Some Initial Reflections on XML Markup for an Image-Based Electronic Edition of the *Brooklyn Museum Aramaic Papyri*

F. W. Dobbs-Allsopp, Princeton Theological Seminary
Contact: chip.dobbs-allsopp@ptsem.edu

Chris Hooker, Princeton Theological Seminary
Contact: christopher.hooker@ptsem.edu

Gregory Murray, Princeton Theological Seminary
Contact: gregory.murray@ptsem.edu

Keywords Aramaic; Brooklyn Museum; critical edition; Elephantine; markup; papyrus; TEI; XML

Abstract:

A collaborative project of the Brooklyn Museum and a number of allied institutions, including Princeton Theological Seminary and West Semitic Research, the *Digital Brooklyn Museum Aramaic Papyri* (*DBMAP*) is to be both an image-based electronic facsimile edition of the important collection of Aramaic papyri from Elephantine housed at the Brooklyn Museum and an archival resource to support ongoing research on these papyri and the public dissemination of knowledge about them. In the process of building out a (partial) prototype of the edition, to serve as a proof of concept, we have discovered little field-specific discussion that might guide our markup decisions. Consequently, here our chief ambition is to initiate such a conversation. After a brief overview of *DBMAP*, we offer some initial reflection on and assessment of XML markup schemes specifically for Semitic texts from the ancient Near East that comply with TEI, CSE, and MEP guidelines. We take as our example BMAP 3 (=*TAD* B3.4) and we focus on markup as pertains to the editorial transcription of this documentary text and to the linguistic analysis of the text's language

About the Authors:

F. W. "Chip" Dobbs-Allsopp is Professor of Old Testament at Princeton Theological Seminary. His research interests include the historical, philological, and literary study of biblical and ancient Near Eastern literature (with special focus on poetry, Northwest Semitic inscriptions) and exploring how new technologies can enhance the editing of ancient Semitic texts. Dobbs-Allsopp's most recent monograph is *On Biblical Poetry* (New York/Oxford: Oxford University Press, 2015).

Christopher Hooker is a PhD candidate at Princeton Theological Seminary.

Gregory Murray is Director of Academic Technology and Digital Scholarship Services at Princeton Theological Seminary Library. He has worked with TEI encoding of humanities texts since 1997 (TEI P3 in SGML) and has extensive experience with text processing and XML technologies, including XSLT and XQuery.

To Cite This Article:

Introduction: Project Overview[i]

A collaborative project of the Brooklyn Museum, Princeton Theological Seminary, and West Semitic Research, the *Digital Brooklyn Museum Aramaic Papyri* (*DBMAP*) is to be both an image-based electronic scholarly edition of the important collection of Aramaic papyri from Elephantine housed at the Brooklyn Museum and an archival resource to support ongoing research on these papyri and the public dissemination of knowledge about them. The collection, consisting of nine whole papyrus rolls (eight of which were still intact, folded, with original cords and sealings upon acquisition) and a large number of fragments (from more than eight other rolls), was bequeathed to the Brooklyn Museum by Ms. Theodora Wilbour in 1947. Ms. Wilbour's father, Charles Edwin Wilbour, had purchased these papyri originally sometime during the period 26 January - 13 February 1893 in Aswan (this according to a notebook entry of his from that time). The papyri were packed in "tin biscuit boxes" and placed in a trunk with other boxes of Egyptian papyri, where they remained (ultimately stored in a New York warehouse) unknown and unread for over half a century—Wilbour died in 1896, without having revealed to anyone the contents of his purchase. As a result of Ms. Wilbour's bequest, an *editio princeps* of these papyri was published in short order by Emil G. Kraeling—*The Brooklyn Museum Aramaic Papyri: New Documents of the Fifth Century B.C. from the Jewish Colony at Elephantine* (= *BMAP*; Kraeling 1953)—some sixty years after their initial discovery. The papyri date to the fifth century BCE and mostly consist of legal documents having to do with the interrelations of two families spanning several generations. Historically, the collection represents the earliest major acquisition of Aramaic papyri related to the ancient military colony at Elephantine.

At the heart of the proposed project is the creation of an image-based, facsimile edition of these spectacular Aramaic papyri. As with all scholarly text editions, whether print-based or digital, the chief task is to provide a reliable and accurate simulation of the underlying source text. MLA's Committee on Scholarly Editing guidelines (= CSE) assert that editors establish reliability by "explicitness and consistency with respect to methods, accuracy with respect to texts, adequacy and appropriateness with respect to documenting editorial principles and practice" (MLA 2011). A primary rationale for undertaking a new critical edition of any text is to improve on what older editions have achieved.

