

ISO
International Organization for Standardization

Information and documentation – ISO 259-3

**CONVERSION OF HEBREW CHARACTERS
INTO LATIN CHARACTERS
Part 3: Phonemic Conversion**

February, 1999

Descriptors: transliteration, phonemic conversion, phonemic (structural) script, Hebrew script

INTRODUCTION

This part of ISO 259 completes the series of standards dealing with the conversion of Hebrew characters into Latin characters. It provides a complete solution for conversion of non-pointed script, which can be reconstructed into the original Hebrew characters by both man and machine. This part of ISO 259 may be used in a simplified version, as ISO 259 has a simplified version: 259-2.

The difference is that:

- (a) with some limitations, a cluster of two identical consonants (4.2) may be written in the simplified version by one character.
- (b) Some on-phonemic vowels (4.3), which are not written in the main version, may be introduced. For details, see section 5.

The main version of this part of ISO 259 (not simplified) is also in compliance with pointed Hebrew script, which may also be reconstructed from the converted text by either a man or a machine. Moreover, pointed Hebrew texts may be reconstructed from the converted texts according to this part of ISO 259 (main version) even when the conversion stems from non-pointed Hebrew texts.

Conversion of pointed Hebrew script according to this part of ISO 259 can be performed automatically by either a man or a machine. No profound knowledge of the language is needed. For reliable conversion of non-pointed Hebrew texts, the user should be somewhat acquainted with the Hebrew language, or be able at least to use a Hebrew dictionary and a grammar manual.

However, the non-pointed script is the method mostly used in modern Hebrew, while pointed script serves mainly for poetry, prayer-books and children's literature. Some exceptions where the non-pointed script is used may be found even in these areas.

Non-pointed script lacks essential ingredients of the word. Most vowels are not written, gemination of consonants is written with one character and small particles are attached to the subsequent word, with no blank space between them. Mechanical transliteration of the characters in non-pointed script into Latin characters does not help a reader who does not know Hebrew and is a heavy burden on a reader who does know it.

The conversion system of this part of ISO 259 is seemingly based on the transliteration approach (see documents ISO 259-2:1994(E), clause 2.2 or ISO 250:1984(E) clause 0.2.1.2). But it is not a transfer of the given signs in the pointed Hebrew script to signs in Latin characters nor of the signs in the non-pointed script. The basis for conversion according to this part of ISO 259 is the structure of the Hebrew word and not the method in which it is written in one of the Hebrew script systems, nor the manner in which it is pronounced. The conversion of this part is a phonemic conversion. Therefore, even though this part of ISO 259 provides a solution mainly for conversion of non-pointed script, it can be adapted to all the types of Hebrew script: pointed, quasi-pointed and non-pointed script.

Figure 1 shows the connections between a word and its written forms. The present approach (259-3) can be seen as a circumvention of the tedious problems which stem from the various realisations of the language in the various methods of writing.

