Hebrew Lexicon Schema

by David Troidl

The Hebrew Lexicon schema is an interim format for entering Hebrew lexical information. It is intended to be easy to use, easy to understand and easy to transform.

The document is contained in a `<lexicon>` element.

```
<?xml version="1.0" encoding="utf-8"?>
<lexicon xsi:schemaLocation="http://www.APTBibleTools.com/namespace HebrewLexicon.xsd"
        xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
        xmlns="http://www.APTBibleTools.com/namespace">

</lexicon>
```

The namespace `"http://www.APTBibleTools.com/namespace"` is not registered anywhere.

Entry

Within the `<lexicon>` element is a sequence of `<entry>` elements. For example:

```
<entry id="H1254a">
    <w pos="v" pron="baw-raw'" xlit="bârâ'" xml:lang="heb">בָּרָא</w>
    <source>a primitive root;</source>
    <meaning>(absolutely) to <def>create</def>; (qualified) to <def>cut down</def> (a
        wood), <def>select</def>, <def>feed</def> (as formative processes)</meaning>
    <usage>choose, create (creator), cut down, dispatch, do, make (fat).</usage>
    <bdb type="root" cite="full" mod="I"><w>בָּרָא</w> <def>shape</def>, <def>create</def>
        <sense><stem>Qal</stem> <asp>Pf</asp>.—<def>shape</def>, <def>fashion</def>,
            <def>create</def></sense>
        <sense><stem>Niph</stem>.—<em>Pass</em>.
            <sense n="1"><def>be created</def></sense>
            <sense n="2">with reference to birth</sense>
            <sense n="3">of something new, astonishing</sense>
        </sense>
        <sense><stem>Pi</stem>.
            <sense n="1"><def>cut down</def></sense>
            <sense n="2"><def>cut out</def></sense>
        </sense>
    </bdb>
    <source><w src="H1254a">ברא</w></source>
    <xref twot="278"/>
    <status p="135">base</status>
</entry>
```

The `<entry>` element has one required attribute: `id`. The `id` is an *augmented Strong number*, where entries in Strong's Hebrew Dictionary are subdivided to be compatible with the Brown, Driver, Briggs *Hebrew Lexicon* and the *Theological Wordbook of the Old Testament*.

Block Elements

The direct children of the `<entry>` element are a choice of block-level elements. They generally appear in the following order, but the order or appearance of any particular element is not required.

W

The `<w>` element appears in several different contexts, normally with attributes characteristic to that context. As the main word entry,

```
<w pos="v" pron="baw-raw'" xlit="bârâ'" xml:lang="heb">בָּרָא</w>
```
it will have the following attributes:

- `pos` – the part of speech.

- `pron` – the pronunciation, as given in Strong's Hebrew Dictionary.

- `xlit` – the transliteration, also as given in Strong's.

- `xml:lang` – the language: `heb`, `arc` or `x-pn` (for proper names).

This will be the only `<w>` appearing as a block level element.

Note

A `<note>` will appear when corrections are required for the data contained in the printed Strong's Hebrew Dictionary. The content of `<note>` is text, but will generally indicate the field the note is intended to correct. Notes can also appear inline.

Source

The first `<source>` element contains the derivation information contained in Strong's. For example: `a primitive root;` The source can contain `<w>`, `<note>` and `<def>` elements, as necessary.

The second `<source>` element, appearing after the `<bdb>` element, contains information about the root of the element, as it appears in BDB.

Meaning

The `<meaning>` element contains the Strong definition of the term. Meaning can contain `<w>`, `<note>` and `<def>` elements, as necessary.

Usage

The `<usage>` element contains the forms the word takes in the Authorized Version, as listed in

Strong's. Usage can contain `<w>`, `<note>` and `<def>` elements, as necessary.

BDB

The `<bdb>` element is a general container for the Brown, Driver, Briggs data in the lexicon. There are four attributes, representing features of the BDB text:

- `type` – distinguishes an `entry` (the default) from a `root`.

- `cite` – distinguishes whether BDB has a `partial` (the default) or `full` list of Scripture references. A full list is indicated in the text by a dagger (†).

- `form` – indicates whether the lexical form appears in the Hebrew Bible. Default is `true`.

- `mod` – contains a Roman numeral, where BDB distinguishes entries that way.

The `<bdb>` element can contain `<w>`, `<pos>`, `<stem>`, `<asp>`, `<def>`, `<em>`, `<ref>`, `<foreign>` and `<sense>` elements in various combinations.