In the case of these papyri there is currently only a single critical (i.e., answering to CSE guidelines) edition available, the *editio princeps* published by Kraeling now almost sixty years ago. The volume in all respects is typical of a standard print-based critical edition, consisting of a long, informative "Historical Introduction" (Kraeling 1953, p. 3-119) and editions of each of the papyri—general description, transcription, English translation, critical commentary. There are also several indices (proper names, words) and a set of photographic plates—one black and white image for each papyrus plus an assortment of other images (e.g., endorsements, unopened papyrus rolls). This edition remains the single most comprehensive treatment of these papyri as a whole, although, of course, scholarship over the last sixty years has greatly improved our understanding of almost every aspect of these papyri. Which is to say, *BMAP*, no matter its historical contributions, can no longer serve as a fully adequate and accurate edition of these texts.

In fact, most contemporary students of these papyri use the edition of these papyri found in the handbook edition of the entire corpus of Aramaic documents from Egypt by Bezalel Porten and Ada Yardeni (= *TAD*; Porten and Yardeni 1986-1989). This latter volume features the insights and readings of the foremost student of the Elephantine Aramaic corpus, Porten, and the exquisite hand drawings of Yardeni. It offers the most accurate rendition of the Brooklyn Museum papyri generally available. But as the title suggests, *TAD* was never envisioned as a critical edition of these texts (e.g., there is no commentary, explicit theory of editing or photographs of the texts and only a very minimal critical apparatus). It is our intention, then, to author a truly critical edition of these papyri, one that aspires to the traditional goals (and

standards) of scholarly editing and that builds on the vast scholarly advances achieved during the period between the appearances of *BMAP* and *TAD* and since.

Reflections on XML Markup

In the process of building out a (partial) prototype of the edition to serve as a proof of concept, we have discovered little field-specific discussion that might guide our markup decisions. Consequently, here our chief ambition is to initiate such a conversation. We offer some initial reflection and assessment of XML markup schemes specifically for Semitic texts from the ancient Near East that comply with TEI (Text Encoding Initiative n.d.), CSE, and the Model Editions Partnership (MEP) guidelines. We take as our example BMAP 3 (=*TAD* B3.4) and we focus on markup as pertains to the editorial transcription of this documentary text and to the morphosyntactic markup (part-of-speech tagging) of the text's language.[ii]

Editorial Transcription

Transcription may be defined as "the effort to report—insofar as typography allows—precisely what the textual inscription of a manuscript consists of" (Meulen and Tanselle 1999, p.201). Our transcription is of a documentary text (i.e., non-literary) in a single copy and is designed to support a facsimile edition, not to stand primarily in its place.[iii] As such, our transcription is what has been traditionally described as a "typographic facsimile," which "attempts to duplicate exactly the appearance of the original source text as far as possible within the limits of modern typesetting technology" (Kline and Perdue 2008, p. 147; cf. Meulen and Tanselle 1999, p. 201-3).[iv] Where we believe explicit editorial comment is warranted (e.g., with regard to scribal alterations or where a reading is graphically ambiguous), users will be pointed to an epigraphic commentary for elaboration and discussion and will always be able to compare the transcription with the digital facsimile.

A chief editorial aim in our transcription, then, is to report "what actually appears in a manuscript" as faithfully as possible (Meulen and Tanselle 1999: 203). Within the limits of folio technology, this has usually required a conscious editorial decision to forego the incorporation of

editorial emendations, for example in order to correct apparent errors. As Meulen and Tanselle (as late as 1999) state, "a text cannot simultaneously be unemended and emended" (1999: 203). In a digital environment this is no longer the case. The <choice> element in the General TEI Guidelines (3.4) specifically

enables the encoder to represent for example a text in its 'original' uncorrected and unaltered form, alongside the same text in one or more 'edited' forms. This usage permits software to switch automatically between one 'view' of a text and another, so that (for example) a stylesheet may be set to display either the text in its original form or after the application of editorial interventions of particular kinds.

This provides us with the very attractive opportunity to both present the textual artifact as it has been historically preserved and to register editorial interventions where we deem them desirable. Our interventions remain minimal, restricted (at this point) to apparent errors.