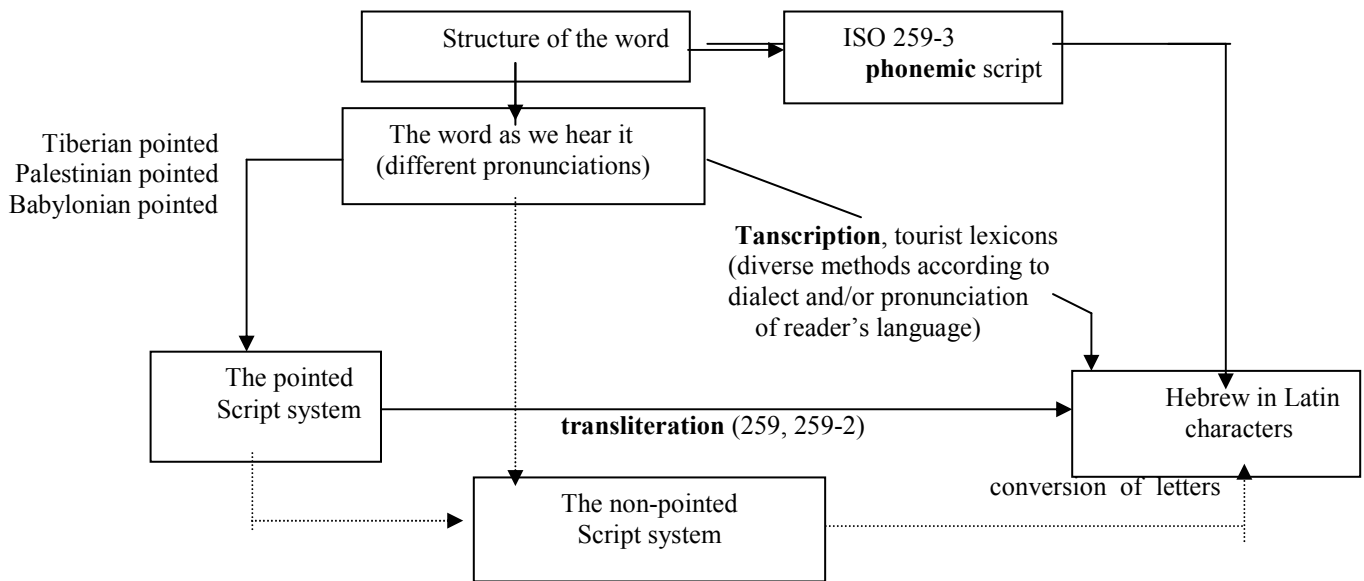


Figure 1 – The Hebrew word and its script systems

This part of ISO 259 aims at achieving the following goals:

- (1) To obtain true representation of every structural element, either a consonant or a vowel;
- (2) To present clear and easy rules for writing;
- (3) To enable simple re-conversion to both pointed and not-pointed versions of original Hebrew script.

By using the conversion system described in this part of ISO 259, not only a non pointed Hebrew text can be perfectly constructed by a person or a computer, but also a pointed Hebrew text (even if it was not pointed in its original form)

1 SCOPE

This part of ISO 259 may be used by anyone who has to write Hebrew words in Latin characters, for example in electronic mail and telegrams, and especially for catalogue files of libraries, geographic maps, indices and various types of books. Some knowledge of Hebrew, or ability to use a Hebrew dictionary and a grammar manual, may facilitate the conversion.

2. NORMATIVE REFERENCES

ISO 259:1984(E): Documentation – Transliteration of Hebrew Characters in Latin characters.

ISO 259-2:1994(E): Information and documentation – Transliteration of Hebrew characters into Latin characters, Part 2 : Simplified transliteration.

ISO 10646-1: 1993(E): Information technology – Universal Multiple-Octet Coded Character Set (UCS) – Part 1: Architecture and Basic Multilingual Plane.

3 DEFINITIONS AND TABLES OF CONVERSION

3.1 Consonants: Each one of the signs in the left column, in Table 1 below, with the corresponding Latin character alongside it.

Table 1 (The short identifier of each character, derived from ISO 10646-1 appears in parenthesis, next to the character)

| I Hebrew Char. (identifier from ISO 106-46) | II Latin char. (identifier from ISO 106-46) | III Notes (on pointing etc.) | I Hebrew Char. (identifier from ISO 106-46) | II Latin char. (identifier from ISO 106-46) | III Notes (on pointing etc.) |
|--|--|---------------------------------------|--|--|---------------------------------------|
| א (05D0) | ` (02BE) or ' (02C0) | | א (05DF) א (05E0) | n (006E) | |
| ב (05D1) ב (05D1+05BC) | b (0062) | (1) | ב (05E1) | s (0073) | |
| ג (05D2) ג (05D2+05BC) | g(0067) | (1) | ג (05E2) | ˘ (02BF) or ' (02C1) | (2) |
| ד(05D3) ד (05D2+05BC) | d(0064) | (1) | ד (05E3) ד (05E4) ד (05E4+05BC) | p (0070) | (1) |
| ה (05D4) | h(0068) | | ה (05E5) ה (05E6) | c (0063) | |
| ו(05D5) | w(0077) | | ו (05E7) | q (0071) | |
| ז(05D6) | z(007A) | | ז (05E8) | r (0072) | |
| ח (05D7) | h(1E25) | (2) | ח (05E9) | š (0161) | (2) |
| ט (05D8) | ţ(1E6F) | (2) | ט (05EA) ט (05EA+05BC) | t (0074) | (1) |
| י (05D9) | y(0079) | | י (02Bc+05D2) | ǰ(01E7) | (2) |
| ך (05DA) כ (05DB) כ (05DB+05BC) | k(006B) | (1) | ך (02BC+05D6) | ž (017E) | (2) |
| ל (05DC) | l(006C) | | ל (02BC+05E6) | č(010D) | (2) |
| מ (05DD) | m(006D) | | מ (05E9+05C2) | j(006A) | |
| נ (05DE) | | | | | |