Sense

The `<sense>` element is the only block-level element that does not appear as a direct child of the `<entry>`. `<sense>` elements appear as children of the `<bdb>` or other `<sense>` elements, to any degree of nesting required. `<sense>` elements have a single, optional attribute: `n`, to distinguish the various `<sense>` elements. Senses, as listed in BDB, can take numbers, letters, Roman numerals, or even Greek letters in parentheses. Often on verb entries, senses are distinguished, instead, by `<stem>` elements that begin the sense content. (It has been suggested that these `<sense>` elements also be given an attribute, for easier processing. In that case, I would suggest using the MorphHB Parsing Scheme codes, at http://beta.openscriptures.org/projects/project/morphhb/wiki/ParsingScheme/)

The `<sense>` element can contain choices of `<pos>`, `<stem>`, `<asp>`, `<def>`, `<em>`, `<w>`, `<ref>`, `<foreign>` and other `<sense>` elements, in any order. In the document, `<sense>` elements are generally indented to indicate their level of nesting.

Xref

The `<xref>` element is a general container for cross reference information. Presently, it has only the `twot` attribute, to contain the corresponding entry number in the *Theological Wordbook of the Old Testament*.

Status

The `<status>` element has a single attribute: `p`, indicating the page number of the BDB entry in my

source text. Individual pages can be accessed as (for page 1) http://en.wikisource.org/wiki/Page:A_Hebrew_and_English_Lexicon_%28Brown-Driver-Briggs%29.djvu/25, where the last number should be the page number plus 24. The content of the element is the current status of that entry in the development of the lexicon. Status values currently consist of:

- `stable`: entries resulting from the automated construction process, with no editing.

- `base`: entries that have basic definitions from BDB.

- `ref`: entries that have scripture references added, but remain incomplete.

- `done`: entries that fully represent the BDB data.

- `new`: root entries from BDB, that have no corresponding Strong entry.

- `added`: root entries that had to be added to accommodate internal BDB information.

Inline Elements

W

The `<w>` element can also appear as an inline element, in several circumstances:

- As an entry word, within the `<bdb>` element, `<w>` will appear without attributes.

- As a cross reference to another lexicon entry, `<w>` will appear with the `src` attribute, an IDREF. Often here the only content will be the entry ID it refers to.

- As a reference to a word that does not appear in the lexicon, it will have the `pron` and `xlit` attributes, and the content will be the Hebrew word.

- As a root listing, it will have the `src` attribute, and contain the consonantal form of the root.

- As a holdover from the original BDB Outline, `<w>` elements can appear with an `origin` attribute. These are features of entries that have not been updated to at least `base` status, and will be clarified either by inclusion under one of the other uses, or elimination.

POS

The `<pos>` element appears only in `<bdb>` or `<sense>` elements, and holds the part of speech. Often this is omitted in the base entries, when it corresponds to the part of speech attribute on the main `<w>` element.

Stem

The `<stem>` element appears in the `<bdb>` and `<sense>` elements, to distinguish definitions or senses by the stem of the Hebrew word, with values such as `Qal`, or an abbreviation of one of the other stems.

Asp

The <asp> element appears in the `<bdb>` and `<sense>` elements, to distinguish definitions or senses by the aspect of the Hebrew word, with values such as `Pf`, `Impf`, or an abbreviation of one of the other aspects.

Def

The `<def>` element is one of the most frequently occurring elements. It is used to set off the various definitions of the word, whether in the `<source>`, `<meaning>` and `<usage>` elements, for the Strong information, or in the `<bdb>` and `<sense>` elements for BDB.

Em

The `<em>` element is simply for emphasis that does not fall under one of the other elements.

Ref

The `<ref>` element contains a Scripture reference. It has a single attribute `r` that holds the reference, in standard SBL form, with the content remaining similar to BDB, using the colon, rather than superscripts. Multiple Scripture references are separated by spaces, as in BDB.

Foreign

The `<foreign>` element contains foreign words and phrases, mainly in Greek, Arabic, Aramaic, Syriac and Ge'ez, in their own script, but with an array of other languages, generally transliterated. Mostly, at least initially, it will serve as a placeholder for missing content, until someone proficient in those languages can fill in the content. The `xml:lang` attribute contains the standard ISO 639 codes. See http://www.sil.org/iso639-3/codes.asp.

The following table contains the languages included in the Hebrew Lexicon so far:

BDB	Language	ISO
Ar.	Arabic	ara
Aram.	Aramaic	arc
As.	Assyrian	akk
Eth.	Ethiopic	gez
𝕲	Greek	grc
Heb.	Hebrew	heb
𝖁	Latin	lat
Pers.	Persian	fas
Skr.	Sanskrit	san
Syr.	Syriac	syr