The (inter-linear) writing of *šhdy* without the final *aleph* in l. 23 in the formulaic phrase *šhdyʾ bgw* "the witnesses herein are" (*BMAP* 2.14-15, 4.23, 5.16, 7.43, 8. 23, 10.18, 11.13; cf. 1.10) is a case in point:

```
(1)    <choice>
          <sic>
            <w type="noun" subtype="mp-cstr" lemma="šhd">šhdy</w>
          </sic>
          <corr>
            <w type="noun" subtype="mp-det" lemma="šhd">šhdyʾ</w>
          </corr>
       </choice>
```

This makes good editorial (and historical) sense. It also means that when we add the morphosyntactic markup we are not left only with the erroneous analysis—in this case, as if the

noun were actually a plural construct. This leads to a general observation, namely, that because of the digital environment the various aspects of our critical edition (e.g., facsimile, transcription, translation) and archival resource (e.g., morphosyntactic analysis) need not fully overlap or seamlessly agree. The various components can be manipulated to multiple (and even conflicting) ends.

The transcription will be offered in two scripts: a transliterated roman script and an Aramaic block script. All the XML markup was composed initially using the roman script. The markup for the transcription in the Aramaic block script (and also for the digital facsimile itself) will be generated from this same file. After some experimentation and following MEP recommendations (Chesnutt, Hockey and McQueen 1999), we follow a gradual markup procedure, attending to one variety (or level) of markup at a time (e.g., editorial transcription, morphosyntactic analysis). The basic elements in our markup scheme are three: the alpha-numeric characters themselves and indications for line and word division:

```
(2)  <ab type="line" n="BMAP.3.R:01">
        <w type="prep" lemma="b">b</w>
        <num type="cipher" value="7">3 3 1</num>
        <pc> </pc>
        <w type="prep" lemma="l">l</w>
        <name type="month">ʾlwl</name>
        <pc> </pc>
        <w type="pron" subtype="_3ms" lemma="hw">hw</w>
```

These correspond to the characters of the Aramaic script and numerical notation system and the two meta-script conventions (line division, word spacing) habitually employed by the scribe (Haggai b. Shemaiah):

(3) BMAP 3.1-3

The writing consists of alphabetic characters and numerical ciphers grouped by word spacing and organized into horizontal lines. We employ the standard transliteration conventions formulated by the SBL (Alexander 1999, §5.1.1.1) for representing the Aramaic script, a linear alphabetic script:

(4)   ʾbgdhwzḥṭyklmnʿspršt

In the Aramaic numeral notation system, for numbers up to ninety-nine, the system is purely cumulative-additive, consisting of signs for 20, 10, and 1. The unit-signs are grouped in threes (since up to nine such signs could be required). We use the corresponding Arabic numerals for the three component signs (20, 10, and 1), plus 2 and 3, depending on how the units are grouped. For example, the number twenty-eight is written out in l. 1 with a cipher composed of the following signs:

(5) BMAP 3.1
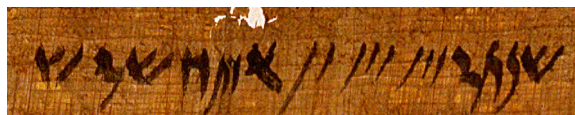
20 3 3 3 2

<num type="cipher" value="28">20 3 3 2</num>

The lines are right-adjusted and line-ends always coincide with graphic word boundaries. These lines are presentational in nature only, and thus bear no semantic significance for what is written. That is, the text is written in a running format with lines ending where they may, constrained only by the width of the sheet of papyrus being used and coincidence of word boundary. We use the "anonymous block" element (<ab></ab>) in TEI with the @type ("line")

and @n (e.g., "BMAP.3.R:01") attributes to signify these lines:

(6) <ab type="line" n="BMAP.3.R:01"> ... </ab>
   <ab type="line" n="BMAP.3.R:02"> ... </ab>
   <ab type="line" n="BMAP.3.R:03"> ... </ab>

As generally in alphabetic writing from the ancient Levant, the Aramaic of the Elephantine papyri is written out with word division. Spacing (a brief segment of the papyrus left uninscribed) is used in these papyri to signify (graphic) word division, a convention of Aramaic scribal practice that becomes prominent in the seventh century BCE (e.g., *KAI* 233 = Assur Ostracon; *TAD* A1.1). We use whitespace wrapped in the "punctuation character" element (<pc> </pc>) to represent this meta-script convention. This turns out to have a number of benefits. First, it underscores the fact that such a use of spacing is a material, meta-script convention (e.g., like the use of commas). The <pc> element according to the TEI guidelines "contains a character or string of characters regarded as constituting a single punctuation mark" (17.1.2). In this instance, spacing is used just like the point or dot in the old Hebrew script or the small cuneiform wedge in the alphabetic cuneiform from ancient Ugarit, and stands in contrast to the *continua scripta* tradition of alphabetic writing without word dividers, as in some Phoenician scripts and in ancient Greek manuscripts. Second, it allows a more perspicuous linguistic description in the coding since a graphic word does not necessarily correspond to a linguistic (or grammatical) word. For example, prepositional phrases with the proclitic prepositions *b-*, *l-*, and *k-* are written out graphically together with their objects, e.g., *lʿnnyh* (*l* + *ʿnnyh*) "to Ananiah" (3.3), *bʾbny* (*b* + *ʾbny*) "in the stone weights" (3.6). Thus, the "word" (<w> </w>) element may be reserved for representing a "grammatical (not necessarily orthographic) word" (17.1.1). This use of the <pc> element also allows the use of whitespace within the XML markup for human readability—that is, only whitespace wrapped in a <pc> element indicates a character from the text, while all other whitespace is insignificant.

