN-AN / 1**
			•	**dingir-dingir / 1**
				dingir-dingir-dingir / 1 (3×)
			•	**dingir-dingir / 60-40**
13	13	480B		**dingir-dingir / 80-1**
13	55	231		**DINGIR-la-NI-tum**
13	58			**da-ti-DINGIR-tu**
13	58	142		**DINGIR-tu / i-ti-DINGIR-ni-da-kul**
13	58	167		**DINGIR-tu / GABA-da-mu**
13	69	13		**DINGIR-be dingir-dingir**
13	69	19		**DINGIR-be bù-la-nu^{KI}**
13	69	129A		**DINGIR-be mul-mul**
13	69	206		**DINGIR-be du-du-lu^{KI}**
13	69	312		**DINGIR-be KALAM-tim**
13	69	319	•	**DINGIR-be ga-ga**
			•	**DINGIR-be kà-na-na-im**
				DINGIR-be kà-na-na
13	69	342		**DINGIR-be ma-tum**
13	69	554		**DINGIR-be sal** (2×)
13	70	75		**an-na / nu**
13	70	148		**an-na / in-na-sum**
13	70	331		**an-na / ŠEŠ** (2×)
13	70	342		**an-na-ma-lik** (2×)
13	70	451		**an-na / ar-ra-du-núm**
13	73	61		**DINGIR-ti-mu-tù** (2×)
13	74	13		**bar-an-bar-an** (5×)
13	84	55		**DINGIR-zi-la-su** (2×)
13	94	61		**DINGAR-tim-mu-tù**
13	99	50		**dingir en / ir-é-ak-da-mu**
13	99	401		**AN / en-àr-da-mu**
13	99	461		**DINGIR-en-ki**
				i-ti-DINGIR-en-ki
				šum-DINGIR-en-ki
13	99	586		**dingir en / za-a-šè**
13	106	172		**DINGIR-gú-bí** (2×)
13	106	328		**gú-gi-a-an / gú-ra-kul^{KI}**
13	129A	142		**DINGIR-mul / i-ti-DINGIR-ʾà-da**
13	129A	206		**nì-DINGIR-mul / du-bù-hu-DINGIR-ʾà-da**
13	129A	231		**nì-DINGIR-mul / ì-giš**
13	129A	335		**nì-DINGIR-mul / da-pi$_5$-nu^{KI}**
13	129A	342		**nì-DINGIR-mul / ma-nu-wa-at^{KI}**
13	129A	575		**KA×?-DINGIR-mul / ur-lum^{KI}**
13	142	73		**iš-má-da-ba-an / i-ti-DINGIR-aš-tár**
13	148	296		**dingir-dingir / in / giš-bar-du^{KI}**
13	151	206		**DINGIR-lugal du-du-lu^{KI}**
13	170	342		**DINGIR-am-ma-rí-gú**
13	172	55		**DINGIR-NE-la**
13	211	144		**bar-an-nita-tur**
13	211	598A		**bar-an-nita 5**
13	231	104		**DINGIR-ì-sa-tù**
13	231	296		**na-DINGIR-ì-giš**
13	231	328		**DINGIR-ià-ra-mu**
13	231	335	•	**AN-NI-da-lum**
			•	**DINGIR-ni-da-kul** (56×)
				i-ti-DINGIR-ni-da-kul (20×)
				DINGIR-ni-da-la
13	231	435		**DINGIR-ì-lam** (12×)
				i-ti-DINGIR-ì-lam (2×)
13	231	579		**DINGIR-i-a-é-šu**
13	237	328		**DINGIR-AMA-ra** (44×)
				DINGIR-AMA-ra-sù (34×)
13	252	6		**DINGIR-išḫara / zu-ra-mu^{KI}**
13	252	84		**DINGIR-išḫara / zi-da-ra^{KI}**
				DINGIR-išḫara / zi-da-la
13	252	122		**DINGIR-išḫara / má-NE**
13	252	564		**DINGIR-išḫara / sikil**
13	307	207		**iš-má-da-ba-an MAR-TUM^{KI}**
13	319	5		**DINGIR-kà-pá-tù**
13	319	328		**DINGIR-ga-ra-i-nu**
13	319	342		**ir-DINGIR-ga-ma-al**
13	319	353		**DINGIR-kà-ša-lu**
13	319	427	•	**DINGIR-ga-mi-iš**
				i-ti-DINGIR-ga-mi-iš
			•	**DINGIR-kà-mi-iš**
				i-ti-DINGIR-kà-mi-iš (2×)
13	319	532	•	**DINGIR-ga-me-iš**
			•	**DINGIR-kà-me-iš** (5×)
13	324	97		**ìl-é-ak-da-mu** (3×)
13	324	335		**DINGIR-ʾà-da** (23×)
				da$_5$-hír-DINGIR-ʾà-da
				du-bu-ḫu-DINGIR-ʾà-da (18×)
				i-in-DINGIR-ʾà-da
				i-ti-DINGIR-ʾà-da (2×)
13	328	104	•	**AN-ra-sa-ap^{KI}**
			•	**DINGIR-ra-sa-ap** (45×)
				en-na-DINGIR-ra-sa-ap (3×)
				eb-du-DINGIR-ra-sa-ap
13	328	342		**ìr-DINGIR-ra-ma-lik** (4×)

Fig. VIII - Ebla Sign Index, 1

Hence publication of the Concordance will normally be in the form of microfiche or computer tape.

