A NOTE TO THE READER

The attached document consists of relevant pages from the original style manual for the Comprehensive Aramaic Lexicon Project as prepared in 1987, before the onset of the project itself, before the development of the actual software to process the texts, and, most importantly, before the internet, HTML, and the currently ubiquitous text software allowing for the easy entry and display of Semitic alphabetic scripts. Its presence here is to aid online users in deciphering coded CAL texts, should they choose to view them in Roman code, as well as to better understand some of the coding conventions they may encounter even when viewing texts in their native scripts.

The following are the major changes that should be kept in mind both for those using the CAL and for scholars submitting materials to the CAL:

A) Texts may now be submitted in any standard font recognizable to current browsers, in particular the unicode Hebrew and Syriac fonts that had not yet become standard when the Project was first envisioned. Idiosyncratic fonts such as those designed for CPA, Samaritan, and Mandaic may be used, but texts will not be displayed in such fonts on the CAL or in CAL publications. It is our view that the use of such fonts in current publications detracts from their usefulness to the average scholar, let alone the casual user, in a very major way.

B) Diacritics included in such modern fonts and understood as such by modern browsers may be included in text submissions. Thus, for example, a *taw* with the "Syriac Feminine Dot" may be indicated by in Roman **tF** [an addition to our coding, by the way] or by Meltho ܬ̈.

C) In view of the some of peculiar behaviors of HTML in different browsers when dealing with right-to-left texts, some policies for displaying fragmentary material have had to be changed. In particular, in displaying texts in "Hebrew" characters, variants are given in between quotation marks [thus e.g.: "#2#אשתנון"אשתני] corresponding to a coded :)$tny/)$tnwn#2#/. Brackets that would normally appear at the end of a line of text have been eliminated in Hebrew displays.

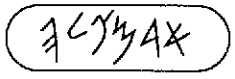
THE COMPREHENSIVE ARAMAIC LEXICON

TEXT ENTRY AND FORMAT MANUAL

by

Stephen A. Kaufman

A Publication of
THE COMPREHENSIVE ARAMAIC LEXICON



Baltimore, 1987

Table of Contents

PREFACE

Practicality necessitates that we publish this *Manual* as early as possible in the life of the CAL project. The provisional nature of the following materials cannot be overemphasized. It is, quite simply, impossible to anticipate at the outset the scope of the problems and variations that will be encountered in the treatment of such an enormous and varied corpus of textual material. Accordingly, the author and other editors of the CAL will be more than pleased to accept comments and suggestions from any scholars and students who have attempted to put the guidelines contained herein into practice and/or those who have made already made extensive use of the text files (i.e. "output" format) of the CAL. Periodic updates may well follow this publication.

I am most grateful to those who have reviewed earlier drafts of this document, Prof. Bezalel Porten meriting particular mention in this regard. Their trenchant criticisms prevented many significant omissions and oversights. I would also like to thank the staff of the Institute for Advanced Studies of the Hebrew University—under whose pleasant auspices the first draft of this work was written in November and December, 1985—and Prof. M. Goshen-Gottstein, who invited me to participate in the Seminar Group for Targumic and Cognate Studies at the Institute.

All errors of judgment and infelicities of style are, of course, totally the responsibility of the author.

INTRODUCTION I

The COMPREHENSIVE ARAMAIC LEXICON (CAL) is a long term project involving an international team of scholars. Physical headquarters for the project are at the Department of Near Eastern Studies of The Johns Hopkins University, Baltimore, Maryland, USA. Its immediate goals are (1) to assemble a machine readable database (i.e. a collection of data encoded on computer readable magnetic media and organized according to certain standard principles) consisting of the entire (or as nearly entire as practicable) corpus of Aramaic texts, in all of the dialects thereof, exclusive of the vernaculars; and (2) to produce from that database a complete, citation-based dictionary of the Aramaic language.

Our goal being an accurate and complete lexicon, it is essential that the machine-readable database be compiled from text editions of high critical merit. Where such an edition is not available, and wherever appropriate and practical, the CAL project will suggest, encourage, and support the preparation of new text editions, in computer-readable (e.g. diskette and, eventually, CD-ROM), printed, or microfiche format or a combination thereof. In almost every important respect, the vast size and heterogeneity of the corpus present significant technical problems for the project, designed as it is to encompass the entirety of Aramaic. Probably few if any other lexical projects have ever had to contend with such a variety of scripts, media, dialects, and genres. In order for the CAL computer programs to process this diverse data accurately and rapidly the machine readable text editions obviously must be prepared according to a standard set of guidelines—a standard "format". It is the purpose of this document to provide the details of that format.

In preparing these guidelines the author has relied heavily upon the preliminary text entry procedures adopted by Michael Sokoloff for his new dictionary of Palestinian Jewish Aramaic and on the work of David Mackenzie, *A Manual of Manuscript Transcription for the Dictionary of the Old Spanish Language* (Madison, 1984).

PRELIMINARY CONSIDERATIONS

Solely from the perspective of the primary, lexicographic function of the CAL, any text format that represents all of the lexicographically relevant information in its source text would be adequate. The editors of the CAL are well aware, however, that the

machine-readable editions prepared for the explicit use of the CAL will be used by scholars for many additional functions. Chief among the extra-lexicographic functions from the point of view of the work of the CAL is that of "editions": As machine-readable text editions are prepared, it is the intent of the CAL project to publish hard- and/or microcopy editions of the more noteworthy of these texts and/or concordances thereof. Thus, at least to a limited extent, the format must be flexible enough to reflect the needs both of those who would edit or publish a text known only from a single exemplar and those engaged in the preparation of a critical edition of a text known from numerous manuscripts of varying quality.