Alterations observed in the source text are mainly of two kinds, additions and deletions, for which the "addition" (<add></add>) and "deletion" (<del></del>) elements from the core

TEI Guidelines (3.4.3; cf. 1.3.1.4) are used. With the <add> element, the @hand (e.g., "scribe," "witness 1") and @place (e.g., "above," "inline") attributes are used, and with the <del> element, the @type (e.g., "erasure") attribute. Examples:

(7)   `<add hand="scribe" place="above"><w>ʾlhʾ</w></add>` (3.10)

              (the scribe added the word *ʾlhʾ* "the god" above the line)

(8)   `<num type="cipher" value="4"><w>3 1<del type="erasure"> 1</del></w></num>` (3.6)

              (the scribe originally wrote the cipher for the number "5"; by erasing the last vertical stroke, he corrected the number to "4")

Additions, which (in this document) are generally inclusions of material accidentally left out initially and written in inter-linearly above the line, are marked approximately at the point in the text where they are inserted (often above spacing between words; cf. Meulen and Tanselle 1999, p. 205). Deletions, which are mostly erasures, are marked at the inter-linear point at which they occur. The markup aims only to report the fact of alteration. All additional editorial comment, including specification of chronological sequence, is reserved for the epigraphic commentary. The presence of the accompanying facsimile offers a helpful clarifying aid, relieving the markup of the need to be overly precise as to point of execution.

      On occasion it is apparent that a deletion and addition have been coordinated. For example, at the beginning of 3.3 the scribe initially wrote *kl nšn 2*, "altogether, 2 ladies." Upon recognizing his mistake (the sellers are husband and wife) he erased the final vertical stroke in the cipher for the numeral "2" (converting it to the cipher for the numeral "1") and added *gbr 1* "1 man" inter-linearly above the line following the word *kl* "all."

(9)   `<w type="noun" subtype="ms-abs" lemma="kl">kl</w>`

              `<pc> </pc>`

              `<add hand="scribe" place="above">`

```
<w type="noun" subtype="ms-abs" lemma="gbr">gbr</w>
<pc> </pc>
<num type="cipher" value="1">1</num>
</add>
<w type="noun" subtype="fp-abs" lemma="ʾnth">nšn</w>
<pc> </pc>
<num type="cipher" value="1">1
    <del type="erasure"> 1</del>
</num>
<pc> </pc>
```

There is no way of knowing the precise sequence in which these alterations were executed (i.e., addition then erasure, erasure then addition), but it is at least clear that they are interdependent. In such cases, the TEI Guidelines allow for the use of the "substitution" (<subst></subst>) element (11.3.3.1.5) to group coordinated alterations. However, since in some cases (as here) the alterations are not proximate, we forego the use of this element, marking only the fact of an addition and deletion and relying on the epigraphic commentary to detail a more precise characterization of the coordination (or of other relevant matters).

There are places where a material reading is unclear for some reason. For example, in 3.2 the wife's name was originally written as *wbl*. Later the scribe adds an additional letter super-linearly, above and to the left of the *bet*:

(10)
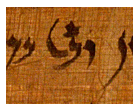
Kraeling construes the letter as an *aleph* and reads *ʾwbl* (*BMAP*, pp. 158-59). In contrast, Porten and Yardeni construe the letter as a *yod* and read *wbyl* (*TAD* B, p. 64). The name is apparently Hurrian (Kornfeld 1978: 113) and is spelled three other ways: *ʾwbl* (3.10), *ybl* (3.25), and *ʾwbyl*