The other two output formats pertaining to graphemic categorization are designed in part to obviate the problem of bulk: they are produced in the form of indices to the Concordance, utilizing at the same time slightly different sorting criteria. The SIGN INDEX lists all the 3-sign units as in the Concordance, limiting however the context to the word(s) within which each unit occurs. Since no references are cited, and since the context portion is narrower and identical occurrences are collapsed, the resulting size is reduced very considerably: the page reproduced here, for instance, corresponds to 10 pages of the Concordance. The basic information is the same, except that entries cannot be verified against references, for which one will have to refer to the Concordance itself (or, indirectly, to the Word Index, for which see presently). The different sorting, on the other hand, provides some additional valuable insight because it shows at a glance the lexical distribution of the 3-sign units, a dimension which is not immediately apparent in the Concordance.

Fig. VIII

The WORD INDEX lists all words alphabetically, and provides textual references, without context. The Word Index can be used for lexical purposes, even though the sorting is preliminary in nature since there is no morphological analysis. But another function is to serve as a complement to the Sign Index, since one can go from an entry in the Sign Index to the corresponding word in the Word Index and thus obtain the listing of the references - even though for a listing of the contexts one will still have to go to the Concordance.

Fig. IX

2.3. Analysis.

The categorization described above provides a primary analysis of the data - a classification of graphemes, with comprehensive documentation, and fixed type of sorting designed to serve as reference tools for human retrieval. Higher level analysis has been started in the form of frequency computations, with more complex elaborations to follow later.

The basic sorting is in the form of four types of tabulation, where for each tabulation only the first few lines are reproduced. The

EBLA WORD INDEX, 1

O = Onomastics; T = Text

Italics : Semitic; Roman : Sumerian; SMALL CAPS : unknown

1. *a* — **O:4** – *A-a-ba*$_4$: 1 – 6G521 V6; *A-ma-lik* : 1 – 5G1324; *A-sí-piš* : 2 – 56273 R4:2; 6G521
2. Á-A^{KI} — **T:2** – 5G1444 60
3. A-A-BÙ-ÌR-KU$_8$ — **O:1** – 5G336 V3:6
4. A-A-DA^{KI} — **T:2** – 5G188 R4:1, V2:7
5. A-AD-DA-SU — **O:1** – 5G1837
6. ^{d}A-ama — **T:1** – 5G1380
7. *ab* — **O:5** – *A-ban-ià*:1 – 6G521 V2; *A-bi-il*:1 – 5G1329; *Ab-ra-mu*:1 – REO1 23; *A-bu*:2 – 5G1267 V8:3, 5G1359 V2:3
8. AB — **T:3** – 5G1696 15; 5G1766 R3:3; 5G2561:2
9. áb — **T:24** – áb:10 – 5G5313 R4:1; 5G10230 R1:1, R2:1, R2:4; 5G2349 V6:5; 6G189 1, 3, 5, 10, 14; gu$_4$-áb:3 – 5G2349 R2:5, R3:7, V5:2; gu$_4$-áb-niga: 3 – 5G2349 R7:3, R8:2, V2:4; gu$_4$-áb-UD-KEŠDA:1 – 5G2349 V6:4; gu$_4$-gu$_4$-niga-áb-áb:1 – 5G2349 R4:4; gu$_4$-niga-áb:6 – 5G2349 R2:1[,R5:7,R6:5,R9:1, 5,V1:5
10. *aba*, aba — **T:1** – 5G2206; **O:9** – *A-ba*:1 – 5G11153; *A-ba-da-du*:1 – 5G2067; *A-ba-da-an*:1 – 5G11044; *A-ba-il*:1 – 5G1324; *A-ba$_4$-li-im*:1 – 5G1382; *A-ba-ma-lik*:1 – 5G1427; *A-ba-sù*$^{?}$:1 – 5G336 R7:4; Lugal-a-ba$_4$:2 – 5G2263; 6G523 R10:14
11. $^{?}$À-BA-BU^{KI} — **T:1** – 5G2309 V3:4; *A-ba-da-du* — cf. *aba, dādu*; *A-ba-da-an* — cf. *aba, dan*; *A-ba-il* — cf. *aba, il*
12. *A-bala*^{KI} — **T:1** – 5G2309 V2:4; *A-ba$_4$-li-im* — cf. *aba, Lim*; *A-ba-ma-lik* — cf. *aba, Malik*; *A-ban-ià* — cf. *ab*
13. A-BA-RA-NU — **O:1** – 5G1669
14. $^{?}$À-BA-SA^{KI} — **T:1** – 5G1558 R6:3
15. ÁBXAŠ — **T:16** – ABXÁŠ: 11 – 5G220 5, 10, 12, 15; 5G273 R2:1; 5G527 R3:10; 5G1293 V5:1; 5G2075 R11:10, 10; 6G247 6, 6; ABXÁŠ-*sù*: 5 – 6G523 R2:1, 13; R3:14; R4:2, 6
16. *a-ba-šum* — **T:2** – 5G2238 R12:6, 12 (cf. GP *Culto* 1979:27f)
17. A-BA-TI^{KI} — **T:1** – 5G1764 R8:25
18. A-BA$^{?}$-TU — **O:1** – 5G1443 V6:13
19. A-BA-Ù — **O:1** – 5G11104 2:6
20. A-BA-ZU — **O:1** – 5G1569
21. *abba* — **O:2** – *Ab-ba-ì-lum*: 1 – 6G521 R6; *Dam-da-il-ab-ba*: 1 – 5G1396; *Ab-ba-ì-lum* — cf. *abba, ilum*
22. ÁB-BÍ-RU-UM^{KI} — **T:1** – 5G1558 R2:2
23. *abda* — **O:1** – *Ab-da-il*: 1 – 5G1287; $^{?}$Á-BÍ — cf. A-DÈ; *A-bí-il* — cf. *ab, il*

