in C. Saporetti, *The Middle Assyrian Laws,* Vol. 2, Malibu: Undena Publications; 1984

fonrt matetr at the end
please use bookmarks

INTRODUCTION

by
Giorgio Buccellati

1. Cybernetica Mesopotamica (CM)

1.1 Preliminaries

It is with special pleasure that the second volume in the series Graphemic Categorization bears the name of my friend and colleague Claudio Saporetti. When the project "Computer Aided Analysis of Cuneiform Texts" began in 1968, it was conceived as a collaborative project open to all scholars interested in data processing applications to Mesopotamian materials. The metaphor of a "Data Bank" was taken literally to reflect an active exchange of materials, in such a way that users would make "deposits" and "withdrawals," contributing thereby to the general enrichment of the field while obtaining at the same time a better utilization of their own data. Claudio Saporetti was the first colleague to develop a close working collaboration with us on this project, adopting and contributing to our standards, adding a vast amount of materials to the data base, and graciously accepting operational delays and system modifications: for his confidence ever since we first formalized our collaboration in 1975 I wish to express to him my sincere gratitude. The Middle Asyrian materials stored in the Data Bank are growing rapidly, and the publication program will continue apace. In the series DSC the next Middle Assyrian corpus will be the texts of the archive of Babu-aḫa-iddina (by H. Freydank and C. Saporetti) and the Assur 14327 Archive (by J. N. Postgate); in the series GC the next Middle Assyrian corpus will be the texts of Tell Billa, to be followed by the texts of selected archives of Assur.

A fuller introduction to the system as a whole will be found in the first volume of this series (GC 1). Here only a few relevant portions of that introduction will be reproduced in order to facilitate the understanding of the formats presented in this volume. The various corpora chosen for the initial volumes of the series are limited in size, a factor which admittedly lessens the usefulness of any concordance: nevertheless, these first volumes will serve to establish in a practical way the scope and goal of graphemic documentation as I have undertaken it, and thus serve as an introduction to the larger corpora which will be presented in later volumes. (For a brief presentation of the nature and goals of graphemic analysis I refer to a special section in the introduction to GC 1, and to the pertinent titles quoted below in section 3.) Needless to say, smaller corpora such as the one of the Middle Assyrian Laws may be integrated at any time, electronically,

within larger corpora of practically unlimited size: thus it will be possible to have at a later date new publications of larger corpora in the same format as the one presented here, or specialized retrieval passes available on demand, or again the full corpora in electronic media.

Many individuals and institutions shared in different ways in the effort to bring this project to fruition. The institutions which have made it possible for us to continue our work over a period of years are listed on the acknowledgment page: to them goes my warmest appreciation for their unfailing support. The individual contributors are mentioned on the title page, since the authorship of a volume of this kind is in fact the result of a very carefully orchestrated scholarly and technical collaboration. In particular, it should be noted that C. Saporetti has full responsibility for the philological part of the work, including the data entry which was done under his direct supervision (the data were sent on tape from Pisa to Los Angeles); I bear on the other hand full responsibility for the choices made with regard to data processing in general and format in particular. In this respect I have made a deliberate effort to try to achieve a proper balance between the amount of information to be published, the degree of differentiation with which it is to be presented and the physical size of the resulting hard copy.

1.2 Graphemic Categorization (GC).

The series Graphemic Categorization (abbreviated GC), is subdivided into five parts or components:

A Sign Index
B Frequency Tabulations (or SIFT for the two parts taken together)
C Word Index with References (WIRE)
D Sign Sequence Graphemic Concordance (SC)
E Word Sequence Graphemic Concordance (WC)

Each component presents the same data from a different perspective. The two concordances include the immediate context within which each sign occurs, while the two indices provide only the elements in question. The Sign Concordance and the Sign Index are based on a sorting by signs which prescinds from morphemic and other boundaries. The Word Concordance and the Word Index, on the other hand, sort the signs with reference to word boundary. Chart 1 presents a graphic synthesis which illustrates the general configuration of Graphemic Categorization as a system and the specific layout of its various components. Note in particular that, while the Word Index is produced exactly with the same sorting criteria as the Word Concordance, the Sign Concordance and the Sign Index differ from each other--as will be explained more in detail in the next sections.

Normally, only three of these components of Graphemic Categorization will be published in book form for any given corpus, namely the Sign Index, the Frequency Tabula-

tions and the Word Index. Together, they provide the most compact and yet comprehensive coverage for graphemic purposes, while at the same time providing a preliminary word concordance. Such is the case for the Middle Assyrian Laws. Only a short sample of the first pages of both the Sign Concordance and the Word Concordance is added here, to provide a more specific illustration of what both formats have to offer. Computer printouts or magnetic tapes of either concordance, whether in full or for selected portions, are available at cost through Undena Publications.

The essential details of format for each of the five components of Graphemic Categorization are explained in the following sections. I begin with a description of the Sign and Word Concordances because logically they come first, even though in the normal sequence of publication these are either missing or are in last place: the indices are in fact derived from the Concordances, although the Sign Index exhibits some significant differences from the Sign Concordance, as we shall see.

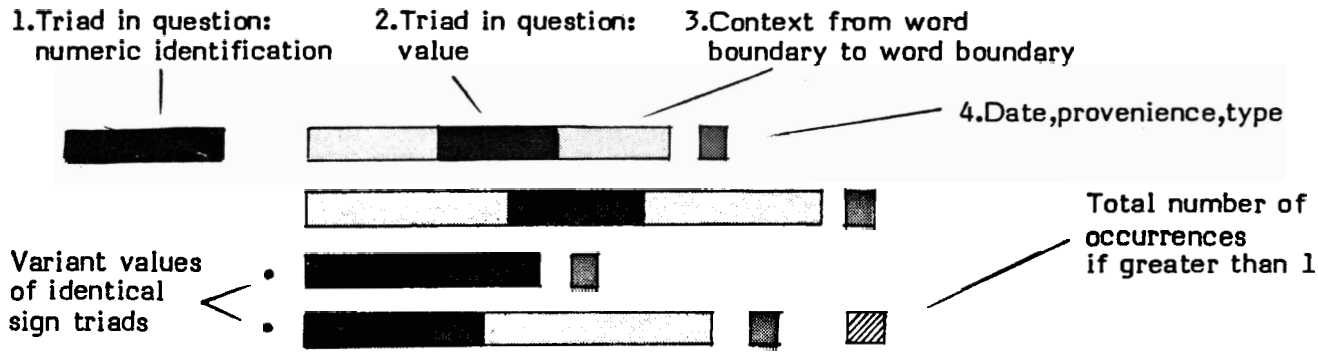
1.3 Sign Concordance

The largest portion of the series GC consists of a concordance of all cuneiform signs occurring in a given corpus, sorted according to the shape of the cuneiform signs. The concordance is subdivided into sections which correspond in number to the number of signs attested in the corpus. The beginning of each section in the concordance is marked by a change in the numerical sequence of signs according to conventional Assyriological usage. Where a sign is not listed in a standard sign list, its place is determined by the normal rules followed in arranging cuneiform signs according to their shape; in such cases no new number is interpolated into the conventional numbering system; rather, an alphabetical index is added to the number immediately preceding in the list.