(*BMAP* 4.3). Graphically, the inserted letter patterns more like a *yod* than an *aleph*, especially in over all size, and its placement (above the line after the *bet*) is also consistent with the *yod* in *ʾwbyl* (*BMAP* 4.3). If the scribe intended an *aleph* as the initial consonant in the name, presumably he would have inserted the letter above the line and closer to the beginning of the name, for which there is plenty of space (most of the super-linear additions in this document are inserted beginning approximately at the point in the text where they would fall most naturally, e.g., *gbr 1* in 3.3; *ʾlhʾ* in 3.10; *zk* in 3.12).
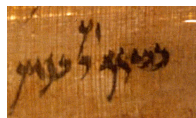
(11) *gbr 1* in 3.3
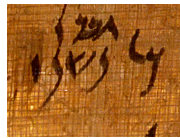
(12) *ʾlhʾ* in 3.10
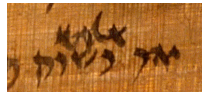
(13) *zk* in 3.12

In such cases, the TEI Guidelines provide for multiple ways of marking such variation. We have opted to use the "apparatus entry" (<app></app>) element, which may be used "whether or not represented by a critical apparatus in the source text," with the parallel segmentation method for coding variant readings (12.2.3), because it provides maximum transparency and flexibility. When there is an editorial preference for a reading (as here) we mark that with the "lemma" (<lem></lem>) element (with the @resp attribute signaling any supporting opinions, e.g., "TAD"). Other readings are marked with the "reading" (<rdg></rdg>) element and the @resp attribute (e.g., "K"). So our markup for this example is as follows:

```
(14)    <name type="person">wb
            <add hand="scribe" place="above">
                <app>
                    <lem resp="TAD">y</lem>
                    <rdg resp="K">ʾ</rdg>
                </app>
            </add>
        </name>
```

If the alternative readings are judged to be equally preferable each is marked with the <rdg> element (with @resp attribute), as (possibly)[v] in the following example (3.24):

(15) BMAP 3.14

```
        <name type="person">
            <app>
                <rdg resp="K, TAD">ḥyḥ</rdg>
                <rdg resp="TAD">ḥyrw</rdg>
            </app>
        </name>
```

Again, the main intent is reportorial in nature. Any supporting rationale will be given in the epigraphic commentary.

Morphosyntactic Markup

Classification of words into parts of speech (or word classes) is not entirely straightforward. The modern linguistic practice has been to use morphological and syntactic criteria for defining parts of speech, which of course vary cross-linguistically and do not necessarily totally overlap even within languages. The appendix contains our working POS classification. The intention here is

not to innovate linguistically. We have attempted to use an intuitive, field-specific sense of the relevant grammatical and lexical categories in use. In general, our default analysis is cued principally by the treatments in standard grammars (e.g., *GEA* (Muraoka and Porten 1998)) and lexicons (e.g., *CAL* (Kaufman et al., n.d.), *DNWSI*) of the various Aramaic dialects. The chief aims in providing such tagging is to ease usability of these documents and to support the linguistic analysis of their language. Furthermore, wanting all markup to be well-formed XML, and thus enabling general portability and use of standard XML parsers for processing and the like, we have utilized only five general TEI elements:

(16) <name> (name, proper noun) contains a proper noun or noun phrase

<num> (number) contains a number, written in any form

<w> (word) represents a grammatical (not necessarily orthographic) word

<m> (morpheme) represents a grammatical morpheme

<abbr> (abbreviation) contains an abbreviation of any sort

Four of the five elements—<name>, <num>, <m>, and <abbr>—are used fairly restrictively. The <w> element, in contrast, does the bulk of the descriptive work.[vi]

Several observations about the markup itself. First, the <name> and <num> elements provided by TEI accord well with the fact that proper names and numbers are generally distinguished lexicographically in Semitic (and Aramaic in particular) from other word categories. Numerals are mostly indicated through ciphers in this document, which we indicate with the @type (="cipher") and @value (e.g., "20") attributes. When a number is spelled out, as in 3.16 (*lʿšrtʾ* "to the ten"), we indicate what kind of number with the @type attribute (e.g., "cardinal") and then wrap it within the <w> element:

(17) <num type="cardinal" value="10">

<w type="noun" subtype="fs-det" lemma="ʿšrh">ʿšrt</w>

</num>

We do something similar with the <abbr> element. Since we are not interested necessarily in expanding abbreviations within the transcription but pointing to a lexical entry, we identify an abbreviation with the <abbr> element and then wrap in the <w> element:

(18)   <abbr>
        <w type="abbr" lemma="r">r</w>
    </abbr> (3.6)

The <w> element, the workhorse of this markup scheme, may appear with as many as three attributes: the @type attribute identifies the relevant part of speech; the @subtype attribute provides pertinent inflectional information (e.g., for nouns: gender, number and state; for verbs: binyan, TAM, person, gender, and number); and the @lemma attribute points to the citation form in a lexicon (module). In the case of homographs, we have followed the ordering found in *CAL*. As a rule, the attributes are used only as relevant (e.g., prepositions, conjunctions and the like require no @subtype entry) and only to the extent relevant (e.g., the @subtype description of nouns with possessive suffixes are marked only for gender and number).[vii] Initially, we have erred on the side of providing more descriptive POS categories, especially when it comes to the various kinds of particles, conjunctions, adverbs, and the like that are used.