Fig. IX - Ebla Word Index, 1

4. FREQUENCY COMPUTATIONS

4.1. SIGNS (SHAPE)

SIGN	OCCURRENCES	% (CORPUS)	READING	OCCURRENCES	% (SIGN)	% (CORPUS)
001	12	.05	AŠ	9	75.00	.04
			KUM	3	25.00	.01
003	15	.06	MUG	15	100.00	.06
004	1	.00	ZADIM	1	100.00	.00
005	213	.95	BA	208	97.65	.93
			PA2	5	2.34	.02
006	46	.20	GU2	1	2.17	.00
			ZU	45	97.82	.20
007	53	.23	KUŠ	8	15.09	.03
			SU	45	84.90	.20
009	26	.11	BAL	5	19.23	.02
			BALA	21	80.76	.09
010	57	.25	GIR2	57	100.00	.25
012	30	.13	KUS	14		.06
			ZAD	16		.07

4.2. SIGNS (FREQUENCY)

SIGN	OCCURRENCES	% (CORPUS)	READING	OCCURRENCES	% (SIGN)	% (CORPUS)
480	897	4.01	1	852	94.98	3.81
			60	43	4.79	.19
			DIŠ	2	.22	.00
537	860	3.85	UDU	721	83.83	3.23
			LU	137	15.93	.61
			DIB	2	.23	.00
335	720	3.22	DA	653	90.69	2.92
			TA2	64	8.88	.28
			XA	3	.41	.01
013	682	3.05	DINGIR	501	73.46	2.24
			AN	176	25.80	.78
			IL3	5	.73	.02
536	669	2.99	TUG2	489	73.09	2.19
			ŠE3	91	13.60	.40
			KU	54	8.07	.24
			DIB2	10	1.49	.04
			TUŠ	9	1.34	.04
			GUS	5	.74	.02
			DUL5	1	.14	.00

READINGS (ALPHABETICALLY)

-4.3. READING	OCCURRENCES	% (CORPUS)	% (SIGN)	SIGN	OCCURRENCES
*A	1	.00	100.00	357	1
*A3	118	.52	30.56	324	386
A	411	1.84	98.32	579	418
A2	4	.01	66.66	334	6
AB	51	.22	46.78	128	109
AB2	30	.13	100.00	420	30
AD	16	.07	7.80	145	205
AG	4	.01	4.65	097	86
AGRIG	4	.01	10.00	452	40
AH	4	.01	33.33	398	12
AK		.28	73.25	097	86
'A		.09		097	6

READINGS (FREQUENCY) PAGE 826

-4.4. READING	OCCURRENCES	% (CORPUS)	% (SIGN)	SIGN	OCCURRENCES
1	852	3.81	94.98	480	897
UDU	721	3.23	83.83	537	860
DA	653	2.92	90.69	335	720
KI	610	2.73	100.00	461	610
2	572	2.56	100.00	570	572
MA	535	2.39	100.00	342	535
MU	534	2.39	99.62	061	536
DINGIR	501	2.24	73.46	013	682
TUG2	489	2.19	73.09	536	669
'A		1.96	100.00		438
		1.	.00		434

Fig. X - Frequency Computations

first two tabulations list graphemes from the viewpoint signs, the first according to the standard sequence of cuneiform signs by shape, and the second in descending order of frequency by sign. The remaining two tabulations list graphemes from the viewpoint of their value, whether phonological or logographic, first alphabetically and then in descending order of frequency. The graphemic significance of this analysis is that it makes possible for the first time the study of the distribution of frequency ranges within large corpora and the comparison of the same corpora among each other at the level of their graphemic inventory (see below, under 2.4.2 for details referring to the Ebla materials).

Besides computations at the level of the inventory, graphemic analysis can be undertaken at the level of the text with the tools described so far. An interesting application is the one which can be defined as COVARIATION RULE. If one observes all the various cases of alternate values for the same 3-sign unit (marked by subdivisions into sub-units within the 3-sign sequence in the Concordance and by bullets in the Sign Index), it will appear that they are generally rare (the apparently high incidence in figs. 000 and 000 is simply due to the fact that the corpus includes texts with transliterations from different sources, left intentionally without normalization, so that the different values reflect alternate choices by different scholars, rather than contrastive graphemic values). The rule may be formulated as follows: in a homogeneous corpus, a sequence of 3 signs reduces to a minimum the polivalence possibilities of each individual sign. The example shown in Fig. XI is one of the very few exceptions, in the Ebla corpus studied here, where a 3-sign unit may possibly be read in three different ways. It is because of this rule that the 3-sign unit has been chosen as the basic unit for the sign categorization system. (The Covariation rule described here is similar in principle to the notion of algorhythm proposed by Reiner 1973).

Fig. X

2.4. Ebla Graphemics: Preliminary Substantive Considerations.

2.4.1. Text level.

The presentation of graphemic analysis given so far, while based for

5. COVARIATION RULE

```
                                                AR3-DA-MU                                    401     335     061
EPNS            16    EN-AR-LI-IM / EN-AR-LI-IM / EN-AR3-*DA-MU / EN-AR3-*DA-MU / EN-A
EPNS            17   N-AR-LI-IM / EN-AR3-*DA-MU / EN-AR3-*DA-MU / EN-AR3-*DA-MU / EN-A
EPNS            18   AR3-*DA-MU / EN-AR3-*DA-MU / EN-AR3-*DA-MU / EN-AR3-*DA-MU / EN-A
EPNS            19   AR3-*DA-MU / EN-AR3-*DA-MU / EN-AR3-*DA-MU / EN-AR3-*DA-MU / EN-A
EPNS            20   AR3-*DA-MU / EN-AR3-*DA-MU / EN-AR3-*DA-MU / EN-AR3-MA-LIK / EN-B
5G.01764R10     21   DU / *DINGIR-*AMA-RA / *AN / EN-AR3-DA-MU / 1 UDU / *DINGIR-*AMA-
5G.02075R 6      4   U8 / 2 UDU / IZI-GAR / *AN / EN-AR3-DA-MU / 1 UDU / *DINGIR-ZI-LA
5G.02238V 1     9'   DU / *DINGIR-*AMA-RA / *AN / EN-AR3-DA-MU / 1 UDU / *DINGIR-*AMA-

                                                HIR2-DA-MU
EPNS            48   HIR-MA-LIK / DA-HIR-MA-LIK / DA-HIR2-DA-MU / DA-HIR2-DA-MU / DA-M
EPNS            49   HIR-MA-LIK / DA-HIR2-DA-MU / DA-HIR2-DA-MU / DA-MA-NA / DA-MA-NA-

                                                MUR-DA-MU
EPNS            64   MU-RA / A-MU-RI3 / A-MU-TU3 / A-MUR-DA-MU / A-NA-GA-LU / A-NA-*LU
```