It does not currently seem advisable, however, to attempt to incorporate into the standard format sigla for every conceivable feature that one might encounter in an Aramaic manuscript, especially as regards the scribal idiosyncrasies found in Syriac codices. Thus, and this is perhaps the most important example, we make no provision for the incorporation of the accent marks of biblical, targumic, and Syriac Bible manuscripts into the machine-readable texts. We have not provided equivalents for the innumerable special symbols that might be used in astrological, magical, mathematical texts, etc., nor indicators of changes of column size, letter height, and the like. We simply cannot expect to be able to provide a scheme, amenable to all of the great variety of Aramaic texts, that would allow an accurate physical reconstruction of the original document solely on the basis of the machine-readable version. Should the occasion warrant it, after all, publications of such documents based on their machine-readable versions can contain photographs and prose introductions to convey this additional information, and as a last resort one can simply insert such symbols into the printed computer output prior to publication. Nevertheless, editors who encounter orthographic peculiarities for which we have made no provision should feel free to consult with the CAL staff in order to develop appropriate *ad hoc* solutions.

GENERAL PRINCIPLES

The following general principles are presupposed by the format scheme herein presented:

1. Wherever possible, within the limits imposed by the size of the set of printable ASCII characters available on all modern computer systems, a single sign in the source document is represented by a single character in the machine-readable version.

2. The machine-readable version of the Aramaic text (i.e. the coded form of the text used internally by the project as opposed to the printed output of the same text in an Aramaic alphabet) should be easily decipherable by anyone knowing Aramaic without intensive, special training. Thus we prefer a transliteration scheme that renders טב by "TAb" rather than "+FB"!

3. Generally the texts are not to be "tagged" in any way, i.e. no grammatical or lexical information is to be added to the raw text as it is transferred from manuscript source to machine-readable version. As specialists well know, in the case of alphabetic Semitic texts the absence of such "tags" makes subsequent machine processing many times more difficult; but given the size of the corpus, manual tagging of the texts would simply delay the completion of the project beyond any foreseeable future. Moreover, it

would effectively rule out the direct use of data acquired through the new optical character recognition (OCR) technology in which the textual data is transferred directly from the printed (or handwritten!) page into the computer rather than via the intermediary of a human typist who enters the data manually. On the other hand, occasionally one encounters a form guaranteed to confuse even the most sophisticated of processing algorithms. Provision is therefore made to allow for the insertion of a tag to facilitate subsequent processing. Obviously, such tags will be deleted from published text editions (and from the machine-readable text itself after initial processing has been completed). After processing, however, all analytical information will be available in the CAL database, and the editors may well see fit to respond positively to requests by qualified parties to be supplied with fully tagged versions of specific texts.

4. The groundbreaking work on computer processing of Semitic texts was accomplished using long-since antiquated mainframe computers with severely limited character sets and development environments. Accordingly and perhaps unfortunately, the fruit of that pioneering work has little to offer us now. Therefore, the CAL is committed to taking complete advantage of the tremendous advances that have occurred in recent years in all areas of computing and to enable colleagues throughout the world to participate in both the input (i.e. preparation of machine readable text for submission to the computer for storage and analysis) and output (i.e. the result of that analysis) stages of the project using the tools already at their disposal. On the other hand, the CAL staff obviously cannot provide special software for even a small selection of the myriad of hardware configurations now in common use by colleagues around the world in order to help them prepare the input for or process the output from the project. Thus the project must be prepared to accept input texts in a variety of formats and media, while output texts must be relatively accessible even to those without special, dedicated software. Wherever possible, however, effort will be made to make such software available for the most frequent hardware configurations. The CAL uses the UNIX computer operating system. The increasing popularity of UNIX in both academic and personal computing environments should enable us to make the results of the CAL staff's development efforts directly available to an even wider audience.

It is the current intent of the project to provide complete software support for editors and contributing lexicographers working in UNIX and PC/MS-DOS computing environments.

CODING PRINCIPLES II

(NOTE: The CAL staff is prepared to provide prospective editors with guidance in the selection and acquisition of appropriate hardware and software tools to facilitate the process of text entry/editing, especially as regards the ability to do right-to-left text entry using a Semitic alphabet. Practically any kind of computer can be used to enter/edit text in the ROMAN character set, however. In any case, we strongly recommend that each editor acquire access to a printer that accepts downloadable character sets, such as the Epson FX-80 line and compatibles. Using such a printer and programs provided by the CAL, left-to-right, ROMAN encoded text can be printed out in an appropriate Aramaic script from right-to-left, a procedure that can be of substantial aid to the process of collation.)

TRANSLITERATION

Consonants

Text may be created on text editors using either Roman characters or Hebrew (Aramaic) characters, on personal computers or mainframe computers, and may use either the ASCII (7 bit) or Extended ASCII (8 bit) character sets. (Other formats must be converted to ASCII before submission.) Input text for any Aramaic dialect will be accepted in any of the following three *consonantal* character sets: ROMAN(I), and two HEBREW sets(II, 7 bit; III, 8 bit):

	א ב ג ד ה ו ז ח ט י ך כ ל ם מ ן נ ס ע ף פ ץ צ ק ר ש ת
I	) b g d h w z x T y K k l M m N n s (P p C c q r $ & t
II	& a b c d e f g h i j k l m n o p q r s t u v w x yy~ z
III	א ב ג ד ה ו ז ח ט י ך כ ל ם מ ן נ ס ע ף פ ץ צ ק ר ש ת

(where א is ASCII 128, ב is 129, etc. through ת=154.)

Using any of these sets, the emphatic "P" (=Greek *pi*) of Christian Palestinian Aramaic (or, if used, in standard Syriac texts) is to be indicated by the same sign used for final *peh* in Jewish Aramaic texts.

The term "Hebrew Set III" refers to the ASCII correspondences, not the shape of the letters. Those entering Syriac texts may well prefer to work with a "Hebrew Set III" font whose letters are shaped according to one of the Syriac scripts.