We treat clitics differently, depending on their kind. Clitics are (phonologically) bound forms ("constrained to occurring next to an autonomous word" (Hopper and Traugott 1993, p. 5)) that have an independent syntactic role and thus may be thought of as standing halfway between autonomous words and fully grammaticalized affixes. Prepositions and pronouns are two word categories that often become cliticized in natural languages. We have marked the proclitic prepositions (*b-*, *l-*, *k-*) and conjunctive *waw* (*w-*) with the <w> element:

(19)   <w type="prep" lemma="b">b</w> (3.1)
        <w type="prep" lemma="l">l</w> (3.1)
        <w type="prep" lemma="k">k</w> (3.23)
        <w type="conj" lemma="w">w</w> (3.23)

This is consistent with the general treatment these elements receive from lexicographers, who habitually provide lexical entries for them in the standard lexicons. By contrast, the pronominal suffixes attached to verbs, nouns, and prepositions have been marked with the <m> element:

(20)   <w type="prep" lemma="b">b<m type="sf-3ms">h</m></w> (3.22)

(21)   <w type="noun" subtype="ms" lemma="ksp">ksp
    <m type="sf-2ms">k</m>
  </w> (3.22)

(22)   <w type="verb" subtype="pa-impf-3ms" lemma="gry">ygrn
    <m type="sf-2ms">k</m>
  </w> (3.19)

The logic here is twofold: one, pronominal suffixes are not treated separately lexicographically in Aramaic (and in West Semitic generally), and, two, they are not considered a part of the standard inflectional feature set for verbs and nouns.[viii] Marking them with the <m> element (instead of with the <w> element) signals both of these distinctions and captures as well these clitics' strong resemblance to other affixes (e.g., suffixes on the Perfect), their lexical status notwithstanding.[ix]

Another area where we default to the lexicographers (at least initially) is in our treatment of compound or pseudo prepositions (*GEA*, 87). For example, following *CAL* (and *GEA*) we consider both *bšm* (l. 13) and *br mn* (l. 21) as fully grammaticalized and thus autonomous lexical items and mark them as such:

(23)   <w type="prep" lemma="bšm">bšm</w>(3.13)
    <w type="prep" lemma="br mn">br<pc> </pc>mn</w> (3.21)

Alternative markup privileging the decomposition of these complex items are readily imaginable and perhaps could be handled alternatively using the <choice> element:

```
(24)    <choice>
            <seg type="pos">
              <w type="prep" lemma="bšm">bšm</w>
            </seg>
            <seg type="pos">
              <w type="prep" lemma="b">b</w
              <w type="noun" subtype="ms-cstr" lemma="šm">šm</w>
            </seg>
        </choice>


(25)    <choice>
            <seg type="pos">
              <w type="prep" lemma="br mn">br<pc> </pc>mn</w>
            </seg>
            <seg type="pos">
              <w type="noun" subtype="ms-cstr" lemma="br2">br</w>
            <pc> </pc>
            <w type="prep" lemma="mn">mn</w>
            </seg>
        </choice>
```

We have used the traditional nomenclature for the various verbal *binyanim* (e.g. Peal, Pael, Afel) in the markup but will write a program that allows users to shift back and forth between this set of terms and the newer, cross-Semitic terms (e..g., G, D, C, as used in *CAL*).

Conclusion

In closing, nothing about what we have just reviewed in terms of XML markup seems to us to be revolutionary, either technically or theoretically. The surprise remains the general absence of a scholarly discussion on such issues in the field. In part we suspect this is because most of the digital-based text projects in the field to date have been dominantly entrepreneurial in motivation and orientation and not conceived as research or scholarship. There are exceptions. For example, there seems to be a live interest currently in leveraging digital resources for syntactic analysis of various (Semitic) text corpora, and there are now a number of sites dedicated to presenting transcriptions of cuneiform literature (e.g., *Sources for Early Akkadian Literature*, http://www.seal.uni-leipzig.de/). But to our knowledge no digital-based project involving texts from the ancient Near East (esp. pre-Hellenistic corpora) have been conceived of from an explicitly articulated editorial perspective.[x] That is, most of the commonly used electronic text resources in the field (e.g., Accordance, Logos, Michigan-Claremont-Westminster Electronic Hebrew Bible) are essentially what is known as "reader editions." They are not critical or scholarly editions and therefore, ultimately, cannot be depended on academically. These "reader" editions have served the field well, showing, for example, the viability and benefit of electronic text-based resources and "tools." Now the field needs to take the next step: to create critical, scholarly editions that will make use of all of the advantages of the currently available electronic reader editions and also be trustworthy and reliable. This is what we are proposing to do with *DBMAP*.