Each section is divided into units which include all passages with the same three signs combination: we call such a combination a "triad". The unit is marked off by the transliterated value of the triad in the center and the numerical value of the triad on the right. If the same triad occurs with different values for any one or more of the component signs, then the transliteration in the center is repeated each time the values change, whereas the numerical value on the right is given only for the first occurrence of the triad.

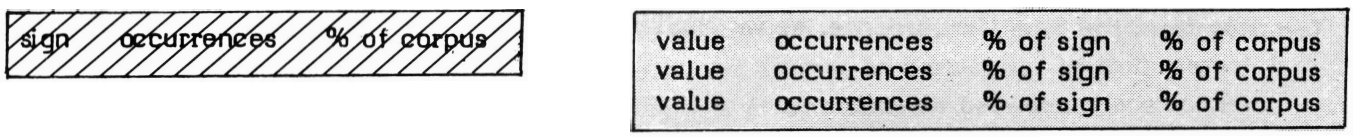
Each unit is subdivided vertically into three major columns. The sorting sequence does not correspond to the sequence of columns from left to right, but rather goes generally in the inverse direction from right to left, as indicated in Chart 1; here, however, we will describe the columns as they appear from left to right. The first column includes, where applicable, information extrinsic to the graphemic data proper, such as date, provenience or type of text; this information is not applicable to the Middle Assyrian Laws, and thus it is not found in this volume. The second column includes the textual references. The third column contains the relevant textual portions, cut at both the left and the right margins without regard for context, and with the triad in question at the

A. SIGN INDEX

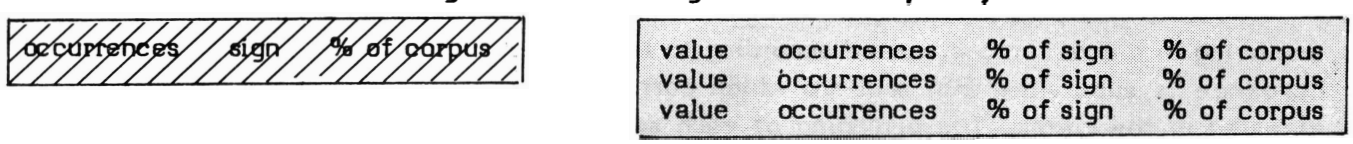


B. FREQUENCY TABULATIONS

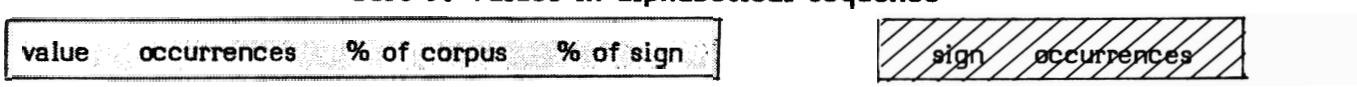
Sort 1. Signs in graphic sequence



Sort 2. Signs in descending order of frequency



Sort 3. Values in alphabetical sequence



Sort 4. Values in descending order of frequency

C. WORD INDEX

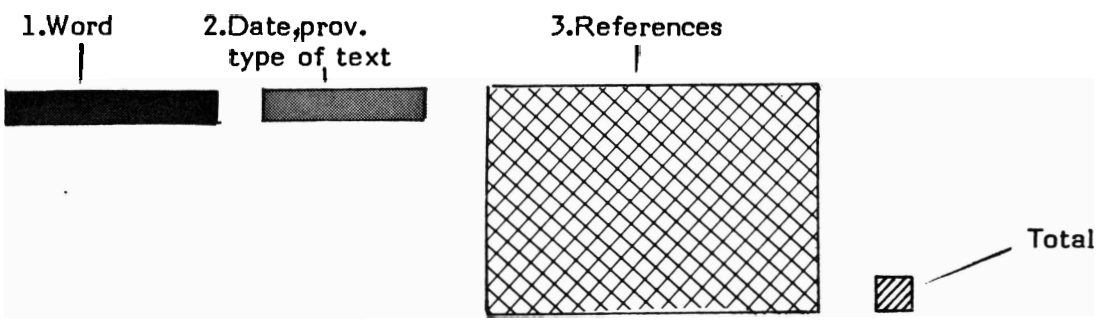
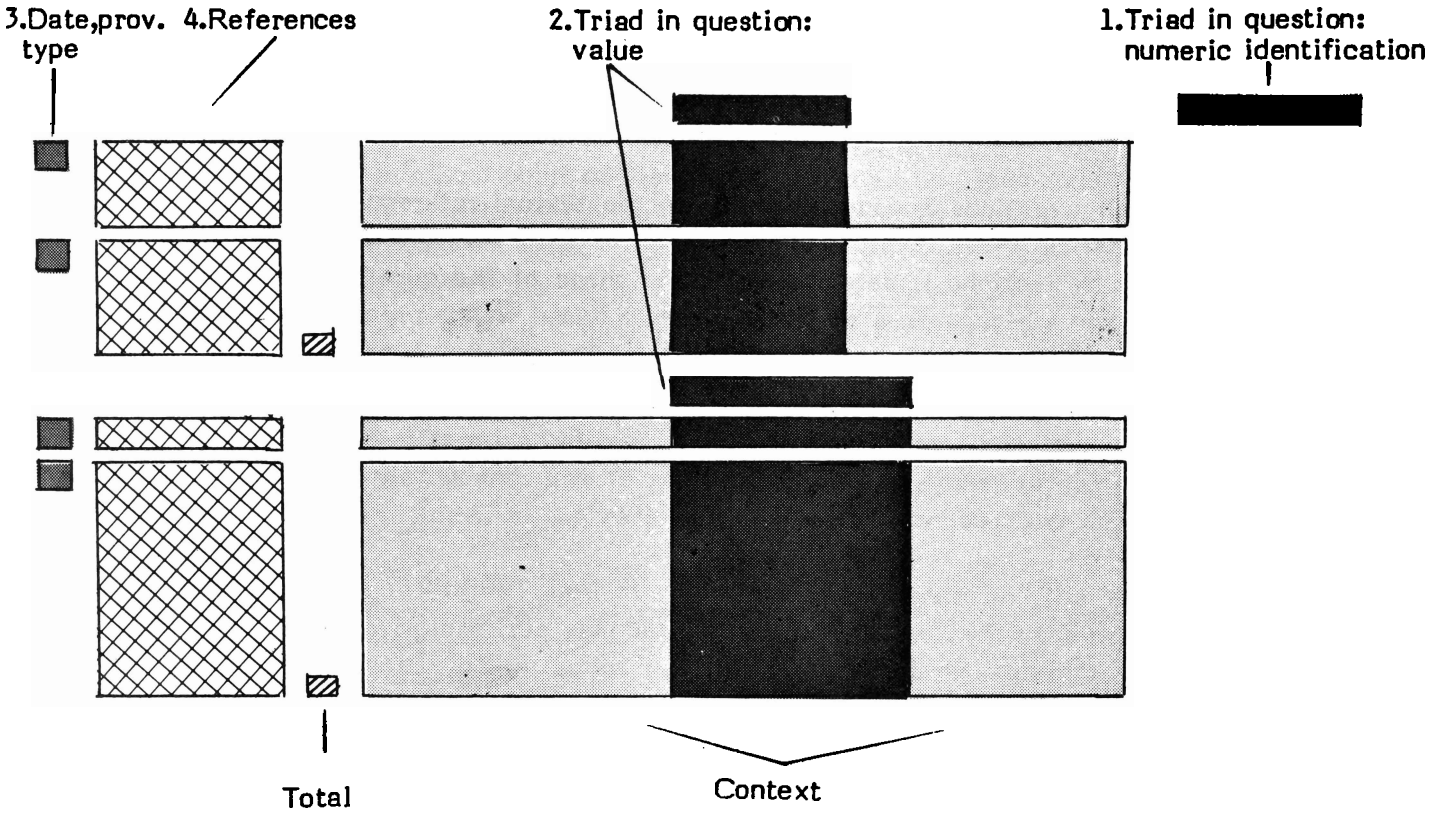


