Electronic Edition of the Midrash Pirqe Rabbi Eliezer: Creating an Encoding Manual

Lewis M. Barth

Hebrew Union College, 3077 University Avenue, Los Angeles, CA 90007, USA.

KEYWORDS: Midrash, Manual, Hebrew

AFFILIATION: Hebrew Union College, L.A.

E-MAIL: lbarth@mizar.usc.edu
FAX NUMBER: (213) 749-1192
PHONE NUMBER: (213) 749-3424

I. Introduction

This paper will deal with structural and encoding issues encountered in the process of creating a manual for encoding an electronic edition of Pirqe Rabbi Eliezer (Pirqe R. El.), "the Chapters of Rabbi Eliezer". This work is a midrashic retelling of significant aspects of the biblical narrative, from the creation story through the Book of Esther. It was written in the Land of Israel probably during the eighth century CE, i.e., in the early Muslim period. The language of Pirqe R. El. is Hebrew with a few non-Hebrew loan words in transliteration.

Pirqe R. El. was exceedingly popular in medieval and pre-modern traditionalist Jewish literary circles. It is preserved in more than twenty complete manuscripts containing fifty-two to fifty-four chapters and more than seventy-five partial manuscripts and fragments. In addition, over thirty printed editions of Pirqe R. El. have appeared since the sixteenth century. Recently, scholarly interest in Pirqe R. El. has focused on literary, historical and interpretative issues.[1]

There is no scholarly edition of this text in the modern sense of this term. An electronic text of Pirqe R. El. exists and is commercially available.[2] However, the present e-text is a semi-critical eclectic edition, based on three manuscripts whose relationship has not been fully determined; markup is limited to identification of citations from biblical or rabbinic literature.

The initial goal of this project was to create a critical edition of Pirqe R. El. The goal has now expanded to include electronic publication of all Pirqe R. El. manuscripts and fragments in two forms: digital facsimiles and transcriptions with hypertext links. There are two reasons for this: 1) the mass of textual material and 2) recent hypotheses regarding the development of medieval Hebrew manuscripts which argue that each manuscript of a work is a completely new literary creation.[3] Thus the need to present visually a representation of each manuscript (at least the major ones), with the possibility of comparing complete readings of specific passages on the fly – something possible only electronically.

This paper will elaborate on some matters raised in the Introduction and concentrate on technical areas necessary for preparing an Encoding Manual for the project.

II. SGML/TEI element and ID attribute designations for divisions of the text

The first issue concerns the question of how the text of Pirqe R. El. should be divided.[4] Such a decision is related to the choice of representing either the units of meaning (Chapters, Paragraphs) or the physical makeup of a text (Pages and Lines). Units of meaning might be designated by the elements DIV (division) in which the attribute TYPE would indicate "chapter," ID the specific chapter number, and P (paragraph) in which the ID attribute would contain the specific paragraph number. Alternately, the physical make up of the text could be represented by the elements DIV in which the attribute TYPE would indicate "folio" and ID the specific folio number, and L (line) in which the ID attribute would contain the specific line number.

Two problems emerge regarding text division, both having to do with SGML/TEI limitations and neither unique to this project. First, it is not presently possible to do concurrent markup, that is, to simultaneously tag material units of meaning and physical layout. Second, the TEI L tag is reserved for a line in poetry, not a physical line of a prose manuscript, i.e., it encloses a unit of meaning which is contained in a physical line even when the meaning may run on to the next line.

The way around this is through the use of various MILESTONE elements: MILESTONE, PB (page break), and LB (line break) which contain attributes to indicate divisions in the text, but cannot contain text.

In regard to encoding manuscripts of Pirqe R. El., or any rabbinic text, after long evaluation, I have concluded that the basic initial encoding must be in units of meaning. Rabbinic units of meaning contain quotes – primarily from the Hebrew Bible – which may flow through two or occasionally three lines in a manuscript. In the present state of software development, it is not possible to place an opening QUOTE tag within a line enclosed by an L tag, and then place its closing QUOTE in another line. Thus one is forced to use units of meaning to divide the text and then insert MILESTONE tags to indicate page and line breaks for each separate manuscript.