Notes

i   This represents a slightly revised version of a paper presented in the Digital Humanities in Biblical, Early Jewish, and Christian Studies unit at the Annual Meeting of the Society of Biblical Literature in San Diego, CA (November 23, 2014). The images used in examples 3, 5, and 10-13 are details of BMAP 3 (47.218.95; =TAD B3.4). InscriptiFact Text ISF_TXT_00055. Photograph by Bruce and Kenneth Zuckerman, West Semitic Research. Courtesy Brooklyn Museum. Reuse of these images is prohibited without permission of the rights-holders. We thank Bruce Zuckerman and Marilyn Lundberg of West Semitic Research and Ed Bleiberg of the Brooklyn Museum for their support of this project more generally.

ii   In what follows, we employ inline markup. As one reviewer of this paper has pointed out, however, other methods of markup, such as stand-off markup (see, for example: http://www.tei-c.org/Activities/Workgroups/SO/sow06.xml; http://www.balisage.net/Proceedings/vol5/html/Banski01/BalisageVol5-Banski01.html), may actually end up being more congenial to our project. We find this an incredibly generative observation and plan to explore further such possibilities as the project moves forward.

iii   Kline and Perdue (2008, p. 147): "Increasingly common are print editions in which a photo facsimile appears as part of a parallel text accompanying a printed editorial transcription. Digital scanning creates wider options for editors who wish to offer such photographic images in online or DVD-based editions, conveniently linked to machine-searchable transcriptions, accessed through automated indexes."

iv   Both *BMAP* and *TAD* (unwittingly) offer approximations of a "typographic facsimile," although neither is consistent on this issue since these volumes are not expressly theorized from an editorial perspective.

v   This is by way of example only and follows the judgment of Porten and Yardeni (*TAD* B, 64)—we have not looked closely at personal names to this point.

vi   Here we emphasize the practical and limited nature of our initial experiment. There are other standards that are both compatible with TEI and promote the use and reuse of textual data across applications, e.g., LAF (Linguistic Annotation Framework, ISO 2012; see Ide and Romary 2004, p. 211-225).

vii   Historically, possessive suffixes were attached to nouns after the case endings in Aramaic, and syntactically, nouns with possessive suffixes are considered determined. Some synchronic grammars of specific Aramaic dialects (e.g., *GEA*, 46; Hug 1993, p. 56) indicate that the suffixes are attached to the construct forms of nouns. Whether this is the right analysis is open to debate, but even if correct, for our purposes, the presence of a possessive suffix implicates the use of the construct state of the noun, and therefore need not be explicitly marked (cf. Bar-Haim, Sima'an, and Winter 1998, p. 7).

viii   Contrast the suffixes on the Perfect form of the verb, which are clearly related historically to the larger pronominal system, with the chief difference that over time they became fully grammaticalized as suffixes, and thus a part of the verb's inflectional morphology.

ix   Contrast Bar-Haim, Sima'an, and Winter (1998, p. 7, 28), who treat pronominal suffixes on verbs and prepositions in Modern Hebrew as word segments, but not possessive suffixes on nouns.

x   For example, SEAL offers this as its main rationale: "to enable the efficient study of the entire early Akkadian literature in all its philological, literary, and historical aspects." The site boasts of new "collations" for the texts presented, but offers no explicit editorial theory for guidance. Presumably this is to be elaborated in the print volumes under production.

Appendix:

Part of Speech (POS) Inventory

<name> (name, proper noun) contains a proper noun or
 	
    noun phrase
    @type="person" "divine" "place" "gentilic"
	@type="pos(sessive)"
	@lemma="(dictionary entry)"
	
	@type="indef(inite)"
	@lemma="(dictionary entry)"

<num> (number) contains a number, written in any
 	
    form
    @type="cipher" "cardinal" "ordinal"
    "fraction" "multiplicative"
	@type="prep(osition)"
	@lemma="(dictionary entry)"
	
	@type="conj(unction)"
	@lemma="(dictionary entry)"

<abbr> (abbreviation) contains an abbreviation of any sort	@type="neg(ative)" @lemma="(dictionary entry)" @type="cond(itional)" @lemma="(dictionary entry)"