Fig. XI - Covariation Rule

the sake of illustration on Ebla material, is in fact of a general import: the tools described are suitable for graphemic analysis on any cuneiform corpus, with but a few adaptations and changes. We will now turn to Ebla graphemics in a more specific sense, showing how the tools of graphemic categorization and analysis may be applied to individual problems. We will look at graphemics first on the text level, and then at the level of the inventory.

As indicated above, special values for given signs may be conditioned by lexical or sequential factors: we will call these PHONO-LEXICAL and PHONO-TACTICAL VALUES. As an example of the former we may consider the geographical name Du-lu^{KI}. The sign DU has been read GUB by G. Pettinato, and the name has been understood as /Gublu/ for "Byblos". The following may be said here from a graphemic point of view. The sign is common at Ebla since it occurs 206 times in our corpus, and nowhere does the value GUB fit the context, except possibly for the word here in question. Structurally, this particular value is possible: there are 53 CVC signs attested in the Semitic portion of our Ebla corpus, and even though only seven of these have a value with a final stop, four of these end in fact in B. The graphemic data, then, tell us that structurally the value GUB is possible, but that distributionally it would be acceptable only as a phono-lexical value, assuming that the reading GUB-*lu*^{KI} can be otherwise be proven correct (which seems unlikely, cf. A. Archi: *SEb* 2, p. 3).

As an example of phonotactical values we may refer to a case already noted by P. Fronzaroli. He rejects (*RAI* 20, §2; *SEb* 1, p. 11) the value /li/ for NI in the preposition NI-*na* (which he reads *i-na*), because the sign NI is never otherwise attested with the value *li* in word initial position. Statements of this type will become more and more difficult to verify against the data in the measure in which the volume of published texts increases: negative evidence, i.e. the non-existence of a given graphemic fact, is difficult to claim without the type of sorting provided by the Concordance of the Sign Index. With these tools, on the other hand, the statement by Fronzaroli can be verified and quantified at a glance: out of 545 occurrences (in our corpus) of the sign NI, 210 occurrences are in word initial positon, and nowhere with a demonstrable value /li/.

6. GRAPHEMIC ALTERNANTS

	(OCCURRENCES)	(%CORPUS)	(%SIGN)	(SIGN)	(OCCURRENCES)
I	287	1.28	100.00	142	287
I3	147	.65	26.97	231	545
IA3	110	.49	20.18	231	545
WA	84	.37	67.20	383	125

	/ya/				
	i-a(C)	*ià-*{CV, #}	*ì-a(C)*	*ia*$_8$(WA)	*Ci-a(C)*
initial	*i-ad-da-mu* *i-a-da-mu*	—	*ì-ap-KI* *ì-a-ba* etc. (10x)	—	?
medial	*ma-i-at*	*d ià-ra-mu* *i-ti-ià-ma-lik*	*d ì-a-é-šu* *a-ì-a-la-du*	*gú-gi-WA-an* *gu-gi-WA-an*	*gú-gi-a-an* *gú-gi-a-nu*
final	—	*en-na-ià* etc. (87x)	*iš$_{11}$-ru$_{12}$-ì-a* *i-du-ì-a*	—	?

Fig. XII - Graphemic Alternants

Fig. XII

A type of GRAPHEMIC PARADIGM includes sets of GRAPHEMIC ALTERNANTS for the same phonemic configuration. An example is presented here in Fig. 000, which contains first an excerpt from the frequency computations drawn from our corpus, and then a graphic paradigm of alternate graphemic renderings of the same phoneme cluster (/ya/), according to positional classes (word initial, medial or final position). From our corpus it appears that initial /ya/ is not expressed by the sign NI alone - except that, it turns out, this value is actually found in other texts not yet entered in our data base (see P. Fronzaroli: *SEb* 1, p. 72). It also appears that final /ya/ is not expressed by the signs I-A, and that the sign WA can be used with the value /yà/ which is otherwise rare and later in time. Again, such paradigms can be drawn up easily and exhaustively from the Concordance and the Sign Index, thus obtaining with a rapid manual search complete patterns of cooccurrence and distribution.

Fig. XIII

Just as easily one can obtain another type of paradigm, the one illustrated here in Fig. XIII, which exhibits various examples of (graphemic) CONCATENATORY RULES. By scanning visually, in the Concordance (from which the excerpt of Fig. VII is taken), or in the Sign Index, through entries pertaining to (CVC)CVC signs, one can rapidly obtain all instances where two identical consonants follow each other at sign boundary. The graphemic question may then be asked as to whether the ...C_1-C_1... sequence is used to express long consonant or not. The examples selected in Fig. XIII exhibit cases where such sequence is in fact indicative of phonemic length, but rather serves the function of phonological indicator for either Semitic words (*Ì-a-bur-ru*$_x$, *Tár-kab-bù*, *Puzur*$_4$*-ra-*) or Sumerian determinatives (*-din gir-ra-*).