One further comment regarding encoding Pirqe R. El. using units of meaning. The manuscripts and

printed editions of Pirqe R. El. divide the text by chapters, but do not contain paragraph divisions. Further, there is no agreed upon "canonical" reference system for Pirqe R. El.[5]

The only fully developed canonical reference system is found in the electronic text created by the "Academy of the Hebrew Language" for its "Historical Dictionary of the Hebrew Language". This electronic edition is based on one primary manuscript selected for its linguistic properties – New York, JTS Enelow 886 (Yemen, 1654), – corrected against four others. Consequently, it either contains material – therefore "paragraphs" – which are only found in manuscripts of the same family, or does not contain material – therefore "paragraphs" – which are found in manuscripts of a different family. Nevertheless, the AHL numbering will generally be used to establish a canonical reference system, though it may be revised as encoding of the different manuscripts proceeds.

III. Abbreviations

As far as I can determine, this is the first SGML/TEI editing project of a medieval Hebrew work. Consequently, the issue of abbreviations and references of all kinds in the electronic context needs to be addressed. Printed editions, and especially translations of Pirqe R. El. contain notes and index references to the Bible (Hebrew Bible, LXX or NT), Apocrypha, Pseudepigrapha, the Dead Sea Scrolls, Rabbinic Literature, and the Church Fathers. Numerous modern scholarly publications (books, journals, etc.) contain references to Pirqe R. El. as well.

Several questions and issues have emerged in regard to abbreviations and references.

First, the text is in Hebrew. Consequently, when a source is originally in Hebrew, should references contain Roman or Hebrew characters for titles of books?

Second, because of the differences between the electronic media printing, standard listings of abbreviations and references cannot always be used, or need to be modified. For example, references to rabbinic tractates in some abbreviation systems use scholarly transliteration, including superscript half circles to represent the Hebrew letters ALEF and AYIN.

Third, because the study of biblical texts is international, Western reference systems are often reflective of different cultural traditions and can even differ within the same language system. So, in English language countries verses from the biblical prophet Isaiah are often referenced in the following ways: Isa and Is (with or without a period). In German, this prophet's name is Jesaja, and referenced Jes. The tendency in recent scholarly abbreviation of scriptural and related titles is not to include a period after the book reference. Thus, Isa 1:5. But electronic search mechanisms can use the period as a delimiter, setting off parts of a reference.

Finally, even in so-called standard works, such as the Bible, differences exists in verse numbering (i.e., various editions of the HB, NT or LXX).[6] Thus, it becomes necessary to indicate the specific edition of the work in a bibliographical note.

How does one proceed without reinventing the wheel? By choosing existing standards and indicating where modification is necessary. The standard for this project will be that of the AAR/SBL requirements, found in SBL: Membership Directory and Handbook, 1994, pp. 224-240.[7] However, superscript for "ALEF" and "AYIN" as well as other diacritic marks for rabbinic texts are omitted. Where the reference contains two words, no space should be placed between the words; ex. "Ros Has" =<Rosh Hashanah> would appear "RosHas." If at all possible, each source reference should be composed of four parts, each part separated by a period; ex. "HB.Gen.20.15."[8] The first part represents the general body of literature (HB=Hebrew Bible), the second the specific text (Gen=Genesis), the third either the chapter (20=chapter 20) or the folio, the fourth either the verse (15=verse 15) or the column.

Reference examples

HB.Gen.20.15=Hebrew Bible, Genesis 20:15.
LXX.Gen.20.15=Septuagint, Genesis 20:15.
NT.Matt.5.6=(Greek) New Testament, Matthew, 5:6.
Rab.mAbot.1.3=Rabbinic Literature, Mishna, Abot 1:3.
Rab.bBer.25.a=Rabbinic text, Babylonian Talmud, Berakot 25a.
QL.1QapGen.1.3=Qumran Literature, Genesis Apocryphon from Qumran Cave 1.