Chart 1. System Configuration
(Note. Numbers refer to sorting hierarchy.)

D. SIGN CONCORDANCE



E. WORD CONCORDANCE

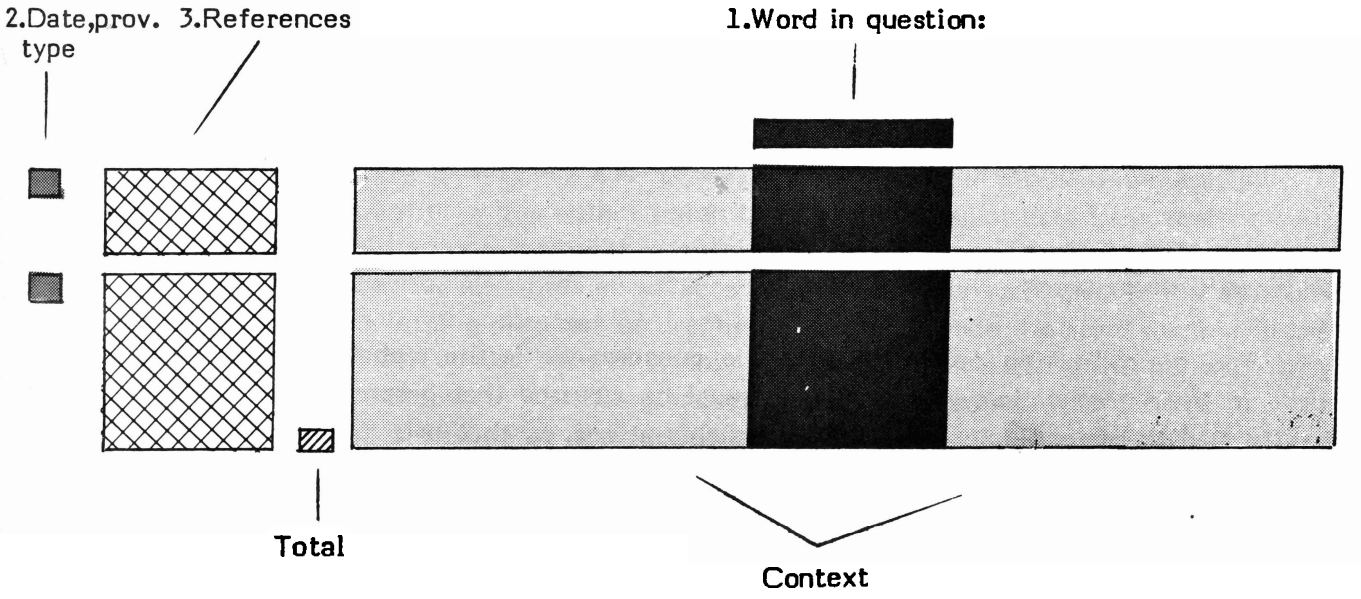


Chart 1. System Configuration (Continued)
(Note. Numbers refer to sorting hierarchy.)

center. (This is the traditional format of computer concordances, in the style of the so-called KWIC Index.) Line boundaries are indicated in the transliteration by a slanted line (/).

Sorting within sections is done according to the textual references, arranged in alphabetical and numerical order.

The sorting by sections is determined by the shape of the cuneiform signs in each given triad. All triads are given in a sequence which is based on the progressive graphic arrangement of cuneiform signs that is standard in Assyriology. A complete theoretical sequence of triads would read as follows:

1	1	1
1	1	2
1	1	3
1	1	4
.....		
1	1	999
1	2	1
1	2	2
1	2	3
1	2	4
.....		
1	2	999
1	3	1
1	3	2

1	3	3
1	3	4
.....		
1	3	999
.....		
1	999	999
2	1	1
2	1	2
2	1	3
2	1	4
.....		
2	1	999
2	2	1
2	2	2

2	2	3
2	2	4
.....		
2	2	999
2	3	1
2	3	2
2	3	3
2	3	4
.....		
2	3	999
.....		
2	999	999
.....		
999	999	999

It should be noted that the choice of a triad, i.e. a sequence of three, rather than two or four, signs is not accidental. One of the results of graphemic analysis about which I have reported elsewhere (Buccellati, 1979 and 1983) is that a triad seems to represent a critical mass, in the sense that it is sufficient to reduce to a minimum the phenomenon of polyvalence within a homogeneous corpus (see also below, section 2.2: Index of Polyvalence). A two sign sequence would not be sufficiently distinctive, i.e. there would be too many cases of polyvalence left. A four sign sequence, on the other hand, would reduce polyvalence almost totally but it would increase by 25% the size of the sign concordance.

1.4 Word Concordance

The main sorting criterion for this concordance is the alphabetical sequence of words in their graphemic configuration. It must be stressed that graphemic analysis is not directly tied to morphological and lexical considerations, so that this is not fully and truly a word concordance, precisely because it lacks those two fundamental dimensions of word analysis. The reason why a word concordance may nevertheless be legitimately made a part of graphemic categorization is that the concept of "word" refers here to the boundaries within which signs are sorted. In this sense then our "word" concordance is properly and specifically a "word oriented" graphemic concordance, and not a concordance of words as

phonemic and morpho-lexical items.

An example will clarify the difference. An Akkadian word such as a:lum might be rendered in cuneiform in different ways, e. g. a-lu-um, a-lum, URU, URU.KI. From a phonemic and morpho-lexical point of view these would all be listed together under a single entry /a:lum/. From a graphemic point of view, on the other hand, they are listed separately as four different entries. A true word concordance, in a phonemic and morpho-lexical sense, will appear in the series Morpho-Lexical Categorization of Akkadian Texts. The Word Concordance is significantly shorter in size than the Sign Concordance. In the Sign Concordance each word appears as many times as there are signs in the word, plus two. In the Word Concordance, on the other hand, any word, whatever its length, appears only once. Thus the difference between the size of the Sign Concordance and the size of the Word Concordance depends on the ratio between the number of signs and the number of words as inventory items. In the case of the Middle Assyrian Laws such a ratio is 8.66 to 1.