2.4.2. Inventory Level.

If we look now at the inventory of the Ebla corpus as a structural whole, alongside the inventories of other cuneiform corpora, we may compare the distribution of frequency ranges within the various corpora, and obtain thereby an insight into the economy of the graphemic system proper to each corpus. A similar approach has already been described for the corpora other than the Ebla corpus in the writer's contribution to the Schaeffer Volume (see below, 3. *Biblic*

		RA-MA-LIK	328	342	575
EPNS	23	MU / IR3-RA-DI-MU / IR3-*DINGIR-RA-MA-LIK / IR3-*DINGIR-RA-MA-LIK			
EPNS	24	*DINGIR-RA-MA-LIK / IR3-*DINGIR-RA-MA-LIK / IR3-*DINGIR-RA-MA-LIK			
EPNS	25	*DINGIR-RA-MA-LIK / IR3-*DINGIR-RA-MA-LIK / IR3-*DINGIR-RA-MA-LIK			
EPNS	26	*DINGIR-RA-MA-LIK / IR3-*DINGIR-RA-MA-LIK / IRO-KAB-DU-LUM / IŠ-A			
EPNS	76	2-IŠ / IG-RI-SU / IG-SU-UD / IH-RA-MA-LIK / IH-RA-MA-LIK / IH-SU-			
EPNS	77	/ IG-SU-UD / IH-RA-MA-LIK / IH-RA-MA-LIK / IH-SU-UP-DA-MU / IK-B			
EPNS	83	-U3-MA / PUZUR4-RA-IA3 / PUZUR4-RA-MA-LIK / PUZUR4-RA-MA-LIK / PU			
EPNS	84	IA3 / PUZUR4-RA-MA-LIK / PUZUR4-RA-MA-LIK / PUZUR4-RA-MA-LIK / PU			
EPNS	85	_IK / PUZUR4-RA-MA-LIK / PUZUR4-RA-MA-LIK / PUZUR4-RA-MA-LIK / PU			
EPNS	86	LIK / PUZUR4-RA-MA-LIK / PUZUR4-RA-MA-LIK / PUZUR4-RA-MA-LIK / PU			
EPNS	87	_IK / PUZUR4-RA-MA-LIK / PUZUR4-RA-MA-LIK / PUZUR4-RA-MA-LIK / PU			
EPNS	88	LIK / PUZUR4-RA-MA-LIK / PUZUR4-RA-MA-LIK / PUZUR4-RA-MA-LIK / QA			
EPNS	89	LIK / PUZUR4-RA-MA-LIK / PUZUR4-RA-MA-LIK / QA2-TA2-NA / QA2-TA2-			
5G.00336R 1	14	HU / IŠ-LA-IA3 / NAGAR / PUZUR4-RA-MA-LIK / UGULA 5 / MI-KA-IA3 /			
5G.00336R 7	10	/ NI-E2-I3-LU / AR-ŠE3 / PUZUR4-RA-MA-LIK / UGULA 5 / AŠ2-BA-NI /			
5G.01332R 1	2	2 MA-NA 50 KU3-GI / PUZUR4-RA-MA-LIK / I3-NA-SUM / *TAR KU3-			
5G.01332V 1	1	3 / A-*BI2 20-4 KU3-GI / PUZUR4-RA-MA-LIK / I3-NA-SUM / I3 MA-NA			
5G.01332V 2	3	-NA *TAR 8 KU3-GI / AL / PUZUR4-RA-MA-LIK			
5G.01333R 2	1	I2 / 20-4 KU3-GI 2? 1D / PUZUR4-RA-MA-LIK / I3-NA-SUM / 3 MA-NA 8			
5G.01333V 2	3	4 *ŠA-*PI KU3-GI / AL / PUZUR4-RA-MA-LIK			
5G.01345R 6	10	*MI2 *TUG2 / *PA-*KAS-*DU / *AH-RA-MA-LIK / E2 MA-NA? / 1 *TUG2 *			
5G.01764R 5	25	IR-*AMA-RA / 2 UDU / *DINGIR-KU-RA / MA-LIK-TUM / NIDBA0 / IN UD			
5G.02075R 4	23	M / NIDBA0 / 5 UDU / *DINGIR-KU-RA / MA-LIK-TUM / NIDBA0 / IN / M			
5G.02238R 6	24	IR-*AMA-RA / 3 UDU / *DINGIR-KU-RA / MA-LIK-TUM / NIDBA0 / IN / *			
5G.02238V 7	12	UDU / DAM / ZU-HU-A-NU / PUZUR4-RA-MA-LIK / ŠU-DU8 / IN UD / *DIN			
		RA-A-GU2	328	579	106
5G.00336V 2	5	A / IR3-NA-*DINGIR-*IM / PUZUR4-RA-A-GU2 / *UR-*PI / E2-DA-ŠA / *			
		BUR-RU0-I3	349	099	231
EPNS	32	BE / I3-A-*BE / I3-A-*BE / I3-A-BUR-RU0 / I3-A-DU / I3-A-DU-UD /			
		KAB-BU3-IN	088	019	148
5G.02238R10	4	A-SA-AP / *A3-TA2-NI-*KI / TAR2-KAB-BU3 / IN UD / *DINGIR-*AMA-RA			
		KAB-BU3-2	088	019	570
5G.02238R 9	29	DI3-TU3 / 4 UDU / E2-NUN / TAR2-KAB-BU3 / 2 UDU / *DINGIR-RA-SA-A			
		KAB-BU3-NIDBA0	088	019	999
5G.02238R10	10	IR-IŠHARA / ZU-RA-MU-*KI / TAR2-KAB-BU3 / NIDBA0 / 1 UDU / *DINGI			
5G.02238R10	15	/ *DINGIR-UTU / *EN-*LI / TAR2-KAB-BU3 / NIDBA0 / MI-NA-*NI / ŠU			

