

**Title:** Draft Disposition of Comments for CD 15924  
**Source:** Michael Everson, Project Editor  
**Date:** 1999-06-16

general for the standards of “Script” itself. With the current definition, arbitrary subset of a script can be “Script”. For example, LATIN CAPITAL LETTERS can be a script.



Comments on the Austrian Yes vote:

These comments are all of an editorial nature with regard to grammatical and orthographic issues in the French version of the standard, and so have not been reproduced here. They are all accepted in principle (but in principle only) pending verification. Extensive verification of French terms will take place before the DIS ballot is sent out.



Comments on the Japanese No vote:

1. It is unclear if there is any rules or procedures for a script to be entered in the standard. The standard, or relating documents, should be able to answer clearly why the “Klingon pIQaD” has entry in this standard, and “Lolo” does not.

**Accepted in principle. Rules for maintenance and registration will be added co A.3.3. In the case specified, Klingon is included because it is unambiguous; Lolo may be a variant of Yi and further research is required to determine if the differentiation is necessary.**

For example, in Japan there is long controversy on whether Japanese “Kanji” and Chinese “Hanzi” is the same script or not (though, historical identity is evident, modern orthography is significantly different in both countries). Without well-established academic authority, or obvious and fair procedures, we can’t even establish our national position.

**Noted. It has been intended to use the principles of script identification and unification used by ISO/IEC JTC1/SC2/WG2 with regard to the entities coded in this standard. With regard to both the UCS and ISO 15924, “Kanji” and “Hanzi” could be indicated with the use of a country code identifier (Han/JP and Han/CN), or a language code identifier, etc. Apparently this kind of thing is done in Internet language tagging.**

We should be careful on linguistic issues because it easily causes nationalistic argument.

**Noted.**

2. Definition of “Script”. With regard to the Maintenance rules, definition of the “Script” should be more specific. The definition in ISO 10646 may be good for the character code standards, but is too

**Accepted in principle. An attempt will be made to make the document refer more clearly to ISO/IEC 10646 practice, collections, etc. But the definition used in this standard should not differ from that found in the UCS. Anyway the editor tried several times but couldn’t come up with a better text. However, the editor disagrees with the comment, and does not think that the 10646 text implies “any arbitrary subset”, in part because it specifies “the written form of a language” and not “some text in the language”. In any case a note has been added to the definition.**

3. A3 code for “Japanese”: Jap Ja 930 (alias for Han + Hiragana + Katakana) should be Jpn Ja 930 as in ISO 690-2.

**Accepted. The editor thinks that ISO 690-2 is a typo for ISO 639-2; in ISO 639-2 “jpn” is used and should indeed be used here.**



Comments on the UK Abstention:

The UK is an O-member of ISO/TC 46/SC 2 and therefore **ABSTAINS** on Committee Draft ISO/CD 15924. However we do submit the comments below from individual UK experts to accompany the UK ballot form, dated 9 December 1998.

*Comments from John Clews, IDT/2/5 Chairman, ISO/TC 46/SC 2 Chairman*

1. Separate versions (English and French) are preferred to a single version with parallel text. Any editions in French have been produced by AFNOR in the past, to ensure maximum readability in French by French native speakers.

**Rejected in principle. Bilingual publications are perfectly acceptable in ISO and the ad-hoc which drew up this CD felt comfortable with bilingual presentation. The French text has been provided and corrected by speakers of both European and Canadian French. We can request AFNOR to review the text if you wish, but the editor notes that ISO does not require BSI or NSAI or CSA or ANSI to check the suitability of English text. Similar standards (ISO 639) are also bilingual.**

2. Section 4.4 is superfluous, and should be removed: Latin script is used internationally. Cyrillic or Greek variants of this are a matter for national standards bodies, if necessary, not an international standard. It refers to “the principles” of this standard, but without enumerating “the principles.”

**Rejected.** This text is identical to similar text in structurally identical standards, namely, ISO 639:1988 (clause 4.1) and ISO 639-2:1998 (clause 4.1).