The concept of "word" used in the Word Concordance and the Word Index is the traditional one in Assyriology (see also below, 2.2: Definitions), with the proviso that determinatives and elements of personal names are treated as separate words as well as components of a larger word. On the other hand, phonological indicators ("phonetic complements"), enclitics and suffixes are not treated separately.

The idea of a word concordance was first suggested by Guy Bunnens during his stay in Los Angeles as a Research Associate on our computer project. I wish to acknowledge here his contribution.

1.5 Sign Index

The first and second sorting criteria of the Sign Index are identical to the first and second sorting criteria of the Sign Concordance -- the graphic sequence of the signs as reflected by the sequential identification number of the standard lists, and the alphabetical sequence of alternate readings for the same triad.

The third sorting criterion in the Sign Index, not found in the Sign Concordance, is represented by the word or words within which the signs are found, listed alphabetically. If the triad in question is contained within one word, only that word is cited. If the triad straddles word boundaries, all words so affected are cited. This sorting criterion adds one interesting dimension for the purposes of graphemic analysis: since the word context is the one which most closely conditions the various graphemic alternations, the corresponding sorting is especially pertinent to an understanding of graphemic rules.

This is further enhanced by the identification of variant readings or values of the same triad by means of a "bullet" (i. e. a solid circle) on the left of the word(s) in which the triad appears. Variant readings are found regularly in different words, and the nature of the presentation given in the Sign Index is especially instructive in this respect.

Where applicable, the last sorting criterion subdivides the occurrences of each word

by date, provenience or type of text. (This sort is not applicable to the Middle Assyrian Laws.) References are omitted from the Sign Index, but the total number of occurrences is given for each major sort.

1.6 Frequency Tabulations

The Frequency Tabulations provide some basic computations which can be of considerable use for a more sophisticated type of graphemic analysis (see for now the preliminary applications given below in section 2.2, as well as the pertinent literature quoted in the bibliography). The frequencies are given per sign and value. The sign frequencies are sorted first by sign (in the standard sign list sequence) and then by number of occurrences in descending order of frequency. For each sign, one will find the total number of occurrences and the percentile value within the corpus. Within each sign, all values attested in a given corpus are listed respectively in either an alphabetical or a frequency sequence; for each value one will find the total number of occurrences, the percentile value within the sign in question, and the percentile value within the total corpus. At the end of the sign frequencies there is the total number of signs as inventory items and of occurrences for the corpus.

The value frequencies are sorted first by value in alphabetical sequence, and then by number of occurrences in descending order of frequency. For each value, one will find the total number of occurrences, the percentile within the corpus and the percentile within the sign in question. The sign in question is identified by its standard numerical value; for comparative purposes, the total number of occurrences of the sign in question is also added. At the end of the value frequencies, there is a total of values as inventory items and of value occurrences within the corpus.

Values used in these tabulations are maximal values only (see below under 2.2: Definitions). Alternate sorts by minimal values will be found in graphic form in Charts 2 and 4. Since I consider the concept of minimal value to overlap with that of grapheme (see below), the tabulation found in these charts may be considered specifically a full listing of graphemes.

1.7 Word Index with References

The sorting criteria of the Word Index are identical to those of the Word Concordance: the words are listed alphabetically according to their graphemic configuration, and for each word the appropriate references are listed in alphanumeric sequence. At the end of each entry, a total is given for all the words in that entry. Thus the Word Index is an abbreviated form, or a summary report, of the Word Concordance, identical in everything except the presence of a context.

There are two main advantages to the Word Index. (1) Since word context is an

important factor that conditions the graphemic value of the signs, the sign environment given in the Word Index is particularly useful for the purposes of graphemic analysis. (2) In conjunction with the Sign Index, the Word Index allows for the most economical and at the same time the most differentiated presentation of the data: full documentary evidence for all sign occurrences, within a minimum of space and with the distinct advantage of dual sorting criteria. For this reason, publication in book form will normally be limited to the Sign Index and the Word Index.

1.8 Diacritics and Special Symbols

In the transliteration, semantic and phonological indicators (i. e. determinatives and phonetic complements) have been transliterated on the line, rather than as superscripts. Semantic indicators are transliterated in capital letters like logograms, and phonological indicators are given in lower case.

The following conventions have been used in the transliteration:

h = ḫ	!	amended
ṣ = ṣ̌	?	questionable
š = š̌	[	broken context
ṭ = ṭ̌	<	sign(s) presumed omitted by scribe
	«	sign(s) erroneously entered by the scribe

(When present next to the reference in the Word Index, these symbols refer to a given portion of the context.)

At the end of the sign sequence, the following numbers have been added: 997 for a single broken sign and 998 for a break that includes an undetermined number of signs.

All rank indices for cuneiform signs are superscripted rather than subscripted, e. g. U^{2}. Superscript numbers (^{2} and ^{3}) are used instead of the acute and grave accents. Determinatives are written on the line, not superscripted, and are fully spelled out in capital letters, e. g. DINGIR, not superscript d. The personal marker ("Personenkeil") is represented by the numeral 1; e. g. 1-DINGIR-a-šur-u^{2}-ma-i. The determinative for feminine names has been transliterated as SAL. Hyphens rather than periods are used between signs in logograms.

2. The Middle Assyrian Laws

2.1 The Corpus

The programs have been applied to the corpus as published in DSC 2, corrected according to the errata sheet found in that volume and also according to the few changes listed below. As stated above, C. Saporetti has retained full control and responsibility with regard to the philological preparation of the data: the initial data entry was carried out in Pisa under his direct supervision, and while the elaboration printed in this volume was produced in Los Angeles under my supervision and according to my own format specifications, he was able to review them for approval as to content before publication.

The list of texts utilized follows; the texts whose standard abbreviation is too long for the format adopted in the listings have been abbreviated as indicated:

AFO12,T.3,2	= AFO12:2	KAV1 VII
AFO12,T.5,1	= AFO12:1	KAV1 VIII
AFO12,T.5,2,I REC	= AFO12:6	KAV2 II
AFO12,T.5,2,I VER	= AFO12:8	KAV2 III
AFO12,T.5,2,II REC	= AFO12:7	KAV2 IV
AFO12,T.5,2,II VER	= AFO12:9	KAV2 V
AFO12,T.6,1 REC	= AFO12:3	KAV2 VII
AFO12,T.6,1 VER	= AFO12:4	KAV3
AFO12,T.6,2	= AFO12:5	KAV4 REC
IRAQ35,T12,4		KAV4 VER
KAV1 I		KAV5 I
KAV1 II		KAV5 II
KAV1 III		KAV6 REC + KAV143 = KAV6R+143
KAV1 IV		KAV6 VER + KAV143 = KAV6V+143
KAV1 V		KAV143 VER + KAV6 = KAV143V+6
KAV1 VI		KAV193

The following corrections, not included in DSC 1, were entered in the data base before producing the listings published here:

KAV1	I	16	taq3	KAV2	II	23	lu ši-la-ta
	II	66	-ma		III	22	ma-la!-e
	III	33,37	taq3-ṭì-bi	KAV5	II	8	⌜i⌝-[da-an]
		72	ša	AfO 12	T.6,1 REC	7	[ša M]U
	V	78	-as2-šir3		T.6,2	8	⌜la⌝-a
		95	šir3				
	VI	14	-qi2				

2.2 Preliminary Graphemic Analysis

A full graphemic elaboration of the data will be presented at a later date in the projected series Graphemic Analysis of Cuneiform Texts: such elaboration will consist of analytical modules which will perform standard operations on the corpora as a whole and on specific subsets of the same corpora; it will also contain an explicit statement of method, an extensive comparison of the various corpora, and a full discussion of the implications and the linguistic significance of the data. In these early volumes of Graphemic Categorization I will anticipate some of these "analytical modules" by offering the results of simple statistical computations on the data contained in each volume with preliminary remarks on their significance. It must be stressed that the full significance of these figures can only be derived from comparisons with other corpora as they will appear in GC.

