ALEF BIT

Hebrew Character Sets: A Review and Update of *Library Automation Worldwide*

Joan Aliprand
Research Libraries Group, Inc.
Mountain View, CA

Why should I care about character sets?

If text in a particular language is to be manipulated on a computer, the computer's character repertoire must be sufficient to write comprehensible text in that language. For the text to be manipulated properly by the computer, the discrete components of the language must be unambiguously identified (encoded). For alphabetic languages, the language components that are encoded are the letters of the alphabet, together with other required characters, such as punctuation, numeric digits, symbols, and (in the case of Hebrew) vowels and other marks of pronunciation. In sum, the character set used in your computer program must be adequate for the language you are going to write.

When data remains within a particular system, or set of identical systems, one encoding scheme is used. But data may be exchanged between computers, where the encoding schemes are not necessarily the same. Kuperman (1987-88) describes the problems that ensue. Agreed-upon standards prevent such problems. "Technical standards ... are established so that people and agencies can share information in a known, consistent manner" (Crawford, 1991, p. 48). Cataloging rules are one type of library standard; character sets used to encode machine-readable library records are another.

Currently, USMARC bibliographic records are encoded in a default Latin script character set and (optionally) in one or more character sets for non-Roman scripts. The use of multiple character sets is cumbersome from a data processing point of view. The deficiencies of multiple character sets led to the development of a single character set that would encode all the major scripts of the world together. The parallel efforts that led to Unicode and ISO's Universal Character Set have been merged in ISO DIS 10646-1.2, which (if approved as an International Standard) will be the character set of the future.

The existence of a single multiscript character set allows development of computer applications generalized to handle any script. Some of the problems of current word-processing programs for Hebrew (as reported in the electronic newsletter E-HUG, for example) stem from the fact that the Hebrew functionality is an addition to a program designed for a left-to-right world. A global word-processing program will not have such problems; facilities to handle scripts running in opposite directions will be an integral part of the design. We will see some exciting products from American software producers before very long!*

**To subscribe to E-HUG, The Electronic Hebrew Users' Newsletter (Edited by Ari Davidow), send a one-line message to: LISTSERV@DARTCMS1.BITNET. The one-line message should read:*
SUB E-HUG First Last,
where "First" is your first name, and "Last" is your last name.

Introduction

The editors of *Judaica Librarianship* received an announcement about *Library Automation Worldwide*, by John Clews (Harrogate, Yorkshire: Sesame Publications, 1988), and invited me to write a review of this book. Rather than reviewing *Library Automation Worldwide* in its entirety, I have chosen to concentrate on its coverage of Hebrew character sets and the MARC Format. (For an explanation of abbreviations such as *MARC*, see Appendix A.)

In the four years since the publication of *Library Automation Worldwide*, there have been major developments in the area of standards for character sets: the issuance of *The Unicode Standard, Version 1.0* (1991-92) and its proposed merger with ISO's *Universal Character Set*, also known as *ISO 10646* (1992). *Library Automation Worldwide*, although highly current when it was published, is now dated. This review includes a discussion of the Hebrew component of the Unicode standard to update the information in *Library Automation Worldwide.*

Hebrew Character Sets in *Library Automation Worldwide*

The earliest Hebrew character set (Figure 63 in *Library Automation Worldwide)* was a combined Latin/Hebrew set in which the 27 letters of the Hebrew alphabet (22 basic letters plus five final forms) replace the lowercase letters and spacing grave [accent] of ASCII. This character set is an approved national standard of Israel, SI 960 (1976), and has been implemented in Israeli-manufactured computer equipment. It was used as the character set in the earliest iteration of the ALEPH System (Lazinger, 1991).

As a 7-bit character set (in which each character is represented as a seven-digit binary number), this Latin/Hebrew set reflects the technical environment in which it was developed, and it allowed Hebrew script to be manipulated by computer. The restriction of Latin script to uppercase was a necessary trade-off; 7 bits allow only 128 unique patterns, so lowercase Latin and Hebrew could not both be accommodated.