Note that there should be a period even prior to the page or folio in a Talmud reference.

Such references are to be used in the attribute "N" for QUOTE and in various notation and bibliographical elements.

IV. Conclusions

SGML/TEI markup is particularly useful for scripturally based text, i.e., texts from the vast literatures of Judaism, Christianity and Islam which frequently cite biblical or koranic verses. There are numerous genres in these religious literatures (exegetical works, homilies, scriptural essays, dialogues, legal texts, etc.). They all have in common the citation of texts sacred to a religious community, the frequent mention of characters, places and institutions found in such texts, plus references to later religious individuals, places and

institutions. In addition, these texts are often macaronic, i.e., they contain more than one human language.

Such texts offer particular problems for electronic presentation, apart from the issues of the non-existence of SGML software for viewing correctly encoded Semitic languages. This paper has focused on technical issues, the solution of which will be indicated in an Encoding Manual used both as a supplement to viewing the electronic text and as a guide for those participating in the encoding process.

V. Bibliography

Ide, Nancy and Jean Veronis, eds., Text Encoding Initiative: Background and Context (Kluwer Academic Publishers: Dordrecht/ Boston/ London, 1995).

Kaufman, Stephen A., The Comprehensive Aramaic Lexicon: Text Entry and Format Manual, Comprehensive Aramaic Lexicon, Baltimore 1987.

Milikowsky, Chaim, The Henkind Talmud Text Database, Lieberman Institute for Talmudic Research (JTS): "Directions for Text Copyists, Hebrew and English versions," Jerusalem and New York, no date.

Moscovitz, Leib, Responsa Version 3.0: User's Guide, Bar-Ilan University, Ramat Gan 1994.

Robinson, Peter, The Transcription of Primary Textual Sources Using SGML, Office for Humanities Communication Publications Number 6, Oxford 1994.

Schäfer, Peter and/or Gottfried Reeg, "Konventionen zur Aufnahmen von Handschriften für die Datenverarbeitung," Berlin, Stand: November 1991.

Sperberg-McQueen and Lou Burnard, ed. Guidelines for Electronic Text Encoding and Interchange (TEI3), Chicago, Oxford, April 8, 1994.

Endnotes

1 See the numerous articles cited in notes by Jacob Elbaum, "Rhetoric, Motif and Subject-Matter: Toward an Analysis of Narrative Technique in Pirqe Rabbi Eliezer," Jerusalem Studies in Jewish Folklore, XIII-XIV, (1992), 99-126. In addition to an early Latin translation, in the twentieth century Pirqe R. Eliezer has been translated into English, French and Spanish.

2 Bar Ilan Database (Responsa Database, Bar Ilan University); STM Database (Polytext, Jerusalem) and Davka database of Rabbinic Literature.

3 Bibliographical references for the debate on this issue between Schäfer and Milikowsky, and the comments of M. Beit Aryeh will be provided.

4 Peter Robinson has written, "perhaps the most important decision an encoder of scholarly text must face is how the text should be divided (Transcription, p. 64)."

5 Traditional citing most often utilizes the pagination of the edition of the RaDaL, the page division of the edition of Higger, or occasionally reference to the "critical edition" of C. M. Horowitz. The problems of all these texts will be discussed in a separate document "Introduction: the Need for a Critical Edition of Pirqe R. Eliezer."

6 My thanks to Robin Cover for reminding me of this.

7 For abbreviations of journals, etc., additional items are found in the Index of Articles on Jewish Studies, (The Jewish National and University Library: Jerusalem, 1995 and earlier), "List of Periodicals and the Collections and their Abbreviations," and International Glossary of Abbreviations for Theology and Related Subjects, ed. Sigfried Schwertner (Walter de Gruyter: Berlin and New York, 1974).

8 The MILESTONE tag LB (line break) will also use a four part structure for the ID attribute. Example: PRE.04.26b.1. This refers to the work Pirqe R. Eliezer.; manuscript 04 (so designated in manuscript database0); folio 26b (+ a = rechto or b = verso); line 1.