Definitions

Definitions are repeated in each of the volumes of the series GC since several terms and concepts are not standard in the field.

Corpus	- repertory of graphs or sign occurrences (in textual data)
Graph	- minimal graphic unit, i.e. each individual sign occurring uniquely in a text (e.g. each occurrence of Sign N. 13, including palaeographic variants)
Grapheme	- minimal unit in the correlation between sets of identical graphs and a single phonemic realization; see also "value"
Inventory	- repertory of signs and graphemes
Logogram	- a grapheme with lexically conditioned correlation to phonemic inventory; either a single logogram or a cluster of logograms may stand for a word
Phonogram	- a grapheme with direct correlation to phonemic inventory (i. e. lexically unconditioned)
Polyvalence	- the state of a sign whereby it can have many values (traditionally called "polyphony"). See also below, under Index of Polyvalence.
Reading	- same as "value"
Repetition	- the amount of inventory words in relationship to the total word count. The index of word repetition expresses the ratio between the two: the higher the index, the more repetitive the word usage, the lower the index, the more dispersion there is in the corpus (an index of 1 would mean that every word in the text is a unique word).

Sign - any set of identical graphs subsuming either one or more graphemes (or value), e.g. Sign N. 13 as it subsumes the values an, DINGIR, etc.

Sign occurrence - same as "graph"

Triad - a string of three graphemes, occurring sequentially in a text

Value - phonemic realization (or "reading") of a given sign; it corresponds either to a grapheme proper (minimal value) or to a grapheme with a graphemic rule embedded in its notation (maximal value), as follows:

minimal - a minimal range of values, corresponding to a grapheme proper; it combines any and all of the following values: phonograms differentiated only by the voice or emphatic dimension of one consonant (e. g. ig, ik, iq); phonograms differentiated only by the vocalic contrast between i and e (e. g. li and le); phonograms without vocalic differentiation (e.g. ah, eh, ih and uh); determinatives and logograms (e. g. DINGIR and DINGIR); thus for example ig, ik, iq, eg, ek, eq are all taken together as a single minimal value or grapheme proper

maximal - a maximal range of values, corresponding to a grapheme proper with one or more graphemic rules made explicit in its notation; it differentiates each and every element of the above mentioned pairs as independent values

Word - a minimal unit defined on the basis of various criteria, including the following: phonemic (presumed stress unit), morphemic (limits on co-occurrence of morphemes, etc.), morphophonemic (alternations resulting from either internal or external inflection), lexical (semantic value). Graphemic factors play only a minimal role in defining a word - e. g. word boundaries are coincident with line boundaries.

The relationship between graph, sign and grapheme may be summed up as follows:

graphs	signs	graphemes proper (=minimal values)	graphemes+rules (=maximal values)
	N.13	DINGIR, AN, an,...	il,ilum ... Anum,šamē,... an
	N.97	aG	ag,ak,aq

A special notation is used here to mark minimal values (graphemes proper), at least in the case of phonograms: a capital letter stands for a set of alternating realizations. Specifically: E stands for either e or i (e. g. lE for either le or li); A stands

for any vowel (e. g. Ab for ah, eh, ih, uh); a capital voiced consonant stands for any consonant in the corresponding phonemic class (e. g. Bu for either bu or pu). Thus EG stands for eg, ek, eq, ig, ik and iq. This notation is applied in most but no all cases where it is pertinent: it is omitted, for instance, in cases where alternate values are rare and the pertinent notation would require using an unusual transliteration (e. g. mE3 for mi).

Basic Numeric Data

Factors. Following are various absolute numeric values (as different from percentile values), expressing the totals for each of the named classes as they pertain to the Middle Assyrian Laws.

Graphs (corpus items, sign occurrences)		11460
Signs	- all signs	194
	- polyvalent signs, maximal	88
	- minimal	55
Graphemes	- graphemes+rules (maximal)	341
	- graphemes proper (minimal)	212
	- logograms and numbers (maximal)	71
	- (minimal)	not available
	- phonograms (maximal)	260
	- (minimal)	141
Triads	- all	6677
	- monovalent	6585
	- polyvalent	92
Words	- total count (except breaks)	4513
	- breaks and initial breaks	509
	- inventory items	1324

Ratios are the relationships between two values, and they are expressed here in the form of either averages or indexes. **Averages** are given here as simple means; I have not computed median averages, which, it should be noted, would provide a better indicator in several cases. (For instance, with a simple mean average for word length one cannot conclude that the higher the average the longer the words - because one or two long words occurring very frequently would yield the same average as several medium length words occurring less frequently.) Two different word averages are given. The first represents the average length of a word as a text occurrence. The second represents the average length of a word as an inventory item - something which has no proper linguistic significance, but is useful for practical purposes, see above, section 1.4). - By **index** I mean a ratio in which 1 stands for the minimum occurrence (that is, the non-occurrence) of the named category, while proportionately higher numbers indicate proportionately higher occurrences of the named category. Thus, 1.83 is the highest index of polyvalence in the Middle Assyrian Laws (for signs with maximal values), while the index of 1 for logograms means total monovalence (but see below).

Average word length	- count: sign count/word count	2.54
	- inventory: sign count/word inventory	8.65
Index of logography	- maximal values/maximal phonographic values	1.31
polyvalence	- signs: maximal values/maximal monovalent signs	1.83
	- signs: minimal values/minimal monovalent signs	1.20
	- phonograms: min. phon./min. monovalent phon.	1.64
	- logograms: min. log./min. monvalent log.	1.00
	- triads: all triads/monovalent triads	1.01
word repetition	- word count/word inventory	3.41

Frequency Distribution

Frequency ranges are scattered along a continuum line, without significant interruptions. The nature of this continuum is illustrated by the curve in Chart 2. Graphemes proper (i. e. minimal values) are used for this chart, and they are subdivided into the three categories of phonograms, logograms and numbers. This yields a new sort in addition to those already provided in the Frequency Tabulations, where only maximal values are used. (Note that the charts too are computer generated, utilizing the plotting program described in CARNES 1/1.)

Within this continuum, I have arbitrarily established three frequency classes with the following parameters:

1. a frequent unit exhibits a frequency of 1% or more of a given corpus;
2. a common unit exhibits a frequency between .1% and .99% of the same corpus;
3. a rare unit exhibits a frequency of less than .1% of the same corpus.