The loss of lowercase Latin letters was overcome in the 8-bit Latin/Hebrew character set issued by the International Organization for Standardization as part of the ISO 8859 family of standards. Eight-bit encoding provides 256 unique patterns. ISO 8859-8, *Latin/Hebrew Character Set* (Figure 64 in *Library Automation Worldwide*), contains a full Latin-character repertoire – equivalent to that of ASCII – in the lower range (first half) of the character set. The upper range (second half) contains the Hebrew alphabet, together with Latin-script symbols and punctuation. The Latin-script symbols and punctuation are derived from those at the same code positions in ISO 8859-1, *Latin Alphabet No. 1*; the differences are that the inverted exclamation mark and the inverted question mark are omitted from 8859-8, and the multiplication sign and the division sign replace the feminine and masculine ordinal indicators. (All the 8859-1 characters that are absent from the 8859-8 repertoire are from Spanish.)

The Standards Institution of Israel is working on an 8-bit character set similar to that of ISO 8859-8. The proposal (Clews, Figure 65) includes not just consonantal letters, but *nekudot* (vowel points) in the upper range as well; the *nekudot* replace the Latin-script symbols and punctuation found in ISO 8859-8.

Library Automation Worldwide contains a quite remarkable array of proposals for standards, as well as character sets that have been published by national or international standards bodies as formally-approved consensus standards. Clews does not, however, include an illustration of the earliest ISO/TC46 proposal (1979) for a bibliographic character set for Hebrew (which was the starting point for development of the RLIN Hebrew character set). The draft ISO set included not just the Hebrew alphabet, vowels, and other non-spacing marks used to write modern Hebrew, but also a fairly complete set of cantillation marks.

The RLIN Hebrew character set (developed by RLG, with the advice and assistance of Bella Hass Weinberg) is included in the work under review (Clews, Figure 66). (The RLIN Hebrew character set has also been published in Aliprand (1986-87).) Clews notes the inclusion of Yiddish digraphs in the RLIN character set, and goes on to say that "other character sets code these pairs as two separate letters" (p. 97). This is not precisely correct: other character sets do not encode the Yiddish digraphs *at all*. As a result, the Yiddish digraphs are "spelled out" as two occurrences of *vav*, two occurrences of *yod*, or the sequence *vav yod*.

In discussing the Yiddish digraphs, Clews adds, parenthetically, "the Library of Congress uses only SI 960" (p. 97). This is totally incorrect. The Library of Congress has never implemented a Hebrew script capability on its own automated system; it uses RLIN as its online system for Hebrew cataloging. Clews is confusing the concept of *character set* (characters with a specific code assigned to each; for example, SI 960) with that of *character set repertoire* (characters alone; for example, the Hebrew consonants). Even considering the characters alone, the statement is wrong. SI 960 includes only uppercase Latin-script characters, symbols and punctuation; the Latin character set used by the Library of Congress is more extensive.

What Clews *really* means is this: in using RLIN, the Library of Congress imitates the unpointed font used for its most recent Hebraica catalog cards. Earlier LC Hebrew cards did include vocalization, however (Weinberg). I assert that vocalization which appears on the source of information *should* be transcribed, because the author or publisher included it for a purpose (e.g., to indicate an uncommon pronunciation), and failing to transcribe it violates AACR2 Rule 1.1B1: "Give accentuation and other diacritical marks that are *present* in the chief source of information" (emphasis mine).

Clews suggests that the RLIN Hebrew character set "could possibly be put forward as an ANSI standard" (p. 97), that is, be proposed and balloted as an American national standard. This was the course followed for ANSI Z39.64, *East Asian Character Code* (EACC, 1990): the RLIN East Asian Character Code (REACC) was the source for this standard.

Endorsement of the RLIN Hebrew character set as a standard for bibliographic use has taken a different path. The character set was adopted by the Library of Congress for the alternate graphic representation of Hebrew data in USMARC records, and appears in the 1990 edition of the *USMARC Specifications*. In addition, the Library of Congress, through NISO, proposed the RLIN Hebrew character set as an international standard for the exchange of bibliographic data in Hebrew. The proposal has languished due to lack of action on the part of ISO/TC46/SC4/WG1 (WG1 currently has only an "acting convenor").