The distinction between *shin* and *sin* is indicated only where so marked in the source text—i.e. in Biblical Aramaic or in a Jewish Aramaic text with Tiberian pointing. Otherwise the code for *shin* is always employed.

The characters in set II, when entered on an 8 bit machine, may have their high bits either on or off.

In addition, a fourth character set MANDAIC (IV), derived from the transliteration scheme (Drower-Macuch) now current in Mandaic studies, may be used for Mandaic text only:

IV **a b g d h u z H T i k l m n s (p S q r $ t** and **D** for "d̲".

Note that the pronominal suffix is coded "**H**" while the normal *heh/ḥet* sign (although historically a *ḥet*!) is simply "**h**"; so, too, if the ROMAN set is used for Mandaic. When using either of the HEBREW sets for Mandaic text, the sign for *ḥet* (i.e. "**ג**" or "**ח**") is to be used for simple "**h**", while the sign for *heh* is used for the pronominal suffix; and in those sets the Mandaic ligature "d̲" is to be indicated by the sign for *dalet* followed by a plus ("**+**") sign. In the ROMAN set, as in the MANDAIC set, the ligature is indicated by upper case "**D**". The double dot diacritic used to indicate Arabic or Persian characters should rarely be needed, but can be indicated by "**:**" after the letter. So, too, is Arabic *ayin* indicated by "**(:**".

Case is always significant. Only one of these character sets may be used in a single file. With the exception of Mandaic "d̲", all *ligatures must be resolved* into their constituent consonants.

Note that texts encoded in Hebrew Set II, *or in any set using vowels*, must still be encoded sequentially "left-to-right" (as will be done naturally by any true "Hebrew speaking" computer even though it displays and prints from right to left) so that there will be a simple one-to-one correspondence between texts encoded in the various sets, except for sequences of numerals and concave/convex editorial markings (see below) which will be reversed in the coded Hebrew form. Thus a text that reads

למלכא 32 שנת <בּ>{ל}

will appear encoded in the physical file as follows:

set I **{l}<b>$nt 32 lmlk)**
II **}l{>a<ypz 23 lnlk&**
III **}ל{>בּ<תנש 23 אכלמל**
IV **{l}<b>$nt 32 lmlka** (of course in Mandaic it would have been spelled "**$nat 32 lmalka**")

On the other hand, text encoded in set III *without vowels* (but with or without other diacritics) may be encoded either left-to-right or right-to-left, the latter to accommodate

those whose computers can display Semitic characters on the screen. (Scholars should feel free to consult with the CAL staff for guidance in finding the best hardware and/or software configuration for their needs.) In such a case (i.e. right-to-left display) the above text would be rendered internally as:

```
)klml 32 tn$<b>{l}
```

Note that the numerals remain in left-to-right order, while the editorial bracket pairs are inverted in terms of the flow of the text.

Final Consonants: The final consonants "KMNCP" of ROMAN set I are to be used only for Jewish Aramaic texts. In the Hebrew sets, to facilitate text entry and correction, final consonants may be used for texts transcribed from other scripts as well, without regard to the use of such distinctions in the source script. NB: CPA texts must be transcribed *without* the use of final consonants.

Vowels

In general vowels are to be indicated only where their linguistic value is assured, i.e. a late vocalization tradition, even on an early text, is not to be rendered. In such cases, what is required is a sign for sign transliteration rather than a phonetic transcription of any kind. There are two sets of vowel signs, depending on the consonant set used. Set A (preferred, because of its transparency) is used with consonantal sets I and III. Set B is used with consonantal set II:

A	B	Tiberian	Babylonian	Palestinian	Jacobite	Nestorian
a	1	◌ַ	◌	◌̄	◌	◌
A	2	◌ָ	◌	◌	◌	◌
e	3	◌ֶ		◌	◌̈	◌
E	4	◌ֵ	◌	◌		◌
i	5	◌ִ	◌	◌	◌̄	◌
u	6	◌ֻ	◌	◌ or ◌	◌́	(e)◌
U	7	וּ◌		◌		
o	8	◌ֹ	◌	◌		(o)◌
O	9	וֹ			◌̇	
:	0	◌ְ	◌̄			
:a	01	◌ֲ				
:e	03	◌ֱ				
:o	08	◌ֳ				

In all cases the vowel indicator is to appear after the consonant with which it is pronounced and before any vowel letter that also indicates that vowel. Thus מַלְכִין is "mal:kiyN". NB: While West Syriac (Jacobite) ܡܘܢ and East Syriac (Nestorian) ܡܘܢ are both "quwm", Babylonian קוֹם is "quwM" while Tiberian קוּם is "qUM"; so, too, with "o" class vowels. The "o" vowel of mixed Syriac texts should be indicated with "w." rather than "ow" (see below, "Diacritics, e".)

A common but somewhat exceptional case is presented by initial *Yod* words in Syriac: "he sat", spelled "yiteb" ܝܬܒ is to be transcribed as pronounced: "iyteb".

those whose computers can display Semitic characters on the screen. (Scholars should feel free to consult with the CAL staff for guidance in finding the best hardware and/or software configuration for their needs.) In such a case (i.e. right-to-left display) the above text would be rendered internally as:

```
)klml 32 tn$<b>{l}
```

Note that the numerals remain in left-to-right order, while the editorial bracket pairs are inverted in terms of the flow of the text.

Final Consonants: The final consonants "KMNCP" of ROMAN set I are to be used only for Jewish Aramaic texts. In the Hebrew sets, to facilitate text entry and correction, final consonants may be used for texts transcribed from other scripts as well, without regard to the use of such distinctions in the source script. NB: CPA texts must be transcribed *without* the use of final consonants.