On this basis, the frequency distribution for the Middle Assyrian Laws yields the classes tabulated in Chart 2. Because of the generally low index of polyvalence, there is a close correspondence between signs and graphemes, as shown by the correlation of the two sets of bar histograms. The data may be briefly interpreted as follows (I limit the following observations to maximal values only).

Signs. The absolute range of occurrences for the Middle Assyrian Laws is between 118 and 723 times for individual frequent signs, between 12 and 108 for common signs, and between 1 and 11 for rare signs.

There are 26 frequent signs, corresponding to a total of 7006 graphs; in percentile terms, this means that 13% of the sign inventory accounts for 61% of the total sign corpus.

There are 91 common signs, corresponding to a total of 4164 graphs; in percentile terms, this means that 46% of the sign inventory accounts for only 36% of the total sign corpus.

There are 77 rare signs, corresponding to a total of 314 graphs; in percentile terms, this means that as much as 36% of the sign inventory accounts for a mere 2.73% of the total sign corpus.

Graphemes. The absolute range of occurrences for the Middle Assyrian Laws is between 123 and 668 times for individual frequent graphemes, between 12 and 103 for common graphemes, and between 1 and 11 for rare graphemes.

There are 22 frequent graphemes, corresponding to a total of 6237 graphs; in percentile terms, this means that 6% of the graphemic inventory accounts for 54% of the total text (or graph) corpus.

There are 119 common graphemes, corresponding to a total of 4469 graphs; in percentile terms, this means that 35% of the graphemic inventory accounts for 39% of the total text (or graph) corpus.

There are 200 rare graphemes, corresponding to a total of 754 graphs; in percentile terms, this means that as much as 59% of the graphemic inventory accounts for a mere 7% of the total text (or graph) corpus.

The significance of these statistics may best be appreciated if they are used as standards of measurements in comparing other corpora (see Buccellati 1979 and 1983 for some preliminary tabulation of such comparisons). In general terms, there appears to be a uniform distributional pattern applicable to all corpora, in such a way that even though the actual graphemes used vary considerably across corpora, the general economy remains the same for each corpus-specific system in terms of frequency ratios: a limited number of graphemes accounts for more than half of all sign occurrences, while conversely about half of the graphemes account for an extremely small portion of all sign occurrences.

In other words, there is an inverse ratio between frequency of inventory and frequency of corpus, and there is a marked symmetry in the relative distribution of the two elements, especially so in the case of graphemes. This is a function of the highly symmetrical overall distribution of graphemes as shown by the curve in Chart 2. A priori, one could have expected a stepped rather than a curved array: for instance, there could be two uniform classes, one of very rare and one of common signs (such a situation would conceivably be the result of a deliberate scribal effort at standardization).

Index of Polyvalence

The polyvalence of cuneiform signs, while pervasive in terms of the Akkadian graphemic inventory as a whole, is in fact greatly constrained within individual corpora. I have computed here several distinct indices of polyvalence. The data are subdivided into classes as shown in Chart 3. The various classes are shown there in decreasing order of polyvalence.