Work on character sets in ISO is handled by two groups: ISO/TC46/SC4/WG1 and ISO-IEC JTC1/SC2/WG4 (IEC = International Electrotechnical Commission; JTC = Joint Technical Committee; TC = Technical Committee; SC = Subcommittee; WG = Working Group). Since publication of *Library Automation Worldwide*, ISO-IEC JTC1/SC2/WG4 has superseded ISO/TC97/SC2/WG4. Clews deplores the division of responsibility for character sets within ISO, and the parallel dichotomy in national standards organizations (p. 165); he is quite right to do so. Librarians should be represented on ISO-IEC JTC1/SC2/WG4 (and the corresponding national groups concerned with computerized character sets, such as ANSI X3L2), instead of creating a competing set of standards.

Effect of the Absence of an International Standard on the Exchange of Hebrew Data

There is no urgent demand for an approved international bibliographic standard for Hebrew, because only two parties can utilize such data: RLIN and ALEPH. Furthermore, exchange of bibliographic data between Israel and the United States (as proposed by Adler, 1987-88) does not require an international character set for Hebrew. The ALEPH system can "convert MARC format records to ALEPH structure" (Lazinger, 1991, p. 283). In mapping Hebrew script data from USMARC to ALEPH, character conversions would be required (for example, to turn the Yiddish digraphs into their component Hebrew letters), and some characters (the *nekudot*) would have to be dropped. In the other direction, ALEPH's Hebrew character set, consisting of the consonantal letters, maps exactly to USMARC/RLIN characters. The difficulties in exchanging data between ALEPH and RLIN lie more in their different approaches to field structure than in their character sets *per se*.

Kuperman has pointed out the need for a common encoding of Hebrew for other applications (such as word-processing):

> The lack of a standardized character set is significant in transferring files between computer programs or between computers. Since there is no standard coding for Hebrew, there is no assurance that a file generated by one computer can be utilized by another computer. This makes it very

difficult to send electronic mail or operate an online data base in this field. Files produced by one system are often gibberish when read by a different system.

(Kuperman, 1987-88, p. 7)

He looks forward to a future standard for Hebrew:

> When Hebrew coding becomes standardized, the display, printing, and transfer of Hebrew data will become routine. Software developers will produce utilities to convert files to the international standard. Standard coding for Hebrew will encourage software developers to produce fonts for screen display and printing to be used with other companies' word processors, whereas at present, display and printing systems are necessarily unique to each system.
>
> (Kuperman, 1987-88, p. 8)

A New Character Set for Hebrew (and Other Scripts)

The most recent events in character-set work are the creation of the Unicode standard, and its merger into ISO's *Universal Character Set* (*UCS*), which is also known by its number, *10646 (ten-six-four-six)*. Clews describes the UCS (as of 1988) in section 11.5 of his book, under the heading *The ISO/TC97 Multiple-Byte Character Set* (p. 157-159). The UCS progressed to Draft International Standard status (*DIS 10646*), but failed to achieve a sufficient number of votes for approval as an International Standard. Following this, an *ad hoc* committee of character-set experts proposed a merger of Unicode and DIS 10646. Discussions between the Unicode Consortium and the ISO working group responsible for the UCS led to a synthesis of the two character sets, which is currently being distributed for ballot as *ISO DIS 10646-1.2* (1992).

A draft of Israel's comments to accompany its vote on ISO DIS 10646-1.2 was distributed to various experts for comment (Rosenne, 1992). Israel intends to recommend that the Yiddish digraphs (currently in the Basic Multilingual Plane because they are in *The Unicode Standard)* be considered merely presentation variants, meaning that there would be no requirement to retain the digraphs when text was subsequently transmitted; unannounced substitution of the individual Hebrew consonants would be permitted. Israel also intends to recommend that the *varika* (a diacritic used in Ladino) be dropped, since the *rafe* (a diacritic used in Yiddish) can be used instead. (U.S. review of ISO DIS 10646-1.2 is the responsibility of the ANSI X3L2 subcommittee.)