<m> (morpheme) represents a grammatical morpheme
 	
    *mainly used (now) for representing object
    and possessive suffixes
    @type="sf-(person, gender, number)"
	@type="inter(rogative)"
	@lemma="(dictionary entry)"
	
	@type="adverb"
	@lemma="(dictionary entry)"

<w> (word) represents a grammatical (not necessarily	@type="interj(ection)"

orthographic) word

@type="verb"

@subtype="(binyan: pe, pa, af/haf, ethpe,

ethpa, ettaf)-(TAM: pf, impf, impv, inf,

part)-(person, gender, number)"

@lemma="(dictionary entry)"

@type="noun"

@subtype="(gender, number)-(state: abs, cstr,

det)"

@lemma="(dictionary entry)"

@type="adj(ective)"

@subtype="(gender, number)-(state: abs, cstr,

det)"

@lemma="(dictionary entry)"

@type="pron(oun)"

@subtype="_(person, gender, number)"

@lemma="(dictionary entry)"

@lemma="(dictionary entry)"

@type="exist(ence)"

@lemma="(dictionary entry)"

@type="part(icle)"

@lemma="(dictionary entry)"

@type="abbr(eviation"

@lemma="(dictionary entry)"

Bibliography

Alexander, P. H., ed., 1999. *The SBL Handbook of Style: For Ancient Near Eastern, Biblical, and Early Christian Studies*. Peabody: Hendrickson.

Bar-Haim, R., K. Sima'An, and Y. Winter, 2008. Part-of-Speech Tagging of Modern Hebrew Text. *Natural Language Engineering* 14(2), pp. 223-251.

Chesnutt, D. R., S. M. Hockey, and C. M. Sperberg-McQueen, 1999. Markup Guidelines for Documentary Editions. 4 July. [online] Available at: http://xml.coverpages.org/MepGuide199909.html [Accessed 19 December 2014]

Donner, H. and W. Rölling, 1966-1969 and 2001. *Kanaanäische une aramäische Inschriften*. 2d ed and 5th ed. 3 vols. Wiesbaden: Harrasowitz. (=*KAI*)

Hoftijzer, J. and K. Jongeling, 1995. *The Dictionary of North-West Semitic Inscriptions*. 2 vols. Leiden: Brill. (=*DNWSI*)

Hopper, P. J., and E. C. Traugott, 1993. *Grammaticalization*. Cambridge: Cambridge University Press.

Hug, V., 1993. *Altaramäische Grammatik der Texte des 7. und 6. Jh.s v. Chr.* Heidelberg: Heidelberger Orientverlag.

Ide, N., and L. Romary, 2004. International Standard for a Linguistic Annotation Framework. *Natural Language Engineering* 10 (3-4), pp. 211-225.

ISO, 2012. Language Resource Management – Linguistic Annotation Framework. ISO 24612:2012. Edition 1. [online] Available at:

http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=37326
[Accessed 19 December 2014] (= LAF)

Kaufman, S., et al. n.d. *Comprehensive Aramaic Lexicon.* Cincinnati: Hebrew Union College.
[online] Available at: http://cal1.cn.huc.edu/. [Accessed 19 December 2014] (= *CAL*)

Kline, M.-J., and S. H. Perdue. 2008. *A Guide to Documentary Editing*. 3d ed. Charlottesville:
University of Virginia. [online] Available at: http://gde.upress.virginia.edu/ [Accessed 19
December 2014]

Kraeling, E. G., 1953. *The Brooklyn Museum Aramaic Papyri: New Documents of the Fifth
Century B.C. from the Jewish Colony at Elephantine*. New Haven: Yale University Press.
(= *BMAP*)

Meulen, D. L. V., and G. T. Tanselle, 1999. A System of Manuscript Transcription. *Studies in
Bibliography* 52, pp. 201-212.

MLA's Committee on Scholarly Editions, 2001. Guidelines for Editors of Scholarly Editions.
MLA.org. [online] Available at: http://www.mla.org/cse_guidelines [Accessed 19
December 2014] (= CSE)

Muraoka, T., and B. Porten, 1998. *A Grammar of Egyptian Aramaic*. Leiden: Brill (= *GEA*)

Porten, B., and A. Yardeni., 1986-99. *Textbook of Aramaic Documents from Ancient Egypt*. 4
vols. Winona Lake: Eisenbrauns. (=*TAD*)

Text Encoding Initiative, n.d. *P5: Guidelines for Electronic Text Encoding and Interchange.*
[online] Available at: http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html
[Accessed 19 December 2014] (= TEI)