Fig. XIII - Concatenatory Rules

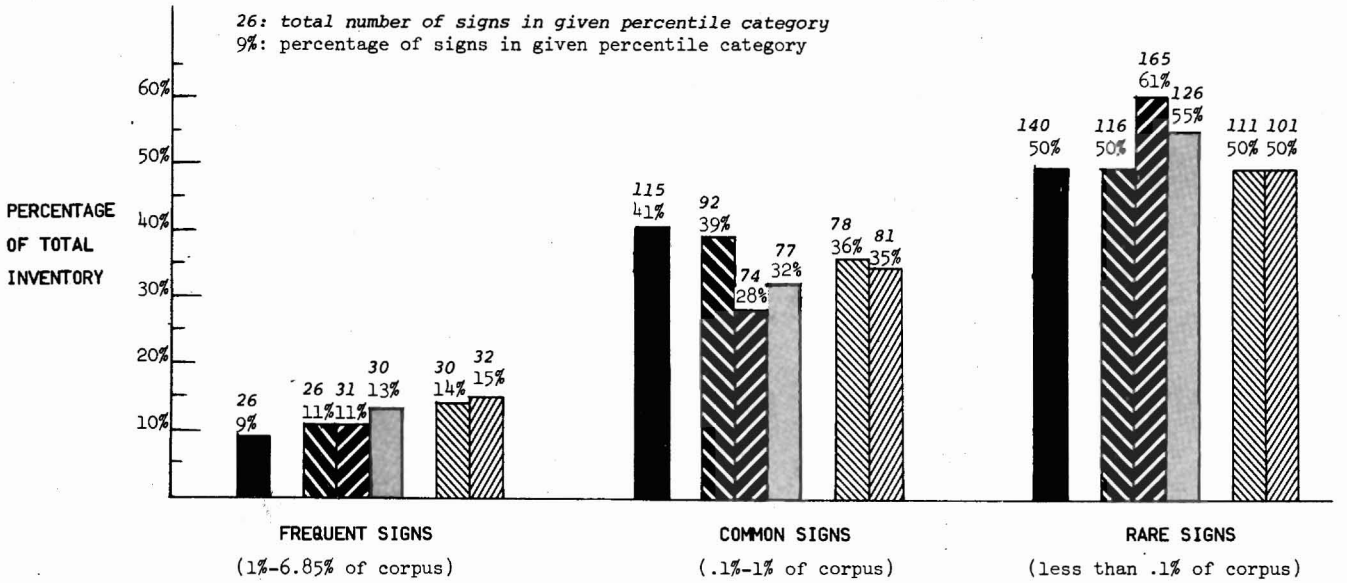
graphical Note). The comparative corpora are listed in Fig. 000. They are heterogeneous vis-à-vis the Ebla corpus, but homogeneous in themselves: they are letters from different areas (Babylonia proper, the North and the West) and different periods (Old Babylonian and the Amarna Age); they are also uneven in size (see the total number of sign occurrences listed for each corpus in Fig. 000). All of this means that there is sufficient "dispersion" to give greater validity to the results, especially the (somewhat surprising) degree of uniformity among the various graphemic inventories.

Fig. XIV The first set of figures pertains to the overall FREQUENCY DISTRIBUTION among corpora. Here the total sign inventory of the six corpora is divided into three categories, which I call the categories of Frequent, Common and Rare signs. The numeric parameters used to define these categories have been set in such a way that frequent signs are defined as the ones whose occurrences cover more than 1% of the corpus; common signs are the ones whose occurrences cover between .1% and 1% of the corpus; and rare signs are the ones whose occurrences cover less than .1% of the corpus. It is apparent from the bar histograms in Fig. 000 that the distribution of frequency ranges remains even in all corpora (even though the actual signs corresponding to those ranges vary considerably from corpus to corpus, as we shall see presently). This may imply that a somewhat uniform principle of economy is at work in all corpora here envisaged - a principle whereby it appears that the thresholds from rarer to more frequent signs remain somewhat constant regardless of either the inventory or the corpus; in terms of the parameters chosen here, about half of the signs used are rare in every corpus considered, and about 10% are frequent.

Fig. XV The INVENTORY OVERLAY tabulated in Fig. XV answers a different question. Given the fact that the various corpora exhibit, each taken by itself, a certain parallelism in their intrinsic graphemic economy, to what extent do they overlap in terms of the actual signs employed? Interestingly, the coincidence factor is greater than might be expected. To this end, I have compared seven corpora, adding Old Akkadian to the six corpora already described, because of its proximity to the Ebla corpus. Since in this particular tabula

8. FREQUENCY DISTRIBUTIONS:

MAJOR PERCENTILE CATEGORIES, WITH COMPARATIVE DATA



Ebla: current corpus (total signs: 22,321 / total inventory: 283)

Old Babylonian: royal letters from Babylon (30,711 / 234)

ABB 1-5 excluding royal letters (96,612 / 275)

Mari letters (153,308 / 230)

Western Akkadian: Akkadian letters from Amarna (57,502 / 221)

Akkadian lettes from Ugarit (12,952 / 214)