The first volume of *The Unicode Standard, Version 1.0* was published in 1991. Unicode, a 16-bit character encoding, covers all the major modern scripts of the world (including Hebrew). A 16-bit encoding provides for over 65,000 unique characters. In the Unicode standard, Hebrew is encoded in the range U+0590 through U+05F5 (the "Hebrew block" is shown in Figure 1), and the New Israeli Shekel is in the Currency Symbols block at U+20AA. The convention "U+" indicates that the four-digit number (written in the hexadecimal notation of 0 through F) is a Unicode value.

Since the Research Libraries Group, Inc. (RLG) is a member of the Unicode Consortium, and I was a member of the Working Group that developed the character repertoire and encoding, it is hardly surprising that almost all characters in the Hebrew block of the Unicode standard also occur in the USMARC/RLIN Hebrew character set. Unicode adds the characters HEBREW POINT METEG (or *siluq*), HEBREW POINT PASEQ (or *legarmeh*), and HEBREW PUNCTUATION SOF PASUQ (U+05BD, U+05C0, and U+05C3, respectively); it also differentiates between the vowel point *ḥolam* and the diacritic *śin dot* (which are graphically identical) by encoding them as separate characters (U+0589 and U+05C2).

The names of most of the Unicode characters in the Hebrew block agree with names and romanizations established by the Standards Institution of Israel; the authoritative information was supplied by Jony Rosenne of IBM Israel. The names of Yiddish digraphs follow the Unicode convention for description of the character, and "double" was thought to be more intelligible to English speakers than *tsvey*. The romanizations of the culturally-correct names of the "double" digraphs are also given (but appear to have been omitted from the *Unicode Name Index*).

059	05A	05B	05C	05D	05E	05F
		◌ְ	׀	א	נ	װ
		◌ֱ	◌ׁ	ב	ס	ױ
		◌ֲ	◌ׂ	ג	ע	ײ
		◌ֳ	׃	ד	ף	׳
		◌ִ		ה	פ	״
		◌ֵ		ו	ץ	◌ﬞ
		◌ֶ		ז	צ	
		◌ַ		ח	ק	
		◌ָ		ט	ר	
		◌ֹ		י	ש	
				ך	ת	
		◌ֻ		כ		
		◌ּ		ל		
		◌ֽ		ם		
		־		מ		
		◌ֿ		ן		

Figure 1. Hebrew block from *The Unicode Standard,* Version 1.0 (1991)

Columns 05B and 05C contain vowel points (including the three *hataf* combinations), other marks of pronunciation (e.g., U+05BF, the last character in column 05B, is the *rafe*), and some distinctive punctuation (the *maqaf* at U+05BE, the *paseq* or *legarmeh* at U+05C0, and the *sof pasuq* at U+05C3).

Columns 05D and 05E contain the letters of the Hebrew alphabet (including the five final forms as distinct characters). Although final forms could be supplied algorithmically (the Unicode encoding of Arabic script presupposes this), Hebraic final forms were treated as separate letters to provide compatibility with existing standards (e.g., ISO 8859-8).

Column 05F contains three Yiddish digraphs, as well as the *geresh,* the *gershayim,* and the *varika* used in Judezmo (Ladino).

The Hebrew block of the Unicode standard does not include cantillation marks, but this has nothing to do with the fact that cantillation marks were excluded from the RLIN Hebrew character set (Aliprand, 1986-87, p. 6). Shortly before the Unicode standard was finalized, the Unicode Technical Committee (UTC) learned that the Standards Institution of Israel was preparing a standard for cantillation marks. The UTC decided that the inclusion of cantillation marks should be deferred, so that the repertoire and order of cantillation marks in the Unicode standard could be aligned with those of the Israeli standard.