Vowels

In general vowels are to be indicated only where their linguistic value is assured, i.e. a late vocalization tradition, even on an early text, is not to be rendered. In such cases, what is required is a sign for sign transliteration rather than a phonetic transcription of any kind. There are two sets of vowel signs, depending on the consonant set used. Set A (preferred, because of its transparency) is used with consonantal sets I and III. Set B is used with consonantal set II:

A	B	Tiberian	Babylonian	Palestinian	Jacobite	Nestorian
a	1	◌	◌	◌	◌	◌
A	2	◌	◌	◌	◌	◌
e	3	◌		◌	◌	◌
E	4	◌	◌	◌		◌
i	5	◌	◌	◌	◌	◌
u	6	◌	◌	◌ or ◌	◌	(o)◌
U	7	◌		◌		
o	8	◌	◌	◌		(o)◌
O	9	◌			◌	
:	0	◌	◌			
:a	01	◌				
:e	03	◌				
:o	08	◌				

In all cases the vowel indicator is to appear after the consonant with which it is pronounced and before any vowel letter that also indicates that vowel. Thus מַלְכִין is "mal:kiyN". NB: While West Syriac (Jacobite) ܩܘܡ and East Syriac (Nestorian) ܩܵܡ are both "quwm", Babylonian קוֹם is "quwM" while Tiberian קוֹם is "qUM"; so, too, with "o" class vowels. The "o" vowel of mixed Syriac texts should be indicated with "w." rather than "ow" (see below, "Diacritics, e".)

A common but somewhat exceptional case is presented by initial *Yod* words in Syriac: "he sat", spelled "yiteb" ܝܬܺܒ is to be transcribed as pronounced: "iyteb".

If initial *Alaf* is present, however, either as a full consonant (ܐܠܦ) or a supralinear mark (ܐܠ̈), the correct coding is "**)iyteb**".

Diacritics

Jewish Aramaic:

a. *dagesh*	.	(period)
b. *raphe* (only if in text)	_	(underline)
c. single abbreviation mark (in early texts a supraliner point)	'	(single quote)
d. double abbreviation mark (i.e. separating initials)	''	(two single quotes)
e. *maqqeph*	+	

Syriac:

e. supralinear point	.	(period)
f. sublinear point	_	(underline)
g. *syame*	:	(colon)
h. single supralinear line	~	(tilde)
i. sublinear line	+	(plus)
j. word length supralinear line	'	(single quote at end of word) (abbreviation)

(supralinear lines spanning more than one word are not indicated.)

Note that in both systems the sigla for spirantization vs. non-spirantization of BGDKPT consonants are the same; hence the somewhat counter-intuitive use of the period to indicate a *supralinear* point in Syriac.

With the exception of the abbreviation indicators (c, d, j), all these characters are to be placed between the consonant and its vowel. Thus:

בְּנֵי is "**b.:nEy**"

Numbers

Numerical symbols may be "translated" into the decimal system and indicated by the ASCII digits "0" through "9"; e.g. "**b$nt 32 lmlk)**". Where it is deemed prudent to indicate the presence of separate symbols for various portions of a composite number, however, the following approach may be used:

"**b$nt 20+10+2 lmlk)**"

i.e. in the text itself the number is indicated by a sign for "**20**" followed by a sign for "**10**" followed by "**2**". This method is especially useful in dealing with fragmentary texts where the individual symbols of some numbers are missing or uncertain.

In some text traditions, numerals are combined with letters into a symbolic unit. Editorial additions should be used to resolve these into their constituents; thus "**`1lp`**" = "one thousand" is to be coded "**`1<)>lp`**"

In a vocalized text using consonant set II such numbers must be preceded by a plus ("**+**") sign in order to distinguish them from the symbols for the vowel signs.

Punctuation

Provision is made for the indication of four levels of punctuation, normally to be used as follows:

| (vertical bar) word divider (primarily in early texts). A space (ASCII 32) should appear on either side of the word divider as an aid to the human reader. Note that all texts must be divided into words. Thus, if a text uses spacing to mark word division (the nearly universal practice in later texts), a space in the original will be reflected by a space in the computerized version; but a text without consistent use of such inter-word spacing will also be rendered using spaces between words. If the latter is the case, this fact should be noted in the "description" field of either the "text header" or the "file header" (see below). In those texts where a physical word divider sign is normally found, spaces must still be used to separate words in those cases where a word divider is not used in the original.

, (comma) minor phrase divider.

; (semi-colon) major phrase divider.

% (percent) verse/sentence/paragraph divider.

In Biblical Aramaic and Jewish Aramaic texts "**%**" is used for "sof-pasuq".
In Syriac texts:

, corresponds to a single dot.
; corresponds to a double dot.
% corresponds to a four/five dot diamond.

NB: With the exception of "**%**", such punctuation marks are to be used only to correspond to punctuation signs present in the original text. The sign "**%**" is the mark of choice to indication text division indicated by format in the original; e.g. spaces between paragraphs in early Jewish Aramaic scrolls. (Cf. below, "Illustrations".) It may also be used to divide textual material that appears in the source on the same logical line but that is not to be treated as syntactically sequential, a typical occurrence in Talmudic mss., for example.

Unless it is the final character on a line, each punctuation mark is followed by a space. But, unless separated in an unusual manner from the preceding word in the source, *no* space precedes the punctuation mark.

EDITORIAL SIGLA

NB: The text base of the CAL is organized on the principal of "logical lines" (see below, "3. TEXT FORMAT"). *In no case may the closing member of a pair of sigla (e.g. } or]) be "held over" to the next logical line.*

Deletions

{ } - text deleted by the editor.
{{ }} - text deleted/erased in the manuscript.

Additions

<___> - text added by the editor.
^___^ - text added in the manuscript either interlinearly or over an erasure.
/___ /- alternative text.
#___# - source of alternative reading.