Fig. XIV - Frequency Distribution

9. INVENTORY OVERLAY AND COINCIDENCE FACTOR

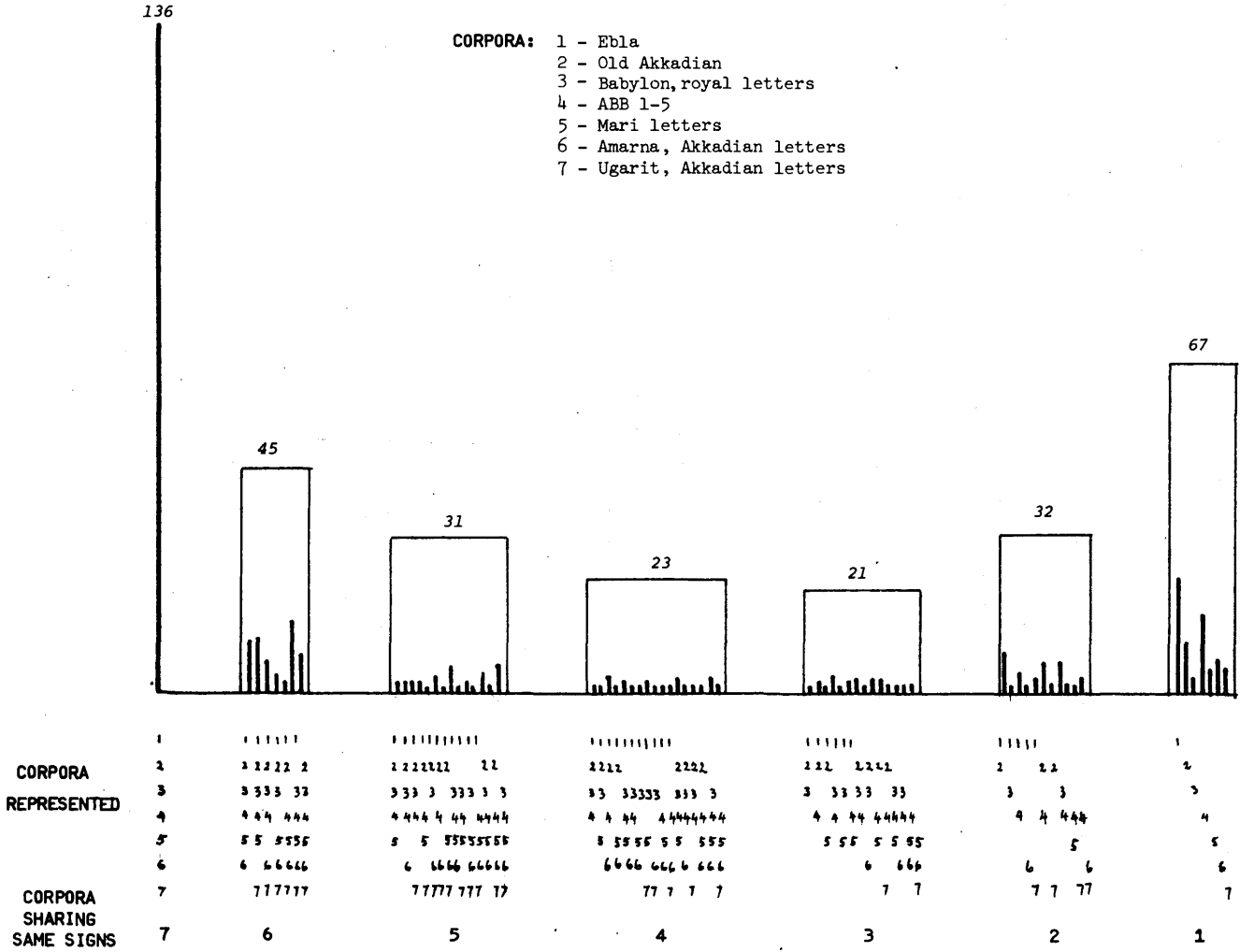


Fig. XV - Inventory Overlay and Coincidence Factor

tion I have considered the various inventories as individual wholes and have not taken frequency ranges into account, Old Akkadian could be included on the basis of the data found in *MAD* 3. The total combined inventory of all seven corpora is 355 signs; 30 of these are number figures, found mostly in the Ebla corpus. Out of this grand total, 136 signs are shared by all seven corpora, and 45 by six corpora; this means that about half of the signs are found in all or most of the corpora. If we consider the ones that are shared by a smaller group of corpora, we can find no evidence of particular clusterings; note especially that there is nothing special to bracket corpora which are in fact closer among themselves, i.e. Ebla and Old Akkadian, the Old Babylonian corpora and the Syrian Akkadian corpora.

It is therefore at a more differentiated level that we have to search for the evidence of contrastive factors among the corpora. We may do so by reintroducing a consideration of frequency ranges, applied this time to individual signs. The picture which emerges is given graphically in Fig. XVI, which tabulates the GRAPHEMIC PROFILE of the 25 most common signs in the six corpora. Besides the comments made already by the writer in the article for the Schaeffer Volume, we will note here, with regard to the new data concerning Ebla, how the latter is generally more at variance with the remaining five corpora than any of these corpora: the Ebla percentiles are generally at one extreme or the other of the bars in the illustration and often at a great distance from the remaining corpora. Admittedly, the graphemic values are mixed, since the graphemic profile produced here refers to signs, rather than values, but the results are certainly indicative of the possibilities of the method. Its discriminatory power is such that it can be expected it may serve, when applied on a narrower scale and with automatic searching procedures, to identify GRAPHEMIC CORPORA, i.e. corpora which share common and distinctive graphemic peculiarities, as different from corpora which are so identified on the basis of linguistic, prosopographic, archaeological or other criteria.