Like USMARC, the Unicode standard encodes non-spacing marks (which, in rendering, are combined with a base letter). Unlike USMARC, the Unicode standard mandates that the non-spacing mark(s) *follow* the base letter. (Hebraica catalogers are used to keying the diacritic for *ḥet* before the letter, for example.) Unlike USMARC, the Unicode standard also encodes precomposed letter/diacritic combinations (for example, U+015B LATIN SMALL LETTER S ACUTE) as distinct characters in the Latin script blocks; this provides round-trip mapping, i.e., complete compatibility, between Unicode and existing ISO standards. Precomposed characters on the model of the Latin letter/diacritic combinations are not encoded in the Hebrew block (nor in the blocks for other non-Latin scripts); there are simply too many possible combinations!

Whether the New Israeli Shekel symbol should be included in the Unicode character repertoire was debated several times by the Working Group (predecessor to the UTC). Some sources asserted that it was merely a decorative element used on Israeli checks. Although the symbol is not widely used in Israel, the final decision was (as noted above) to include it in the Currency Symbols block.

The *Unicode 1.0 Draft Standard* was sent out for worldwide review in December 1990. A number of respondents requested encoding of dotted *śin* and *shin* as discrete (i.e., precomposed) characters. This expressed need was given serious consideration by the Unicode Technical Committee (which superseded the informal Working Group upon incorporation of the Unicode Consortium). The UTC decision was not to add these precomposed characters. There are no precomposed characters in the Hebrew block, and so pointed text can be converted to its unpointed equivalent by removing the non-spacing marks. If *śin* and *shin* were precomposed characters, there would need to be another step to convert them to the undotted equivalent.

The encoding of *śin* and *shin* as HEBREW LETTER SHIN followed by the appropriate dot (i.e., each as two Unicode characters) does not mean that the dotted letters cannot be treated as single units in functions such as keying. A keyboard, for example, could have separate keys for undotted *shin*, for *shin*, and for *śin*. *Śin* and *shin* can be distinguished from each other in string-matching, even when encoded with multiple characters; in other words, the sequence unmarked SHIN followed by SIN DOT is not the same as unmarked SHIN followed by SHIN DOT.

In addition to the panels of characters showing code values, *The Unicode Standard, Version 1.0* includes text and tables to aid implementers of the standard. One aid is the *Bidirectional Algorithm,* which consists of rules governing the presentation of mixtures of scripts of opposite directionality (e.g., English and Hebrew). When text is encoded in Unicode, the characters are in "logical" order, i.e., corresponding to the order in which text is typed on a keyboard after corrections such as insertions, deletions, and overtyping have taken place. The Bidirectional Algorithm provides rules so that text in "logical" order can be presented correctly for reading, in the absence of any other information about how the data should be displayed.

It should be noted that "logical" order is the only possible order for bidirectional text *that is to be exchanged among systems*. Other ordering methodologies, while satisfactory within a self-contained system, are not unequivocally compatible with all systems. The *USMARC Specifications* endorse "logical" order.

The USMARC Format in *Library Automation Worldwide*

Character sets are not an end in themselves: their purpose is to allow text to be represented in machine-readable form. Despite its main title, the focus of *Library Automation Worldwide* is almost exclusively on character sets; the book's subtitle, *The Development of Character Set Standards*, is more representative of its scope.

Character set standards are intimately linked to alternate graphic representation in the *USMARC Format for Bibliographic Data* (1988-) and the *USMARC Format for Authority Data* (1987-). The small amount of information on USMARC is buried in Clews' section on EACC, and cannot be found (except by browsing) since *Library Automation Worldwide* lacks an index. The only detailed information on how non-Roman character sets are used in USMARC appears in Clews' Figure 116, *Character Sets and Escape Sequences in CJK MARC Records.* The source of the information in this figure is not the Library of Congress (which one might expect), but OCLC. There is really no excuse for relying on a secondary source when the primary source is known: the "Footnotes" (i.e., the sources of information at the end of the book) include several Library of Congress publications, including the 1987 edition of *USMARC Specifications for Record Structure, Character Sets, Tapes.*