NOTES

Note 1. The same conversion applies also to this character when used with the Hebrew Dagesh (05BC).

Note 2. Alternative signs may be used for E-mail or when the user cannot get the proper signs of the table above for conversion, as following:

א → x, ב → @, ג → &, ד → \$, 'צ ,׳ט ,׳ג → c', z', g' respectively.

In order to get the special signs for these letters, as given in the table above, a macro can be downloaded from the site of SII.

Note 3. Additional signs (aids “-”, “_”) will be examined in 4.1, 4.3.2).

Note 4: The letter א (05D0) should always be converted into ` (02BE) or ' (02C0), even when it is mute, but א as the first consonant of a word may be omitted: `amar or amar (אמר).

3.2 Vowels: Table 2 shows the vowel signs with the corresponding signs in both non-pointed and pointed Hebrew writing systems.

Table 2

| I The vowel sign in Pointed script | II The Hebrew character in non-pointed script | III The vowel in this standard |
|--|---|--------------------------------------|
| ם (x+05B2), ם (x+05B7), ם (x+05B8), (in last position): ם (05D4+x+05B8) | No sign (in last position): ם (05D4) | a (0061) |
| ם (x+05B1), ם (x+05B6), ם (x+05B5), (in last position): ם (05D4+X+05B6), ם (05D4+X+05B5) | No sign (in last position): ם (05D4) | e (0065) |
| ם (x+05B4), ם (05D9+X+05B4) | No sign ם (05D9) following ם | i (0069) |
| ם (x+05B3), ם (x+05B8), ם (x+05C2), ם (05D5+05C2+x) | No sign ם (05D5) following ם | o (006F) |
| ם (x+05BB), ם (05D5+05BC+x) | No sign ם (05D5) following ם | u (0075) |
| ם (05D9+x+05B5), ם (05D9+x+05B6) | ם (05D9) following ם | ei (0065+0069) |

(*) The user should substitute the sign ם with a suitable Hebrew letter, and the sign x in parenthesis with the corresponding identifier from ISO 10646-1.

4. CONVERSION RULES

4.1 The attached particles

The attached particles – ב-,ה-,ו-,כ-,ל-,נ-,ש (with corresponding identifiers 05D1, 05D4, 05D5, 05DB, 05DC, 05DE, 05E9 respectively) which are attached to the subsequent word in the Hebrew script, shall be converted with a trailing hyphen: b-, ha-, w-, k-, l-, mi-, še-.

Note 7: The ל at the beginning of an infinitive is not an attached particle, but rather a verbal prefix. Its conversion should not be followed by a hyphen (ligmor, ldabber, comp. l-beiti).

4.1.1 The same applies to cases where there is more than one attached particle before the word: e.g. w-ha-bayt, mi-še-ba’u.

4.1.2 Instead of b-ha-, ba- may be used; instead of k-ha- → ka-; instead of l-ha- → la-, e.g., ba-bayt (=b-ha-bayt), la-ginna (=l-ha-ginna).

4.1.3 No change shall be made in the conversion of attached particles even when there is a different point-sign in the pointed script,

e.g. שדרוּשׁ → w-daruš.
דְּרוּשִׁים → w-drušim ,
בְּרַצוֹן → b-racon
בְּרִצּוֹנִי → b-rconi.