3. In section 4.5, a prospective Registration Authority should be named, either in the draft standard, or in an accompanying document. There is nothing else about choice of a Registration agency, or the means for nominating a new registration agency in case the first one gives up.

**Accepted.** Everson Gunn Teoranta is proposed to be the RA, and a modified version of Annex A as found in ISO 639 will be added. Procedures for nominating an RA are understood to be the usual ones followed by ISO/CS. ISO 639 doesn't make any specific reference to these procedures either. TC46/SC2 will be responsible for finding a new RA should the appointed RA relinquish its responsibility, in accordance with ISO procedures.

4. Section 4.6.2 is irrelevant to the use of Script codes, and should be removed.

**Rejected.** Catalogues may represent information in the original script, whether on paper or in electronic format, or they may do so in transliteration, in which case a script code is useful for identifying the script used in a publication.

5. Section 4.6.3 is incomplete, and there is no indication of what is intended. In addition, subdivision a) is superfluous: there are no other subdivisions.

**Accepted in principle. It will be corrected and completed.**

6. In tables 1-5, ISO/TC46/SC2 should decide on one single code. A 3-letter code is to be preferred, as it gives greater chances for matching similar language lists, particularly in ISO 639-2, and matching names of languages in English (and also in other languages).

**Rejected.** Provision of 3 codes is intended to facilitate users, who may, for example, have fixed-width fields in their databases (e.g. allowing only 2 characters for language codes, and preferring the same for script codes). The choice should be given to users as to which format to use (just as with ISO 639). Numeric codes for language codes are provided to facilitate users who may not have Latin characters available.

7. Many of the codes lack predictability: it will be more widely used if language codes are predictable.

**Rejected.** The commentator has not defined predictability in any way and has not indicated why it should be preferred. An English-speaker might predict eng to be used for English, but a French speaker would

predict ang (< *anglais*), and an Irish speaker might predict brl (< *Béarla*). None of them would predict bod for Tibetan, though this is used in ISO 639-2/B for that language; in ISO 639 bo is used. Clause 4.1 explains the derivation of the codes in this standard.

It is recommended that ISO/TC46/SC2 check a list of names of scripts, to see how many can be identical to the first three letters of names of languages, in the three official languages of ISO, and adjust codes accordingly – with those scripts in heaviest use taking the 3-letter code that matches the first three letters of the name of that script, in cases where there are any matches on the first three letters.

**Rejected.** Using the three official languages of ISO are English, French, and Russian would produce ang for English (English, anglais, английский) which would introduce incompatibility with ISO 639-2 practice, where eng is used for English and ang for Anglo-Saxon or Old English.

ISO/TC46/SC2 should also make a comparison with codes used in ISO 639-2, and adjust codes accordingly.

**Noted.** Clause 4.1 specifies that 639-2 codes are used where possible. Of course an independent check could be useful to the editor (who is satisfied with the codes as they are).

8. The current tables 1-5 should be preceded by a “quick look up” table that lists only the approximately two dozen scripts which are in use in official languages today: this would be what was required by most users, and would be simpler to use than the larger tables, which should ideally follow such a “quick look up” table.

**Rejected.** 15924's sister standards, 639, 3166, and 4217 do not make this kind of distinction. All scripts should be considered equal, and the needs of users should not be guessed at with a quantitative approach.

9. Otherwise, this CD has a lot to recommend it, and I would hope that it is progressed as an international standard, as long as these comments can be taken account of.

**Noted.**



*Comments from Anthony P. Stone, IDT/2/5 Member, Project Leader, ISO/TC46/SC2/WG12*

Although not a librarian, I believe the idea will be useful in library systems.

The 3-letter format does not usually add much to the intuitive meaning of the 2-letter format, and I think it

would be confusing to have both (with 2 letters one can code 676 scripts). The 3-digit format has a useful structure and perhaps computer experts should decide whether this format would be important in data processing.

I would prefer to see only the 2-letter and 3-digit formats included. I am not strongly opposed to 3-letter codes if there is a special use for them, but am more opposed to the inclusion of both 2-letter and 3-letter codes.