The author fails to explain the differences between the USMARC specification for alternate graphic representation and OCLC's implementation of it (as described in Figure 116). Three such differences are:

a) Figure 116 says that "each 880 field will be linked . . . to a bibliographic description field," but fails to mention that an 880 field may be linked to a hypothetical Roman field that does not occur in the record.

b) Figure 116 lists the fields for which alternate graphic representation (exemplified by the presence of subfield 6) is valid: 100-262, 265, 310-584, 600-653, 700-840, 886, and 936. (The meaning of the field tags is given in Appendix B.) In 1988, the permitted range of fields for alternate graphic representation was 1XX through 8XX. ("1XX" stands for all MARC tags in the range 100 through 199; "8XX" for tags from 800 through 899.) In USMARC, it is legitimate for a 300 field to have an alternate graphic representation; the absence of field 300 from the list of fields in Figure 116 must represent a feature of OCLC. Field 936 is in the range for locally-defined fields, and is therefore OCLC-specific.

c) The statement, "Nonfiling indicators have no significance in field 880 . . . " is incorrect. Cases are known of fields that contain predominantly CJK data, but begin with an English article. The statement that "the value [of the indicators in the 880 field] should be the same as in the linked bibliographic field" is patently untrue with respect to Hebrew, since the definite article is written with one Hebrew letter, but with two letters and a hyphen when Romanized. Thus the nonfiling indicator is set to 1 for the field with Hebrew script and to 3 for the Romanized one (when the first word in the field begins with the definite article).

Miscellaneous Comments on Coverage of Hebrew in *Library Automation Worldwide*

Clews consistently fails to identify the sources of the general information on each script (Figures 61 and 62, which list "Hebrew consonants" and "Hebrew vowels," respectively, are obviously reproduced from a book). On page 93, the *Matres Lectionis* (Hebrew letters that function as vowels) are referred to as "w" and "y"; on pages 94 and 95, romanized names of Hebrew letters are used (with inconsistent capitalization).

Yiddish is given short shrift. Clews writes, " . . . for Yiddish, some consonants have been used alphabetically to express vowels" (p. 93). While this is true, he fails to point out another feature of Yiddish: the use of vocalic marks to distinguish the vowels *pasekh alef* and *komets alef*. The source of the statement, "In the Jewish Autonomous Region of the USSR, Yiddish is now a completely alphabetical Hebrew script" (p. 93) is not given. *[Soviet Yiddish orthography does not preserve the consonantal spelling of Hebrew words, but it does include some vowel points. -B.H.W.]*

The statement, "Other Hebrew character sets follow the same coding for Aleph to Taw" (p. 95) is incorrect. All character sets for this alphabet have the same *sequence* of consonants (i.e., *alef* to *tav*), but the coding range for the Hebrew consonants is 6/0 through 7/10 in some sets, and 14/0 through 15/10 in others. (For an explanation of the row/column referencing convention, see Aliprand, 1986-1987, p. 15, Note 3.)

Conclusion

When it was published in 1988, *Library Automation Worldwide* was a unique and definitive resource on character sets, which made difficult-to-find material readily accessible. The text of the book is well organized. Not all sources of information are acknowledged; this is unfortunate, as there is no way to evaluate the reliability of certain statements. The sources that are cited should have been gathered in a bibliography, arranged either in author/title order or by script, subarranged by author/title.

The constellations of discrete 7- and 8-bit character sets are fading, and the day of the universal character set is upon us. *Library Automation Worldwide* documents what has been (and, in some cases, what might have been); it can also function as a character quarry for the scripts that are not yet in *The Unicode Standard* (Maldivian, Amharic, Burmese, and Khmer).

By bringing many character sets together, Clews has illustrated the discrepancies that exist between so-called standards, and the duplication of effort in standards-making. He makes the excellent point that librarians should not create their own character sets, but should make sure that unique bibliographic needs are met by computer industry sets. RLG's membership in the Unicode Consortium and representation on ANSI X3L2 constitute the beginning of this trend.