4.2 Gemination of consonants

4.2.1 A gemination is a cluster of two identical phonemes. Such a cluster applies to consonants only, and appears only after a vowel. In the pointed script it is signified with a point in the middle of the letter (“dageš ḥazaq”). In the non-pointed script it is not represented at all, while in the converted script it is represented by a double character (e.g. the word בְּרִצּוֹן in pointed and בְּרַצוֹן in non-pointed script, is converted into huqqim).

4.2.2 A double letter shall be written to represent gemination in every case where: (a) the pattern of the word requires gemination (e.g., naggar – נָגַג); (b) there is a complete assimilation of **n** into the subsequent consonant (e.g., לְפֹל which is converted into yippol, instead of the original yinpol), since the **n** assimilates into the **p**, (like the **r** in ‘irregular’, or the **l** in ‘illegal’ in English); (c) in many cases where the root of the word contains two identical consonants (the second and third character of the root, e.g. הִקְקָה or הִקְקָה is converted into huqqa, of the root ḥqq); and (d) in a few other words, e.g. הִקְקָה – yiqqaḥ instead of הִקְקָה – yilqaḥ.

4.2.3 In cases such as mentioned in 4.2.2 we use a double character even if gemination is not indicated by a dageš ḥazaq in the pointed script, i.e. when it is one of the consonants א,ה,ח,ע,ר (in the pointed script, their characters cannot have a “dageš ḥazaq”), e.g. מִתְנַחֵל: mitnahhel (compare מִתְנַפֵּל – mitnappel). The same applies to a gemination at the end of a word, which in the pointed script is always written with one character with no “dageš”, such as הִקְקָה (קֶקֶה). The “dageš” is clearly seen, however, in the inflection, e.g. הִקְקָה or הִקְקָה in the non-pointed script: huqqim.

For pronunciation of **b,k,p** and bb,kk,pp, see Annex A.2.

4.2.4 However, gemination shall be converted as one character if it is at the beginning of a word after an attached particle ending with a vowel before the hyphen, e.g. הַבּוֹר → ha-bor (not ha-bbor, nor hab-bor.).

4.3 Non-phonemic vowels

4.3.1 There are several non-phonemic vowels in Hebrew which are enumerated in the following sub-clause. For strict linguistic purposes, such as sorting patterns, they should not be transcribed, but for regular use a simplified version may be adopted, in which at least some of them shall be written. See section 5.

4.3.2 Shewa “ְ” appears in the pointed script under a consonant letter to signify “no vowel” after the consonant. Being non-phonemic in nature, shewa should not be converted, whether it is never pronounced, as in מִדְּבָר : [midbar], whether it is pronounced as [e] in some dialects only, as in מִפְּרִי : [praxim] or [peraxim], מִגְּמֵר : [gamru] or [gameru], or whether it is practically always pronounced [e], as in מִיֵּלָדִים : [yeladim]. Needless to say, in the non-pointed script there is no Shewa at all.

4.3.3 However, when there are two adjacent identical characters in the Hebrew script, and the first of which is with a shewa in the pointed script, the two characters might be considered as forming a gemination (see 4.2.1). In this case, an underscore should be inserted between them, e.g. sob_bu = [sovevu] (סובבו) , hal_luyah = [haleluyah] (הללוייה) .

4.3.4 A segolate vowel

The vowel sign “ֿ” under the last-but-one character of a word when the last character is not ה, is never accented (unless the last-but-two character is pointed with shewa). This sign is known as “segol”, and such a word is called “segolate”, e.g. מְגֹל. In terms of the structure, a segolate word ends with a cluster of two consonants. Other segolates are words which have the vowel sign “ֿ” under the last-but-one character if the last-but-two character does not have the shewa sign “ֿ”, e.g. צִוֶּה . It occurs when the last character is ע,פ,ה or when the last-but-one character is ע,פ,ה,א .