Notes on usage:

An editorial emendation is indicated by a combination of curly brackets ("braces") and angle brackets; e.g. <]>{א}טעי means that the manuscript has טעיא but טעין is to be read; in ROMAN: "T(y{)}<N>". *Note that the deleted text precedes the editorial addition.* Where such an emendation is based on another manuscript of the same text, such information is indicated with "pound" signs "# #" immediately following the emendation; thus:

למתאסייא {יתיה}#01#

means that the basic manuscript has the second word יתיה but that it is to be deleted (and will be neither concorded in the database nor lexically analyzed by the CAL) on the authority of the alternate manuscript whose siglum is "01" (see below, "3. TEXT FORMAT, Text Header"). Where more than one variant ms. supports a reading, the respective sigla are separated by commas, e.g. #02,M2#.

Where variant readings are of equal value, *and/or both are to be processed by the CAL*, paired slashes are used, whether the source of such alternation lies in alternate manuscript traditions (a fact that must be indicated), e.g.:

(bdw/(brw#M2#/

or is simply due to orthographic uncertainty (in which case there is no source-of-variant indicator):

(bdw/(brw/

Note that usually the entire word is repeated between the slashes, not just that portion of the word that is uncertain, and that there is no space either before or after the slash separating the variants. If the variant is longer than a single word or there is more than one variant portion in a single word, then the beginning of the parallel portion of the main text must be indicated by a "backslash" "\". Again, full words are preferable.

Thus both:

a. /#03#חיתו ברא/א דחקלא

and

b. /#03#חית ברא/חיתא דחקלא\

mean that the main text has חית ברא while variant ms. #03 has חיתא דחקלא, and each coding is treated identically by the computer, but coding b. is clearly superior for the needs of the human editor and proofreader. Where emendations are involved, however, minimal material should be included within markers. Thus in the previous example, if the variant were preferable to the main text the preferred coding would be:

```
xyt{ br)}<) dxql)>#03#
```

In Biblical Aramaic *ktib/qre* variations are treated like any other variant. The *ktib* is given in *consonantal* form only and the *qre* variant marker is simply "**#q#**"; e.g. at Daniel 2:26:

```
h)ytK/#q#ha)iytAyK:/
```

Cancelling a Scribal Deletion: As mentioned above, material within deletion braces (**{ }** and **{{ }}**) is normally neither concorded nor lexically analyzed. Often, however, text that has been deleted by a scribe is of philological merit and should be analyzed. In such cases the double braces indicating scribal deletion are to be surrounded by angle brackets, e.g:

```
<{{)mrt qdM}}>
```

If only a portion of the deletion is correct, repeat the correct segment within brackets; thus

```
{{)mrt qdM}}<qdM>
```

means that the scribe has marked "**)mrt qdM**" as deleted but that "**qdM**" is correct. Frequently scribally deleted text is illegible; this is easily indicated by "**{{...}}**".

Marginalia are assigned the default manuscript variant siglum "**M0**", and are to be cited within slashes like any other alternate reading. Where the marginal note is an addition to the main text there will obviously be no base text whose alternate it is, so the proper coding is simply, e.g.:

```
w)mr \/mlk) l(bdh/#M0#
```

where the "\" notes the point in the main text where the marginal addition is to be inserted. In manuscripts where various marginal and/or interlinear hands can be distinguished, these hands must be assigned distinct sigla and the referent of such sigla indicated in the proper place in the "File Header" (see "3. TEXT FORMAT"). Thus, once "**M2**" has been appropriately "defined" in the File Header,

```
w)mr ^mlk) l(bdh^#M2#
```

indicates an *interlinear* addition by the second "marginalia" hand. Similarly, **{{mlk) l(bdh)}#M2#** would indicate scribal deletion by the same hand.

Glosses

A gloss is an explanation of a term in the text rather than an addition or variant, and may be marginal or interlinear. Glosses are indicated immediately after the term glossed (without spaces) as follows:

^GL text of gloss^

If the text of the gloss is not in Aramaic, the normal sigla are used to indicate the language of the gloss (see below). If the glossator's hand can be identified, then an appropriate Arabic numeral follows "**G**"—e.g. "**^GL2 text of gloss^**" indicates a gloss by the second hand.

Broken Text

If a letter is partially broken or effaced *but the reading is certain* on the basis of *physical* evidence, no special siglum is used. Where the reading is uncertain but substantial physical evidence of the sign is present, a *question mark* "**?**" is placed after the letter. (Where two or three different readings are possible, they are given within slashes like any variant, the questionable letter indicated by "**?**" in each variant.) A broken sign for which no reading can be suggested at all may be indicated by a period.

All other broken text is placed within *square brackets* "**[ ]**". The following rules apply:

1. Square brackets (like all paired marks) must always be paired on a single "logical line".
2. The size of a broken passage is indicated by a period for each unknown letter. At the edge of a column of text, where the precise size of the gap is unknown, it should be estimated.
3. All textual restorations based primarily on context rather than physical remains of signs are to be given within square brackets. *Complete words* in such restorations *will be concorded*, but will not be used in the lexical database of the CAL. Care must be taken in order to indicate such a "complete word" to the computer. In the sequence "**[...m]lk)**" the computer will not see a complete word. This must be indicated rather as "**[.. m]lk)**", where the space within the brackets indicates that the *mem* begins a new word.

The extent to which missing material is restored within the square brackets will vary from editor to editor, but should generally be limited in extent and must be consistent in practice. Although the KWIC concordance will always indicate that a given word is part of a hypothetical restoration, if too much of the material is hypothetical the value of the concordance as a research tool is considerably diminished.