The likely establishment of a universal character set standard in 1992 has stimulated much discussion about its effect on other standards and technologies. Rather than creating standards that relate to outdated and cumbersome character-set methodologies, library standards makers need to view the future and consider how libraries will be affected by the Universal Character Set.

Judaica librarians will not be immediately affected by these developments. Their needs are already being met by library systems, such as RLIN and ALEPH, that accommodate input and searching in Hebrew and Arabic as well as Latin script. The effect of the global character set will be seen first in software from U.S. companies (for example, word-processing programs) able to handle bidirectional text and a variety of scripts elegantly and easily.

References

[AACR2] *Anglo-American Cataloguing Rules.* 2d ed., 1988 revision. Chicago: American Library Association, 1988.

Adler, Elhanan. "The Use of Israeli Machine-Readable Cataloging by American Libraries: A Proposal," *Judaica Librarianship*, vol. 4, no. 1 (1987-1988), p. 23-26.
Correspondence in response to this article appeared in *Judaica Librarianship*, vol. 4, no. 2 (1988-1989), p. 138-9.

Aliprand, Joan M. "Hebrew on RLIN," *Judaica Librarianship,* vol. 3, no. 1-2 (1986-1987), p. 5-16.

Clews, John. *Library Automation Worldwide.* Harrogate, Yorkshire: Sesame Publications, 1988.

Crawford, Walt. *Technical Standards: An Introduction for Librarians.* 2nd ed. Boston : G.K. Hall, [1991].

[EACC] American National Standards Institute. *East Asian Character Code for Bibliographic Use.* New Brunswick: Transaction, 1990. (ANSI Z39.64).

[ISO 8859-1] International Organization for Standardization. *Information Processing – 8-bit Single-Byte Coded Graphic Character Sets. Part 1: Latin Alphabet No. 1.* Geneva, [1987]. (ISO 8859-1).

[ISO 8859-8] International Organization for Standardization. *Information Processing – 8-bit Single-Byte Coded Graphic Character Sets. Part 8: Latin/Hebrew Alphabet.* Geneva, [1988] (ISO 8859-8).

[ISO 10646] International Organization for Standardization-/International Electrotechnical Committee. Joint Technical Committee 1. Subcommittee 2. *Universal Character Set.* Geneva, 1992. (ISO DIS 10646.2).
Draft was approved, comments were resolved, and the standard will be published late in 1992. Its basic Multilingual Plane is based on the Unicode Standard.

[ISO/TC 46 proposal] International Organization for Standardization. Technical Committee 46. Subcommittee 4. Working Group 1. *Hebrew Alphabet Character Sets for Bibliographic Use.* 2d draft. Prepared by Robert J. M. Gabriel. Ghent, 1979. (ISO/TC46/SC4/WG1 N98).

Kuperman, Aaron Wolfe. "Hebrew Word Processing: A Review of Available Products," *Judaica Librarianship*, vol. 4, no. 1 (1987-1988), p. 6-12.

Lazinger, Susan S. "ALEPH: Israel's Research Library Network: Background, Evolution, and Implications for Networking in a Small Country," *Information Technology and Libraries*, vol. 10, no. 4 (Dec. 1991), p. 275-91.

Rosenne, Jony. *Draft vote on ISO/IEC DIS 10646-1.2.* Electronic mail sent 29 Mar 92 20:11:39 IST.

[SI 960] Mekhon ha-teḳanim ha-Yiśre'eli. (Standards Institution of Israel). *7 Bit Coded Character Set for Information Processing Interchange (Translation).* Tel Aviv: The Standards Institution of Israel, 1976.
The translation was published by the Standards Institution of Israel, but states that "Only the original standard in Hebrew is authentic."

The Unicode Consortium. *The Unicode Standard: Worldwide Character Encoding,* Version 1.0. Reading. MA: Addison-Wesley, 1991-
Volume 1 was published in 1991; Volume 2 (covering East Asian ideographs) was published in 1992. The Unicode Technical Committee says that Version 1.1 of *The Unicode Standard* will match the character repertoire and code point assignment of the Basic Multilingual Plane of IS 10646.