Other segolate words are those with a “ֿ” under the last-but-one character, which occurs when the last-but-one character is י, e.g. מְיִל. There are few other words which are segolates with “ֿ” when the last character is י (See 7).

A segolate vowel, which is never accented, is not part of the structure of the word. Rather, it is part of its realization. As such it is not phonemic, and there is no need to signify it in the converted Phonemic script, thus גֹלֶם – golm, צֹהַר - cohṛ, חַיִל – hayl, as well as rakkebt - (רַכְבֶּת), maḥbert (מַחְבֶּרֶת) etc. but see 5.3.2 for simplified version.

4.3.5 ‘Furtive Pattah’ is the same vowel a (“ֿ”) when it is written in the pointed script under ה,ה or ע which is a last character. It signifies an unaccented a, which is inserted before the last ḥ, h or ‘ when the preceding accented vowel is other than a, as in רוּחַ [=ruaḥ]. Other examples: צוּמַח [=comeaḥ], חוּחַ [=hoaḥ]. This vowel is not phonemic. However, for simplification it may be written for regular use, but for strict linguistic purposes it is better omitted: ruḥ, comeḥ, macmiḥ, as well as tameh, maddu’ etc. but see 5.3.1.

4.3.6 ‘Compound Shewa’ is a combination of shewa and one of the following vowels: “ֿֿ, ֿֿ, ֿֿ”. It replaces a shewa when the preceding consonant is either א, ה, פ or ע. Each of these combinations is written as one sign (ֿֿ, ֿֿ, ֿֿ respectively). In non-pointed script ֿֿ and ֿֿ are not represented but ֿֿ is written as ׀ (see table 2 above).

A few examples:

אמונה = 'emuna
חרקים = ħaraqim
חרוגים = harugim
עונשים = 'onasim

Compound shewa is converted into its parallel vowel (see table 2), but if the text is to be processed for linguistic purposes, the above mentioned words may be written without these vowels: 'muna, ħaraqim, hrugim, 'nasim.

5. SIMPLIFIED VERSION

5.1 For non-linguistic use a simplified version may be used.

5.2 Gemination of consonants may be represented with one letter only, e.g. mitnahel (מתנהל), instead of mitnahhel) ħuqim (חוקים), instead of ħuqqim), but gemination of b,k,p after a vowel should always be represented by a double letter, e.g. tikkon = [tikon] (תיכון), but tikon = [tixon] (תיכון), both words are equally spelled in regular Hebrew script. See ANNEX A.2.1(a). However, b,k,p in a gemination at the end of a word may also be written with one character, e.g. לבב –lebb or leb. See 4.2.3.

5.3 Non-phonemic vowels may be written as follows:

5.3.1 A 'furtive pataḥ', e.g. ruah (רוח), instead of ruḥ, macmiaḥ (מצמיה) instead of macmiḥ. See 4.3.5.

5.3.2 A segolate vowel, such as golem (גולם), instead of golm), rakkebet (רכבת), instead of rakkebt), cohar (צוהר), instead of cohr). See 4.3.4.

5.3.3 Whenever a non-phonemic [e] is inserted in a word (see Annex A.4) a user who is a Hebrew speaker may add it for regular use. Practically, e may be inserted in every cluster of the types mentioned in A.4 whether or not it complies with the consonants mentioned there.

5.3.4 A vowel which is a compound shewa in the pointed script is written with the parallel vowel character, i.e. a,e,o for א, ע, ו, respectively. This is recommended for regular use even in the main version, see 4.3.6.

6. RE-CONVERSION INTO NON-POINTED SCRIPT.

6.1 Consonants

6.1.1 Each consonant is converted into its Hebrew correspondent letter, according to Table 1.

6.1.2 Gemination of consonants in the phonemic script is re-converted into one character in the Hebrew script according to Table 1. The second character of a gemination is not re-converted.

6.1.3 For y between a consonant and a vowel or between two vowels which neither of them is o or u, another ם is added, unless a hyphen separates the preceding vowel from the y, and unless the y precedes an a or e which is the last character in the word.