Spacing

As regards the use of any of the editorial sigla, one cannot overemphasize the importance of the proper use of spacing. Since "braces" mean "omit this material," *spaces enclosed in braces will not be "seen" by the processing programs*, resulting in the erroneous concatenation of separate words. For example, consider the following simple

text:

brh dy mlk)

If the editor wishes to indicate that the correct reading is *br mlk)*, then the proper coding is:

br{h dy} mlk)

If the braces are placed like this:

br{ dy }mlk)

the computer will analyze the text as a single word: *brmlk)*.

Carelessly added spaces will also lead to incorrect conclusions. E.g. (cf. the above example) "**(bdw /(brw/**" does not mean that "**(brw**" is an alternate reading of "**(bdw**" but that one reading is just "**(bdw**" and the other is "**(bdw (brw**" (or even "**(bdw(brw**!"), for as given "**/(brw/**" is the variant of a space, not of the first word.

Abbreviations

Common abbreviations may be rendered as in the manuscript; e.g. "**)' r'**" = "**)mr rby**", "**)mrw rbnN**", etc.) In most cases, however, the editor should indicate his understanding of the abbreviation by adding the omitted text within angle brackets (after the abbreviation mark, if there is one): "**y$'<r)l>**", "**$<ql>**".

The "at" (@) Sign

This sign (ASCII 64) serves to connect material physically separated in the source. It has two major uses:

1. Where a single lexeme is written as more than one word in the source, "**@**" is used instead of a space to represent the space in the ms.; e.g. "**kl@qwbl**". Usage here is to be guided by the CAL's *Outline Lexicon*.
2. In inscriptional material a word may be split at the end of a physical line and continued on the next. In this case "**@**" must be appended to the end of the first line to indicate that its final word is incomplete. (If the first line is broken at its end, the "**@**" sign appears *outside* of the closing "**]**".)
3. See above, "Numbers", for use of this sign as a "multiplier."

The "`" Sign (ASCII 96)

On rare occasions a single "word" in the source may result from the conflation of two or more usually separate words, whether due to unusual scribal practice or to error. In the former case the front quote (reverse apostrophe "`") should be added to separate the individual words. In the latter, the missing space should be added in angle brackets as would any other editorial addition.

Catchwords and "Custodes"

In some scribal traditions, incomplete lines are filled out by inserting the first few letters of the initial word of the following line ("custodes"). Such letters may be omitted from the transcription of mss. of canonical texts but should be included—enclosed in braces—in editions of unique texts where the possibility of misinterpretation exists. End-of-page catchwords are always included—likewise enclosed in braces to indicate their exclusion from the flow of the text—unless, of course, the first word of the following page is otherwise unattested.

EXTRANEOUS MATERIAL

Aramaic text is frequently imbedded in text in another language—particularly Hebrew, and Aramaic manuscripts may contain material from other languages cited in both Aramaic and other scripts, and may contain symbols and illustrations; all such material extraneous to the interest of the CAL but worthy of indication in the machine-readable version of the text. The following principles apply:

1. Text in Aramaic characters but not in the Aramaic language (e.g. Hebrew, Karshuni) is to be enclosed in regular (double) quotation marks. Quotation marks may also be used to surround items in Syriac text (e.g. the names of letters) not to be taken as words, marked in the original ms. with a supralinear line. Material within such quotation marks will not be concorded nor lexically analyzed but will remain available for recall in the CAL's database. Typically, then, quotations from earlier texts already part of the database (particularly quotations of the Peshitta in Syriac texts) should also be excluded using quotation marks. (If such quoted material is a significant variant to the original text, however, it should be given a variant number and cited in the file of the original text.) As with all editorial marks, *paired quotation marks may not extend over a single logical line*. Great care must be taken to assure that the closing quotation mark is not forgotten, lest the computer "forget" to analyze all the subsequent text of that logical line!
2. As explained in the following paragraphs, all other material is to be indicated using paired asterisks; thus:

a.	***I.....I***	= illustrations.
b.	***S** S....**S***	= symbols.
c.	***G** text **G***	= Greek text.
d.	***A** text **A***	= Arabic text.
e.	***P** text **P***	= Persian text.
f.	***F** text **F***	= any other foreign text.
g.	*****text<lemma*****	= grammatical tag.

(In a-f the upper case letters are required.)

Information within the paired asterisks is to be coded so as to appear properly from *left to right* on the editor's screen, i.e. it will be in inverted order in left to right HEBREW coded texts (sets II and III).

Illustrations

Simple:

Wherever the text is combined with illustrations in such a way that the interpretation of the flow of the text is dependent on its relationship to such illustrations, the location of the illustration should be indicated in the machine-readable version. Within the opening and closing "**I**'s", periods or a verbal description, indicate the extent of the illustration. Normally the illustration will interrupt several lines of text; thus the format:

```
11  text1   *I.....I*   text2
12  text3   *I.....I*   text4
13  text5   *I.....I*
14          *I.....I*   text6
```

should be used where "text2" follows upon "text1", "text4" follows "text3", etc. (For the meaning and use of the line numbers in the left most column see "3. TEXT FORMAT".) Were the text in the above example to be read in columnar fashion, however, the following procedure would be used so that the graphic relationship of the material can still be determined from the machine-readable version:

```
11  text1
12  text3
13  text5
20  *I 7 characters × 4 lines I*
31  text2
32  text4
34  text6
```

In this case the line numbering indicates that the illustration constitutes a column in the original, separating text columns 1 and 3.

Complex:

A complex illustration is one that includes text within it. The location of the text should be described in the illustration indicator and the text itself cut off from the framework textual material. For example, if the illustration in the above example contained some text, it could be indicated as follows:

```
20  *I occupying space of 7 characters × 4 lines, including
20  the following text I*
21  % included text %
```

Note the delimiter "percent" signs on either side of the included text. These indicate that this text stands alone and is not to be read sequentially either with what precedes or what follows.