USMARC Format for Authority Data : Including Guidelines for Content Designation. Prepared by Network Development and MARC Standards Office. Washington: Cataloging Distribution Service, Library of Congress, 1987-

USMARC Format for Bibliographic Data: Including Guidelines for Content Designation. Prepared by Network Development and MARC Standards Office. Washington: Cataloging Distribution Service, Library of Congress, 1988-

[USMARC Specifications] *USMARC Specifications for Record Structure, Character Sets, Tapes.* Prepared by Network Development and MARC Standards Office. 1990 edition. Washington: Cataloging Distribution Service, Library of Congress, 1990.
This is the latest edition. An earlier edition was published in 1987.

Weinberg, Bella. Personal communication.
An example of a Hebraica card featuring vowel points is LCCN 25-7512, for the book: Kasovich, Israel Isser. *Dos Eybige folḳ* . Nyu Yorḳ : [s.n.], 1925 (Nyu Yorḳ : Pinsḳi-Mazl Pres).

Appendix A: Abbreviations

ANSI: American National Standards Institute, the U.S. standards-making body

ANSI X3L2: "X3L2" has no meaning. Committee X3 deals with computing; Subcommittee L2 with character sets. ANSI X3L2 is the national equivalent of the international group ISO-IEC JTC1/SC2/WG4.

ASCII: American Standard Code for Information Interchange (ANSI standard X3.4)

CJK: Chinese/Japanese/Korean (*CJK* is a trademark and service-mark of The Research Libraries Group, Inc.)

DIS: Draft International Standard

EACC: East Asian Character Code (ANSI standard Z39.64)

ISO: International Organization for Standardization (the abbreviation ISO is from the Organization's name in French)

ISO-IEC JTC1/SC2/WG4: International Organization for Standardization/ International Electrotechnical Committee. Joint Technical Committee 1. Subcommittee 2. Working Group 4

ISO/TC46: International Organization for Standardization. Technical Committee 46

ISO/TC46/SC4/WG1: International Organization for Standardization. Technical Committee 46. Subcommittee 4. Working Group 1

ISO/TC97/SC2/WG4: International Organization for Standardization. Technical Committee 97. Subcommittee 2. Working Group 4. (Superseded by ISO-IEC JTC1/SC2/WG4)

MARC: Machine-Readable Cataloging

NISO: National Information Standards Organization, formerly ANSI Committee Z39

OCLC: Online Cooperative Library Center, Inc., and its bibliographic system

REACC: RLIN East Asian Character Code

RLG: The Research Libraries Group, Inc. (the organization that owns and operates RLIN)

RLIN: Research Libraries Information Network (bibliographic system)

SI: Standards Institution of Israel

UCS: Universal Character Set

USMARC: United States [format for] Machine-Readable Cataloging

UTC: Unicode Technical Committee

Appendix B: USMARC Tags

USMARC fields are identified by means of a "tag," a sequence of three numbers. The following fields are referred to in the text by their tags:

100 Main Entry – Personal Name

262 Imprint Statement for Sound Recordings (Pre-AACR2)

265 Source for Acquisition/Subscription Address

300 Physical Description

310 Current Frequency

584 Accumulation and Frequency of Use Note

600 Subject Added Entry – Personal Name

653 Index Term – Uncontrolled

700 Added Entry – Personal Name

840 Series Added Entry – Title (Pre-AACR2)

880 Alternate Graphic Representation

886 Foreign MARC Information Field

936 Dates or Volume Designations of Pieces Used for Cataloging (OCLC-defined field; the RLIN equivalent is the PUC field, which applies only to pre-AACR2 cataloging)

Joan Aliprand is Library Systems Analyst at The Research Libraries Group, Inc. in Mountain View, CA. She was responsible for the development of the Hebraica character set for RLIN, and contributed to The Unicode Standard, Version 1.0.