Fig. XVI

3. Bibliographical Note.

The more important recent publications on cuneiform graphemics in

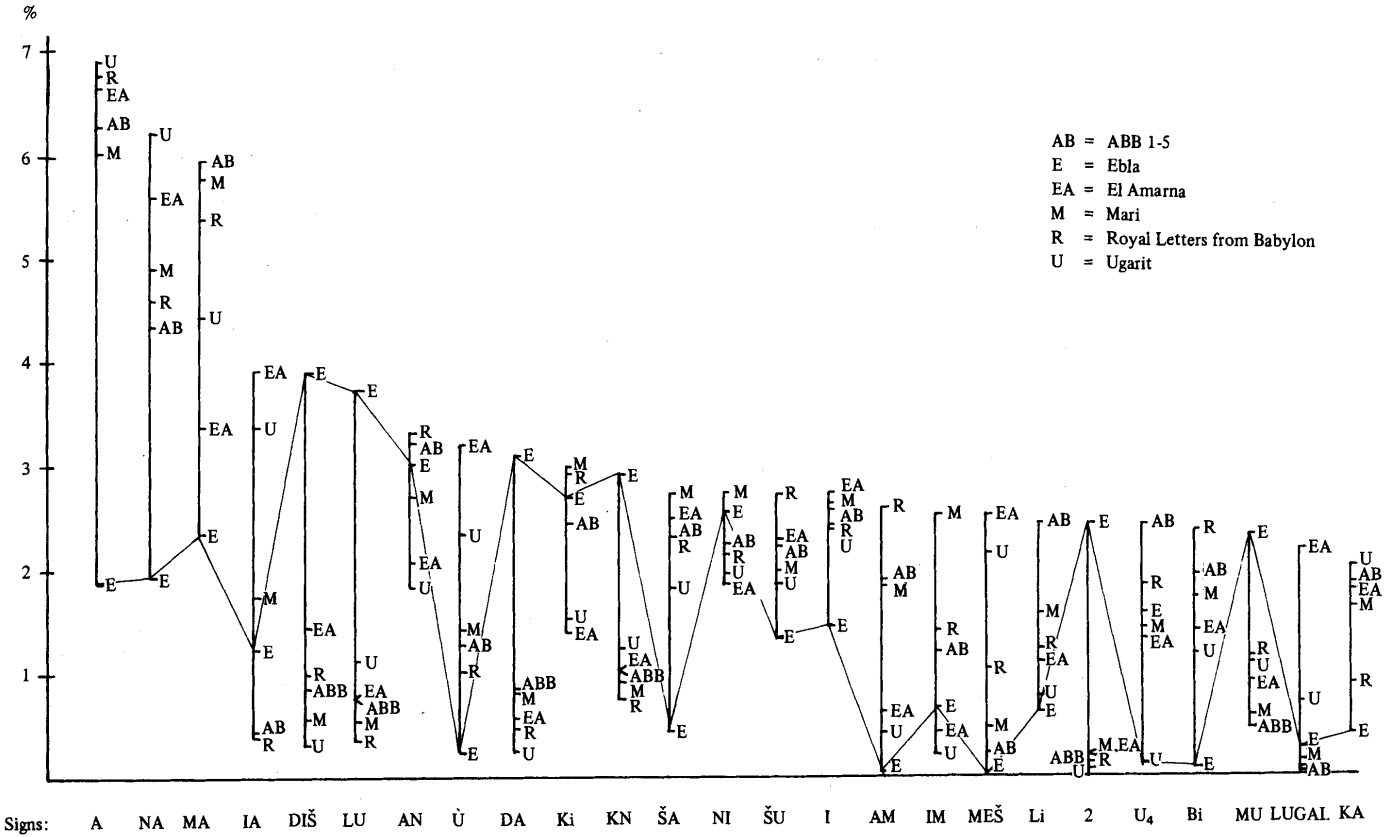


Fig. XVI - Graphemic Profile (25 Most Common Signs)

clude:

Civil, M. 1973 *The Sumerian Writing System: Some Problems: Approaches to the Study of the Ancient Near East: Or.* NS 42 and Los Angeles, Gelb Volume, 21-34.

Gelb, I.J. 1970 *Comments on the Akkadian Syllabary: Or.* NS 39, 516-46.

Reiner, E. 1973 *How We Read Cuneiform Texts: JCS* 25, 3-58.

Renger, J. 1971 *Überlegungen zum akkadischen Syllabar: ZA* 61, 23-43.

Von Soden, W. and W. Röllig, *Das akkadische Syllabar*, Rome 1976[3].

Borger, R. *Assyrisch-babylonische Zeichenliste*, Neukirchen 1978.

Concern for Ebla graphemics has been especially voiced by:

Fronzaroli, P. 1979 *La contribution de la langue d'Ebla à la connaissance du Sémitique archaïque: RAI* 20.

On the research project Computer Aided Analysis of Mesopotamian Materials, directed by the writer, see:

Buccellati, G. 1977 *The Old Babylonian Linguistic Analysis Project: Goals, Procedures and First Results: Proceedings of the International Conference on Computational Linguistics, Pisa 1973*, 385-404.

Buccellati, G. 1979 *Comparative Graphemic Analysis of Old Babylonian and Western Akkadian*, Schaeffer Volume: *UF* 11, 89-100.

Buccellati, G. (ed.) 1979- *Cybernetica Mesopotamica*, Malibu.

Gaebelin, P.W. 1977 *Graphemic Analysis of Old Babylonian Letters from Mari*, Ph.D. Dissertation, Malibu.

Kelly-Buccellati, M. 1977 *Towards the Use of Quantitative Analysis in Mesopotamian Sphragistics: Mesopotamia* 12, 41-48.

Kelly-Buccellati, M. forthcoming *Iconographic and Formal Analysis of Mesopotamian Cylinder Seals: A System Description: Proceedings of the Ugarit Conference*, Lattakiah 1979.

Kobayashi, Y. 1975 *Graphemic Analysis of Old Babylonian Letters from Southern Babylonia*, Ph.D. Dissertation, UCLA.

4. Acknowledgements.

Data entry for the Ebla corpus utilized in this article was done under the primary responsibility of Matthew L. Jaffe and Judith R. Paul, as was the operation of the pertinent programs. The analysis of the data benefitted from a graduate seminar given by the writer at UCLA in the Winter and Spring 1980, at which some ten graduate students contributed with their research and participation.

Initial programming for graphemic analysis was the work of John Settles, but the active implementation of the various programs which are currently utilized for our research is the work of David Holz gang, assisted by Judith Paul.

Institutional and financial support for this research came from the National Endowment for the Humanities, from the Office for Academic Computing at the University of California, Los Angeles, from the In ternational Institute for Mesopotamian Area Studies, and from the Ambassador International Cultural Foundation.