Symbols

Many texts contain symbols of various kinds in-line with the Aramaic text. Each such symbol is indicated by ***S***, or, where multiple symbols occur, by a series of "**S**'s" between a set of asterisks; e.g. ***S S S S***. If the symbol is between words then there must be a space outside of both asterisks. If the symbol comes within a word, then there must be no space between the asterisks and the letters of the word; e.g. "**ml*S*k)**". For symbols indicating numeric values, see above, "Numbers". (For an example see text 016 of EXAMPLE A in the Sample Texts.)

Greek

Greek text is transliterated using the following code:

Letters:

A B G D E Z H Q I K L M N C O P R S T U F X Y W
a b g d e z h q i k l m n c o p r s t u f x y w

Diacritics:

Accents (coded after their vowel):

acute - "**/**"
grave - "****"
circumflex - "**=**"

Breathing (coded before the vowel /r): "**(**" "**)**" (also coronis)
iota subscript (after vowel):"**|**"

Punctuation, if any, is indicated by the equivalent ASCII sign.

Arabic

Arabic text is transliterated (*not* transcribed) using the following code:

Consonants (in the order of the Arabic alphabet):

) b t j g x X d J r z s $ S D T Z (G f q k l m n h w y
(**H** = *ta marbuta*)

Vowels: **u i a**; with *tanwin*: **U I A**.

Diacritics: ~ (tilde) = *tashdid*.
(As in Aramaic text, diacritics follow their consonant and precede the vowel.)

Persian

Persian is transliterated as follows:

) b p t H j c x X d J r z # s $ S D T Z (G f q k g l m n w h y

Tags

If the text within a pair of asterisks does not begin with upper case **I**, **S**, **G**, **A**, **P**, or **F** or a numeral (see "3. TEXT FORMAT, Text"), then the concordance software assumes it is a grammatical tag and will use it to analyze the *previous* word(s). Such tags should be used sparingly enough so as to avoid unnecessary complications and delay in the process of text entry. Typically, only texts from late and inconsistent scribal traditions will require much tagging. The tag consists of three parts: (1) A morphologically equivalent version of the form, that is one that the computer presumably can handle with fewer difficulties; (2) the character "**<**" meaning

"derived from"; (3) the lemma (headword) under which this form is to be "booked." In a Jewish Palestinian Aramaic inscription, for example, one might well encounter the expression דהב דהב meaning "who gave gold." Unless the editor knows that this construction is already attested in the analyzed files of the CAL, such a confusing expression should be tagged; viz:

dhb*dyhb<yhb* dhb*dhb<dhb)*

Editors using tags should refer to the CAL's *Outline Lexicon* to determine the correct lemma for each entry. In the case of homonymous lemmas, the appropriate lemma number follows the lemma and is preceded by a "pound" sign: e.g.

wmr*w)mr<)mr#2*

NB: A form should be tagged only the first time that it occurs in a single file. Under no circumstances should *more* than 50 tags occur in one file (for subsequent tags will be ignored). If a given file contains frequent attestations of unusually written words whose lemma is not yet found in the *Outline Lexicon*, the editor may submit a list of these forms and their analysis in a separate text file on the same media. (Such a file should be named "README" and may also or instead contain other information that the editor thinks may be of value to the staff of the CAL during the input and analysis stages of processing.) Again, no more than 50 such items should be listed. As a rough guideline, a given form should occur at least four times in a single file before it should be included in such a list.

PROPER NOUNS

To facilitate the extraction of proper nouns from the text base, they are marked in the machine-readable text as follows:

Personal names:	preceded by "-" (hyphen).
Geographical names:	preceded by "=" (equal).

If the interpretation of the text is uncertain, the proper noun marker (hyphen or equal sign) is followed by a question mark.

Divine names and gentilics will be included in the CAL and should not be marked (with the exception of the Tetragrammaton, abbreviations thereof, and אלהים in Jewish texts, which may be marked with a hyphen).

If the name is preceded by an inseparable preposition, the hyphen or equal sign is placed *between* the preposition and the name. If a marked name contains any spaces, such spaces must be replaced by the "@" sign.

COMMENTS

Where absolutely necessary the editor may incorporate comments into the machine-readable text. This feature should be used only in exceptional instances. Comments must be complete lines of text and are indicated by the presence of an "!" (exclamation point) as the first character in the line. Normally, comments are solely for the aid of the editor him/herself and the CAL staff, and (like grammatical tags)

will be removed from the text during the first stages of processing. In rare cases the editor will wish to insert a comment that will appear in a subsequent, published version of the text. Such comments are introduced by *two* exclamation points; e.g.

"!! 3 folios missing here"

APPENDIX A

DIALECT CODES

- 10 Old Aramaic
  - 11 Fakheriyeh
  - 12 Samalian
  - 13 Standard Syrian
  - 14 Mesopotamian
  - 15 Deir 'Alla
- 20 Imperial Aramaic
  - 21 Neo-Babylonian Period
  - 22 Egyptian (informal)
  - 23 Egyptian (formal)
  - 24 Persian in Egypt
  - 25 Persian in Persia
  - 26 Samaria
  - 27 Inscriptions - East
  - 28 Inscriptions - West
- 30 Biblical Aramaic
  - 31 Ezra
  - 32 Daniel
- 40 Middle Aramaic
  - 41 Palmyrene
  - 42 Nabatean
  - 43 Hatran
  - 44 Qumran
  - 45 Other
- 50 Palestinian
  - 51 Jewish Literary Ar.
  - 52 Inscriptions
  - 53 Galilean (Yerushalmi talmud and midrash)
  - 54 Targumic
  - 55 CPA
  - 56 Samaritan
- 60 Syriac
  - 61 Inscriptions
  - 62 Peshitta
  - 63 Eastern
  - 64 Western