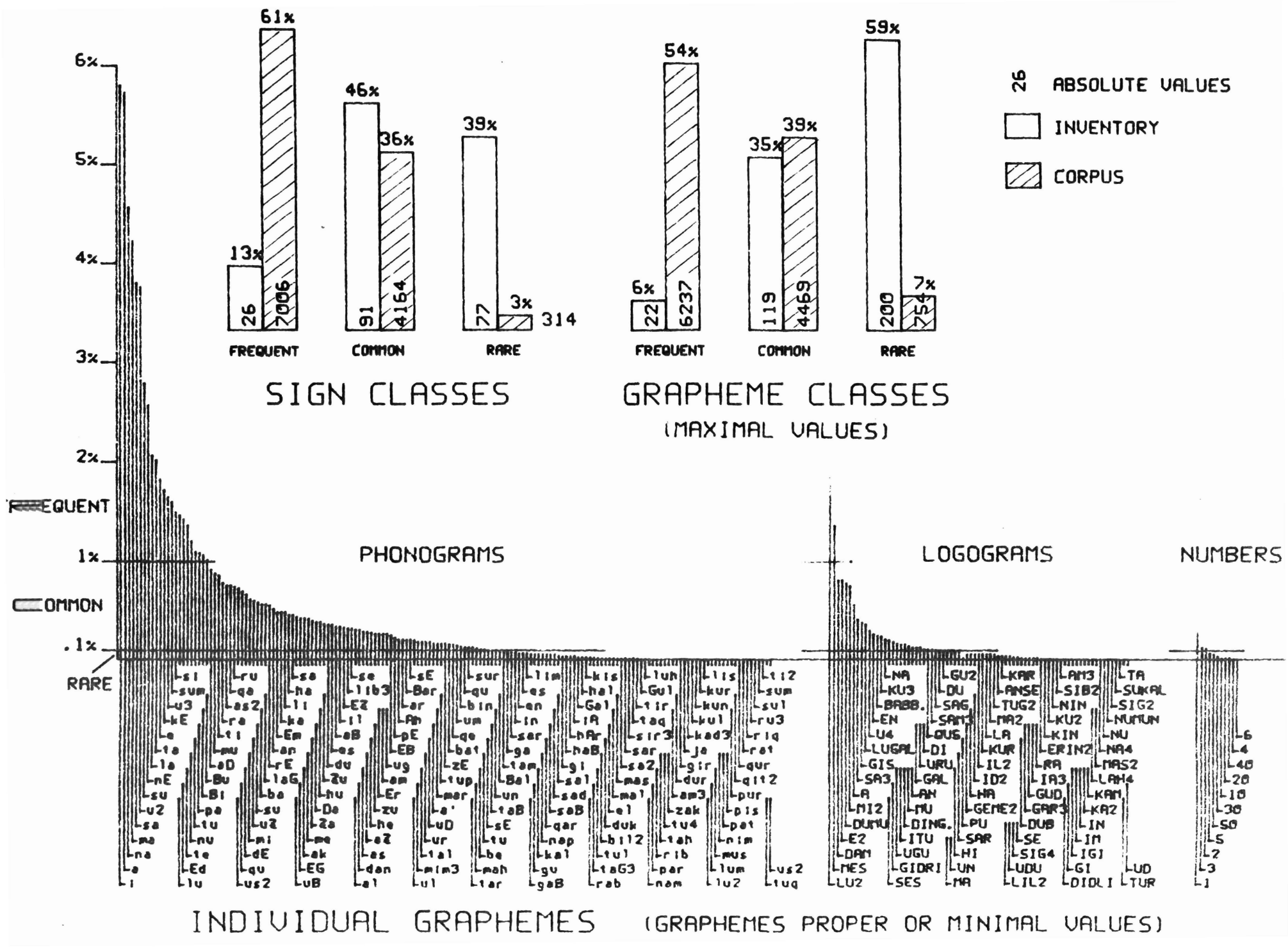


Chart 2. Frequency Distribution

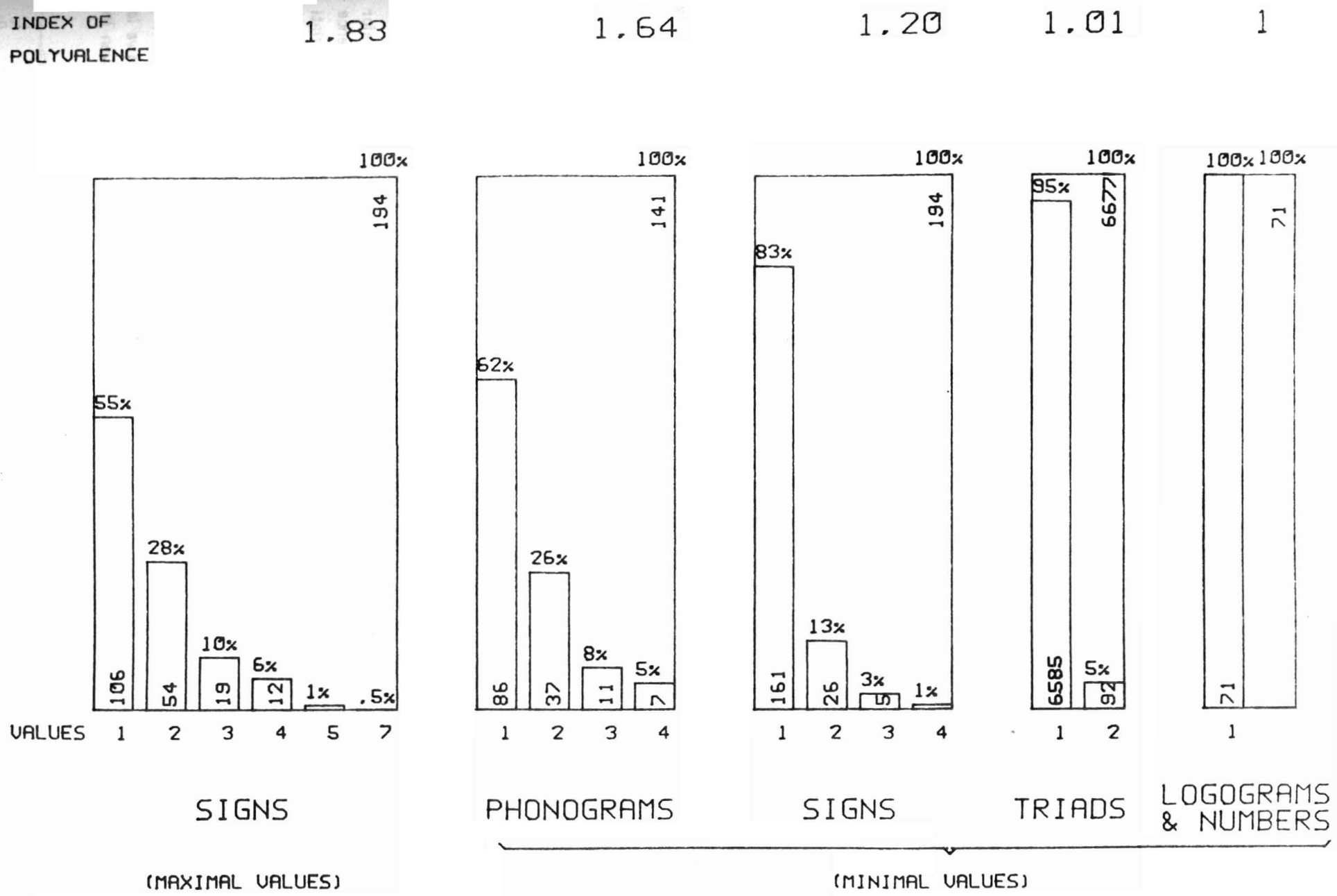


Chart 3. Polyvalence Ratios

A polyvalence ratio expresses the relationship between the total number of inventory items and the total number of values. (An index is the same ratio, except that one of the members of the pair is equated to 1.)

One computation is based on signs. There is a total of 106 monovalent signs (with maximal values) and a total of 157 monovalent signs (with minimal values) to a total of 194 inventory items: the respective indices are 1.83 and 1.20. (Again, an index of 1 would mean that all signs are monovalent.)

Another computation is based on phonograms only. There is a total of 86 monovalent phonograms to a total of 141 phonograms (minimal values): the index is 1.64.

The computation based on logograms shows the lowest possible index of polyvalence, i. e. an index of 1, which means total monovalence. Note that this is because the computation is based on minimal values. With maximal values (e. g. a:l, a:lum, a:lam, etc. for URU) logograms would in fact exhibit the highest polyvalence index.

The computation based on triads shows a more signifcantly low index of polyvalence. There is a total of 6,585 monovalent triads to a total of 6,677 triads (whether minimal or maximal values): the index is 1.01. It is immediately apparent that when triads are considered, the system reaches a state of almost complete monovalence (since here too an index of 1 would mean that all triads are monovalent).

We may call the first index an index of sign polyvalence, and the second an index of triad or contextual polyvalence: the higher the index, the more differentiated (and possibly the more heterogeneous) are the texts.

Graphemic Profile

The last chart (N. 4) provides a complete graphemic profile of the corpus. The graphemes are given as graphemes proper, i. e. with minimal values. In this sense, this represents the graphemic inventory proper of the Middle Assyrian Laws. The graphemes are subdivided into phonograms, logograms and numbers, and they are listed alphabetically within each category. The notation of graphemes is explained above in this section.

The full significance of such a profile can only be gauged from an overlay with profiles of other corpora. To this end, not only signs attested in one corpus, but all major signs attested in groups of corpora should be considered, with indication of non-occurrence in specific corpora as appropriate. Such a procedure (already exemplified in Buccellati 1983) will be followed in later volumes of this series, and especially in the volumes of Graphemic Analysis.

2.3 List of Signs in the Corpus

A complete list of all signs occurring in the corpus of Middle Assyrian Laws is provided in Appendix 1, both in numerical and in alphabetical sequence. This is given in