6.1.4 For w between vowels which neither of them is o or u, another ן is added, even if a hyphen separates the preceding vowel from the ן.

6.2 Vowels

6.2.1 Vowels a and e as last characters in a word are re-converted by ם. but when any of them is in the middle of the word, it is omitted in converting into Hebrew characters. a is also omitted in suffix -ka (inflection of nouns, such as kalbka – כלבך) and in suffix -ta (verbs in past tense, 2nd p.,m.s., such as gamarta - גמרת).

6.2.2 Vowels o,u are always converted as ן, but see exceptions in section 7.

6.2.3 Vowel **i** is converted by **י** in the following cases:

- (a) If **i** is the last vowel in the word.
(b) When one consonant, or a gemination of consonants, both with a vowel following it, follows the **i**.
In other cases the **i** is omitted when converting into Hebrew characters.

7. EXCEPTIONS

For 4.3.4: Segolate words ending with the consonant **y**:

אופי → 'opy אפי
דופי → dopy דיפי
יופי → yopy יופי
פלי → pely פלי
פתי → pety פתי
שופי → šopy שופי

For 6.2.2: A few words ending with **o**:

- (a) פה → po
איפה → 'eipo
כה → ko
פרעה -- > par'o

(b) Words with mute **'** :

זאת → zo't
לא → lo'
צאן → co'n
ראש → ro'sh
שמאל → jmo'l

(c) inflection of some verbs with **'**:

תאהב → to'hab
יאבד → yo'bad
תאמר → to'mri

and other forms in future tense of the verbs 'aba, 'abad, 'ahab, 'axaz, 'amar, 'apa.

ANNEX A
(Informative)

Pronunciation rules for “general Israeli” Hebrew

A.1 These rules may be needed for oral communication with Hebrew speakers or with librarians.

A.2 b,k,p

A.2.1 The characters b,k,p are pronounced v,x,f respectively, when:

(a) The character comes after a vowel and is not a double character, e.g. mibcar → [mivcar] (but nappax – [napax], see section A.3)

(b) The character is doubled at the end of the word e.g. lebb → [lev]

(c) The character comes after a consonant-character which is at the beginning of a syllable, e.g. bki → [bxi]

(d) The word is in an inflectional form where a basic vowel that complies with condition (a) is omitted (not realized), e.g. malkei (מלכי) pronounced [malxei] which is an inflection of mlakim (מלכים) = [mlaxim], garpu (גרפו) pronounced [garfu] which is an inflection of garap (גרפ) = [garaf].

A.2.2 Except for the above conditions, the characters p,k,b are pronounced [p,k,b], respectively.

A.3 Gemination is pronounced as a single consonant (like in English and French and not like in Italian or Arabic), e.g. naggar → [nagar]

A.4 Cluster of consonants

A.4.1 A vowel [e] is inserted between a cluster of two consonants at the beginning of a word (in pointed script there is a shewa under the first character):

(a) If the first character of the cluster is **y,m,n** or **r**

e.g. ילדים yladim → [yeladim]
מנהל mnahhel → [menahel]
נמלה nmala → [nemala]
רצפים rcapim → [recapim]

(b) If the second character of the cluster is **’**, **h** or **’**,

e.g. בהמה bhema → [behema]
תארים t’arim → [te’arim]
בעירה b’ira → [be’ira]
sometimes also if the second character is **h**:
דחיקה dhiqa → [dehiqa]

(c) In several other cases, according to the pronunciation of the individual.

A.4.2 A vowel [e] is inserted between the second and third consonant when three consonants come in succession in the middle of the word.

e.g. : מפתחות mapthot – [=maftehot]
עכברי 'akbrei – [=‘akberei]

It does not apply to cases where the first consonant of the cluster of the three consonants is **n**, e.g. סנדלר sandlar --> [sandlar]

A.5 The phonetic values of the conversion signs.