  (non-specific Syriac use just "60")
- 70 Babylonian
  - 71 Talmudic
  - 72 Gaonic
  - 73 Jewish Magic Bowls
  - 74 Mandaic
- 80 Other
  - 81 Late Jewish Literary
- 90 Modern Aramaic
  - 91 Ma'lula et al.
  - 92 Turoyo et al.
  - 93 Eastern dialects

TEXT CATEGORIES

- 000 Bible Text
- 100 Bible Exegesis
- 200 Monumental
- 300 Legal
- 350 Administrative
- 400 Religious
- 420 Liturgical
- 450 Theological/philosophical
- 500 Historical
- 550 Epistolary
- 600 Literary
- 700 Magic
- 800 Scientific
- 900 Philological

APPENDIX B

Printable ASCII and Extended ASCII characters used in the CAL formats: (excluding the Arabic, Greek and Persian character sets, p. 17)

#	Character	Functions	Page Ref.
32	(space)	word divider, field divider	9, 13, 23
33	!	introduces a comment	18
34	"	surrounds foreign or quoted text	15
35	#	surrounds variant code	10
36	$	*shin* in ROMAN & MANDAIC sets	5
37	%	major punctuation mark	9
		surrounds Aramaic in descriptions	16, 24, 27
38	&	*sin* ROMAN; *aleph* in HEBREW II	5
39	'	abbreviation mark	8
40	(	*ayin* in ROMAN & MANDAIC sets	5, 6
41	)	*aleph* in ROMAN	5
42	*	surrounds extraneous material	15, 24
43	+	*maqqeph*, sublinear line,	8, 6
		in MANDAIC d̲	
		before numerals in II	8, 29
44	,	minor punctuation mark	9
45	-	preceding personal names	18
46	.	*dagesh*, supralinear point,	8
		extent of break/illustration	12, 15
47	/	surrounds alternative text	10
48	0	0, *shwa* and *hataf* vowels	7
49	1	1, short *a* vowel	7
50	2	2, long *a* vowel	7
51	3	3, short *e* vowel	7
52	4	4, long *e* vowel	7
53	5	5, short *i* vowel	7
54	6	6, short *u* vowel	7
55	7	7, Tiberian *shuruq*	7
56	8	8, short *o* vowel	7
57	9	9, Tiberian *holem* with *waw*	7
58	:	*syame* (Syriac), *shwa* (Jew.Ar.)	7, 8
		dbl. dot diacritic (MANDAIC)	6
59	;	minor punctuation mark	9
60	<	opens editorial additions	10
61	=	preceding geographical names	18
62	>	closes editorial additions	10
63	?	questionable letter	12

#	Character	Functions	Page Ref.
64	@	lexically ignored "word" divider,	13, 18, 30
		line connector in Variant List	25
65	A	long *a* vowel, "Arabic"	7, 15
67	C	final *sade* in ROMAN	5
68	D	MANDAIC d	6
69	E	long *e* vowel	7
71	G	"Greek"	15
72	H	final h in MANDAIC	6
73	I	"Illustration"	15
75	K	final *kaf* in ROMAN	5
77	M	final *mem* in ROMAN	5
78	N	final *nun* in ROMAN	5
79	O	Tiberian *holem* with *waw*	7
80	P	final *peh*, CPA emphatic *peh*	5
83	S	*sade* in MANDAIC	6
84	T	*tet* in ROMAN and MANDAIC	5, 6
85	U	Tiberian *shuruq*	7
91	[	begins break in text	12
92	\	point in main text where variant begins	11
93	]	ends break in text	12, 13
94	^	encloses scribal addition	10
95	_	sublinear point, encloses	8
		underlined text (descriptions)	24
96	‘	separates words written together	14

		Roman	Vowel	HEBREW II	MANDAIC	
97	a		short *a*	ב	א	5, 7
98	b	ב		ג	ב	5
99	c	צ		ד		5
100	d	ד		ה	ד	5
101	e		short *e*	ו		5, 7
102	f			ז		5
103	g	ג		ח	ג	5
104	h	ה		ט	ח	5
105	i		short *i*	י	י	5, 7
106	j			ך		5
107	k	כ		כ	כ	5
108	l	ל		ל	ל	5
109	m	מ		ם	נ	5
110	n	נ		מ	נ	5
111	o		short *o*	ן		5, 7
112	p	פ		נ	פ	5
113	q	ק		ס	ק	5
114	r	ר		ע	ר	5
115	s	ס		ף	ס	5
116	t	ת		פ	ת	5
117	u		short *u*	ץ	ו	5
118	v			צ		5

#	Character	Roman	Vowel	HEBREW II	MANDAIC	Page Ref.
119	w	ו		ק		5
120	x	ח		ר		5
121	y	י		ש		5
122	z	ז		ת	ז	5
			Functions			
123	{		opens deleted text (single = editor, double = ms)			10
124	|		word divider			9
125	}		closes deleted text (single = editor, double = ms)			10
126	~		marks *sin*, supralinear line			5, 8
128	א		HEBREW III			5
129	ב		HEBREW III			5
130	ג		HEBREW III			5
131	ד		HEBREW III			5
132	ה		HEBREW III			5
133	ו		HEBREW III			5
134	ז		HEBREW III			5
135	ח		HEBREW III			5
136	ט		HEBREW III			5
137	י		HEBREW III			5
138	ך		HEBREW III			5
139	כ		HEBREW III			5
140	ל		HEBREW III			5
141	ם		HEBREW III			5
142	מ		HEBREW III			5
143	ן		HEBREW III			5
144	נ		HEBREW III			5
145	ס		HEBREW III			5
146	ע		HEBREW III			5
147	ף		HEBREW III			5
148	פ		HEBREW III			5
149	ץ		HEBREW III			5
150	צ		HEBREW III			5
151	ק		HEBREW III			5
152	ר		HEBREW III			5
153	ש		HEBREW III			5
154	ת		HEBREW III			5