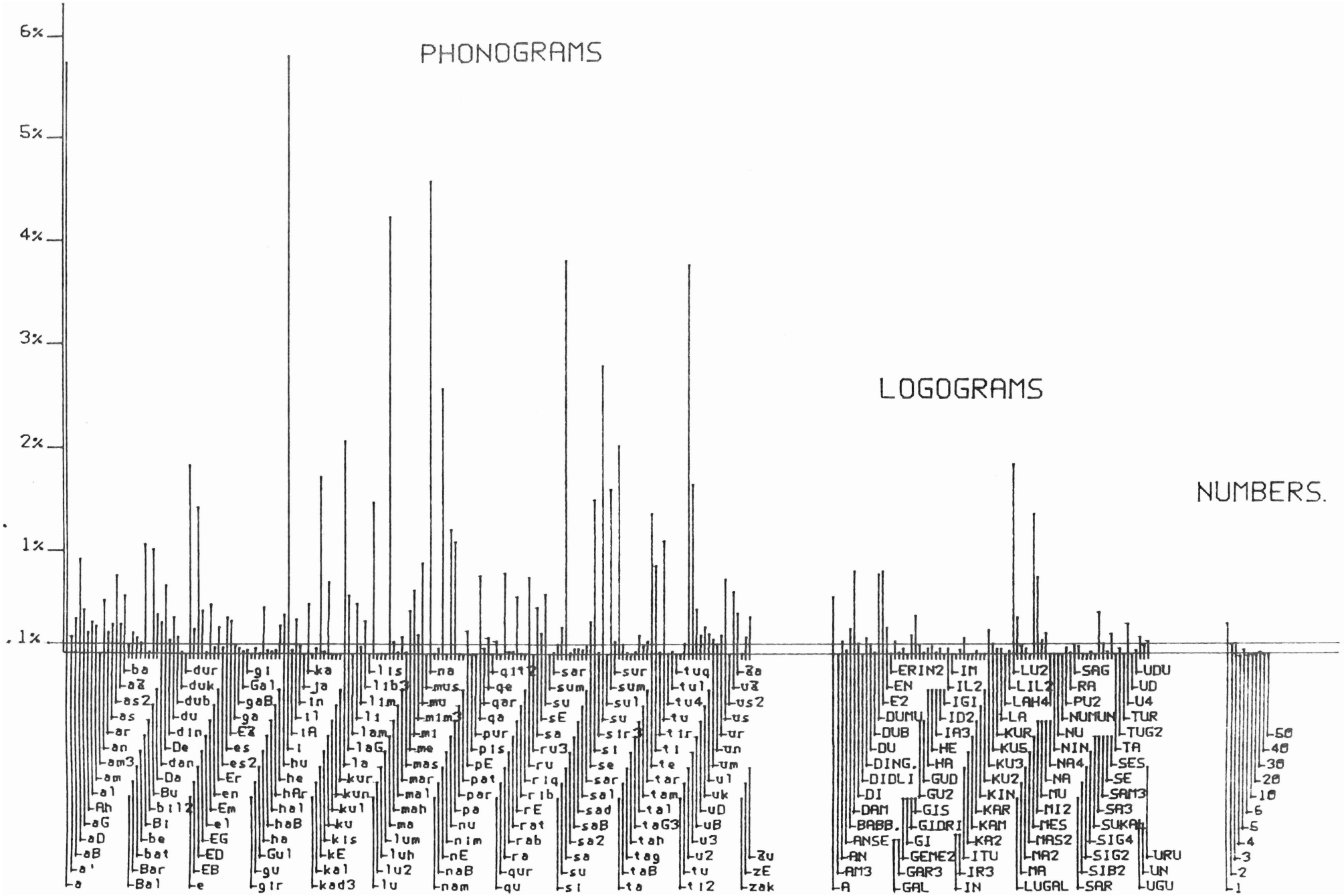


Chart 4. Graphemic Profile

compact format on the very last page of the volume so that one may easily refer to it in order to identify the signs from their numeric value. A single standard value has been assigned to each sign: these values do not have any graphemic significance, but stand simply as labels for the individual signs.

3. References.

Buccellati, G. 1979 "Comparative Graphemic Analysis of Old Babylonian and Western Akkadian," Schaeffer Festschrift, UF 11 (1979) 89-100.

-- 1983 "Studies in Ebla Graphemics, 1," Studi Eblaiti, 5 (in press)

Gaebelein, P. W. 1977 Graphemic Analysis of Old Babylonian Letters from Mari, Ph. D. Disseration, University of California, Los Angeles

Kobayashi, Y. 1975 Graphemic Analysis of Old Babylonian Lettrs from Southern Babylonia, Ph. D. Disseration, University of California, Los Angeles

GC2

CYBERNETICA MESOPOTAMICA
GRAPHEMIC CATEGORIZATION
VOLUME TWO

THE MIDDLE ASSYRIAN LAWS

C. SAPORETTI

CYBERNETICA MESOPOTAMICA

Electronic Data Processing of Mesopotamian Materials
Prospectus of Series in the Collection

MECHANISMS

EMP	Encoding Manuals and Programs
TM	Theory and Methods

DATA

DSC	Data Sets: Cuneiform Texts
DSA	Data Sets: Artifactual Materials

RESULTS

Categorization

GC	Graphemic Categorization of Cuneiform Texts
HC	Historical Categorization of Cuneiform Texts
LC	Lexical Categorization
MLC	Morpho-Lexical Categorization
AAC	Artifactual Attributes Categorization
SR	Stratigraphic Record

Analysis

GAC	Graphemic Analysis of Cuneiform Texts
HAC	Historical Analysis of Cuneiform Texts
MLA	Morpholexical Analysis of Akkadian Texts
SCA	Syntactical and Compositional Analysis of Akkadian Texts
FIAM	Formal and Iconographic Analysis of Mesopotamian Materials
SDA	Stratigraphic and Depositional Analysis

Published under the auspices of
IIMAS
The International Institute for Mesopotamian Area Studies

CM

CYBERNETICA MESOPOTAMICA

ELECTRONIC DATA PROCESSING OF MESOPOTAMIAN MATERIALS

GIORGIO BUCCELLATI, EDITOR
JUDITH R. PAUL, ASSISTANT EDITOR

ANALISI ELETTRONICA DEL CUNEIFORME:
COMPONENTE ITALIANA DI CYBERNETICA MESOPOTAMICA

A CURA DI CLAUDIO SAPORETTI

PUBLISHED UNDER THE AUSPICES OF
IIMAS
THE INTERNATIONAL INSTITUTE FOR MESOPOTAMIAN AREA STUDIES

GC

2

GRAPHEMIC CATEGORIZATION

VOLUME 2:
THE MIDDLE ASSYRIAN LAWS

(SIGN INDEX, FREQUENCY TABULATIONS AND WORD INDEX)

Corpus established by **CLAUDIO SAPORETTI**
System design and introduction by **GIORGIO BUCCELLATI**

Programming: **JOHN L. SETTLES, DAVID A. HOLZGANG, JUDITH R. PAUL**
Data entry: **CLAUDIO SAPORETTI, MANUELA SASSI, ANGELO GHIROLDI**
Data coordination: **MATTHEW L. JAFFE**

UNDENA PUBLICATIONS
MALIBU 1984

GC 2 is the fourth volume published within the framework of *ALTAN - Analisi Linguisitica dei Testi Assiri e Nuziani.*

ISBN: 0-89003-120-7

TABLE OF CONTENTS

ACKNOWLEDGMENTS

Publication of this volume was made possible through a special grant from

Ambassador International Cultural Foundation, Pasadena

Research was made possible through grants from

Consiglio Nazionale delle Ricerche, Roma
The National Endowment for the Humanities, Washington
The Research Committee, The University of California, Los Angeles
Ambassador International Cultural Foundation, Pasadena

Research, programming and data processing was carried out at

Laboratorio per le Tecnologie Applicate ai Beni Culturali,
Area della Ricerca del CNR, Roma
Laboratorio di Linguistica Computazionale, Pisa
Office of Academic Computing, The University of California, Los Angeles
IIMAS - The International Institute for Mesopotamian Area Studies, Los Angeles