A.5.1 Most consonant signs represent a pronunciation similar to many languages. The following signs, however, should be noticed:

| | |
|---------|--|
| ' | -(represents ʔ) – glottal stop |
| h or x | -(represents π) – like ch in Scottish ch, Russian x, Spanish j. |
| ‡ or @ | -(represents ʈ) – equivalent to t. |
| ‘ or & | -(represents ɮ) – glottal continuant voiced phone like Arabic (0639) |
| c | -(represents ʒ) – German z, English ts in cats. |
| š or \$ | -(represents ʃ) – English sh, French ch. |
| ś | (represents ʃ) – equivalent of s. |
| ġ | - (represents ʒ) – as g in English giant. |
| ž | - (represents ʒ) – as j in French. |
| č | - (represents ʃ) – as ch in English china |

A.5.2 The vowel signs represent five vowels pronounced like the Italian vowels. **ei** is also pronounced [e]. The addition of letter **i** is aimed at having it correspond to the spelling with Hebrew characters (See table 2).

ANNEX B (Normative)

B.1 Standards on conversion of systems of writing

This part of ISO 259 is one of a series of International Standards, dealing with the conversion of systems of writing. The aim of this part of ISO 259 and others in the series is to provide a means for international communication of written messages in a form which permits the automatic transmission and reconstitution of these by men or machines. The system of conversion, in this case, must be univocal and entirely reversible.

This means that no consideration should be given to phonetic and aesthetic matters nor to certain national customs: all these considerations are, indeed, ignored by the machine performing the function.

The adoption of this part of ISO 259 for international communication leaves every country free to adopt for its own use a national standard which may be different, on condition that it be compatible with the International Standard. The system proposed herein should make this possible and be acceptable to international use if the graphisms it creates are such that they may be converted automatically into the graphisms used in any strict national system.

This part of ISO 259 may be used by anyone who has a clear understanding of the system and is certain that it can be applied without ambiguity. The result obtained will not give a correct pronunciation of the original text in a person's own language but it will serve as a means of finding automatically the original graphism and thus allow anyone who has a knowledge of the original language to pronounce it correctly. Similarly, one can only pronounce correctly a text written in, for example, English or Polish, if one has some knowledge of English or Polish.

The adoption of national standards compatible with this part of ISO 259 will permit the representation, in an international publication, of the morphemes of each language according to the customs of the country where it is spoken. It will be possible to simplify this representation in order to take into account the number of the character sets available on different kinds of machines.

B.2 General principles of conversion of writing systems

B.2.1 Definition and methods

B.2.1.1 The words in a language, which are written according to a given script (the converted system), sometimes have to be rendered according to a different system (the conversion system) normally used for a different language. The procedure is often used for historical or geographical texts, cartographical documents and in particular bibliographical work where characters must be converted from different writing systems into a single alphabet to allow for alphabetical intercalation in bibliographies, catalogues, indexes, toponymic lists, etc.. It is indispensable in that it permits the univocal transmission of a written message between two countries using different writing systems or exchanging a message the writing of which is different from their own.

It hereby permits transmission by manual, mechanical as well as electronic means.

The two basic methods of conversion of a system of writing are transliteration and transcription. For some scripts a third method may be used: a phonemic conversion (B.2.1.5). This document, ISO 259-3, is a phonemic conversion.

B.2.1.2 Transliteration is the process which consists of representing the characters¹ of an alphabetical or syllabic system of writing by the characters of a conversion alphabet, this being the easiest way to ensure the complete and unambiguous reversibility of the conversion alphabet in the converted system.

In exceptional cases, e.g. when the number of characters used in the conversion system is smaller than the number of characters of the converted system, it is necessary to use digraphs or diacritical marks. In this case, one must avoid as far as possible arbitrary choice and the use of purely conventional marks, and try to maintain a certain phonetic logic in order to give the system a wide acceptance.

¹ A character is an element of an alphabetical or other type of writing system that graphically represents a phoneme, a syllable, a word or even a prosodical characteristic of a given language. It is used either alone (e.g. a letter, a syllabic sign, an ideographical character, a digit, a punctuation mark) or in combination (e.g. an accent, a diacritical mark). A letter having an accent or a diacritical mark, for example a,e,o, is therefore a character in the same way as a basic letter.

However, it must be accepted that the graphism obtained may not always be correctly pronounced according to the phonetic habits of the language (or of all the languages) which usually use (s) the conversion alphabet. On the other hand this graphism must be such that the reader who has some knowledge of the converted language may mentally restore unequivocally the original graphism and thus pronounce it.

B.2.1.3 Retransliteration is the process whereby the characters of a conversion alphabet are transformed back into those of the converted writing system. It is the exact opposite of the transliteration process in that the rules of a transliteration system are applied in reverse in order to reconvert the transliterated word to its original form.

B.2.1.4 Transcription is the process whereby the sounds of a given language are noted by the system of signs of a conversion language.

A transcription system is of necessity based on the orthographical conventions of the conversion language. Transcription is not strictly reversible.

Transcription may be used for the conversion of all writing systems. It is the only method that can be used for systems that are not entirely alphabetical or syllabic and for all ideophonographical systems of writing like Chinese.

B.2.1.5 Phonemic conversion

In scripts where texts often do not represent certain essential phonemes such as some vowels in Hebrew, Arabic or Persian, or such as gemination of consonants (which is not represented by double

letters in these languages), a third approach may be taken, in which we add some graphemes into the target script to supply the missing phonemes of the source script.

This approach is similar to transliteration, but it stems from the phonemic structure of the words rather than from their written form in the source script.

B.2.1.6 To carry out romanization, the conversion of non-Latin writing systems to the Latin alphabet, either transliteration or transcription or phonemic approach may be used, depending on the nature of the converted system.

B.2.2 A conversion system proposed for international use may call for compromise and the sacrifice of certain national customs. It is therefore necessary for each community of users to accept concessions, fully abstaining in every case from imposing as a matter of course solutions that are actually justified only by national practice (for example as regards pronunciation, orthography, etc.).

When a country uses two systems univocally convertible one into the other to write its own language, the system of transliteration thus implemented must be taken a priori as a basis for the international standardized system, as far as it is compatible with the other principles exposed hereafter.

B.2.3 When necessary, the conversion systems should specify an equivalent for each character, not only the letters but also the punctuation marks, numbers, etc. They should similarly take into account the arrangement of the sequence of characters that make up the text, for example the direction of the script, and specify the way of distinguishing words and of using separation signs, following as closely as possible the customs of the language(s) which use the converted writing systems.

B.2.4. When romanizing a script which does not have upper-case characters, it is usual to capitalize some words, following national usage.

B.3 Principles of conversion for alphabetical writing systems

B.3.1 The conversion may be made at various levels

The first level is that of completely reversible **stringent transliteration** which is necessary to attain, in full, the aim given in clause 0.2.1.2. This conversion applies all principles of transliteration without exception. However, whenever it is useful to distinguish the end or beginning of a syllable (a morpheme or a word) variants may be used. The conventional systems of stringent transliteration shall be applied as such without any change to meet national or regional customs as regards pronunciation or orthography. They permit the univocal international transmission of messages by mechanical or electronic means.

To permit an international unequivocal communication, International Standards on transliteration shall apply by priority the principle of stringent conversion. They can then be used as a basis for the establishment of rules for simplified conversion and for preparation of national standards.

The second level is that of **simplified conversion**. The simplification may be made necessary, for example, by the use of machines that do not accept all the alphabet characters required for stringent conversion. The method of conversion may allow national or regional variants, which may not permit complete reversibility. The simplified conversion may be the subject of International Standards or agreements.

The third level is that of **popular conversion** which, for example, should enable the same foreign names to be written in a uniform manner in the newspapers of a given country. It is obliged to take into account phonetic or graphic practices and therefore can only be national.

B.3.2 In cases where the same characters appear in one alphabet used with some differences by different languages, these characters would be transliterated in the same way, irrespective of the language they belong to.

B.3.3 If the converted alphabet gives a different form to the same character according to its place in the word (as is the case for example in the Arabic, Hebrew and Greek alphabets), the conversion alphabet will use only one character of constant form.