**Rejected. Note that this is in variance to the comments of another UK expert, so it is clear that the choice of 2-letter, 3-letter, and 3-digit numeric codes should be maintained.**

p.ii, Note 1. If a script variant can be determined by the set of letters in its alphabet, then I should think it would be important to include the variants of Arabic script used for Urdu (Au ?) and Persian (Ap ?).

**Rejected. Language selection cannot always make this distinction. For instance, Laf and Lat both have been used to write German, Danish, Swedish, Estonian, and Latvian.**



*Comments from Mrs I. Tahan, IDT/2/5 Member, The British Library, OIOC - Hebrew Section*

Comments on Hebrew script codes as presented in the above document:

- The best format of script code is, in my view, the three letter code Heb. It's a code easily remembered and recognised, more so than the "wanting" He. The three-digit code 125, would probably be suitable in various computer applications, but less so in printed bibliographies, catalogues, card records.

**The numeric codes are available for users who may not have the Latin alphabet available to them or who may not be able to adapt the standard because their language is not alphabetic (e.g. Chinese or Canadian Syllabics).**

- The inclusion of all three formats in any one application would be most confusing. The best method would be using specific script codes for specific applications.

**An application should state which format or formats it supports.**

- When asking us whether this draft standard would be useful to library systems, what does the latter term refer to? Does the term imply libraries, automated databases, printed bibliographies? Please clarify.

**"Library systems" implies electronic and printed formats.**

Question: where does Yiddish fit in? It uses the Hebrew script but is in fact a language in its own, combining an old German dialect with Hebrew-Aramaic, Romance and Slavic words. It's been used by Ashkenazi Jews for about 1000 years.

**Yiddish (yi, yid) uses the Hebrew (He, Heb) script, though in a book the editor has on his shelf, Yiddish is written in the Latin script throughout. To indicate this would be a possible use of this standard.**



*Comments on the late Korean comments:*

Comment 1. Definition of Code and Code element (or codewords). In this CD, usage of the term code is different from the usage in other documents (e.g., ISO 3166). In other documents, code is a set of code elements or codewords. However, in this CD, code means code element in other documents.

Suggestion: Define code as a set of code elements (as in ISO 3166) and define code element or codeword to refer to an individual value. If we follow this new definition, we need to modify many places. Some examples are shown below:

- 1) In the title, we change code to codes since we have three codes in this CD: 2-letter, 3-letter, and numeric codes.
- 2) In page 2, left column, line 7: "... addition of codes for the following script variants" should be changed to "... addition of code elements for the following script variants"
- 3) In the titles of sections 1, 2, and 3 (and in many other places), script codes should be changed to script code.

**Rejected. Definitions follow those in the more closely-related standard, ISO 639-2:1998, which are preferable in this context to those in ISO 3166. Note that French differentiates between *code* and *codet*, though English has only one term for these. In the title it is the French *code* which is used.**

Comment 2. Numeric code elements [page 3] Numeric script code has ranges for types of scripts. If I remember correctly, at TC46/SC2 London meeting in 1998, we decided to sort script names in English and then assign numeric code elements increasingly, since classification of scripts is controversial. We would like to know why this decision has not been reflected in this CD.

**The editor does not recall an agreement to sort names in English in London, and would certainly have opposed it on grounds of ISO linguistic policy. In any case a different decision was taken at the Athens meeting, and reconfirmed at the Paris meeting. The statement that "classification of scripts is controversial" has not been unsupported by argument or evidence. Generally speaking the classification follows**

**that which is generally agreed upon by experts (cf. the following references in the bibliography: Gaur, Robinson, Haarmann).**

Suggestion: Sort script names in English and then assign numeric code element increasingly, since classification of scripts is controversial.

**Rejected. Sorting script names in English is not acceptable in a bilingual standard. (Why not sort them in French?) Support must be given for the suggestion that the classification is “controversial”. In Athens and Paris the value of mnemonicity in a classified arrangement was affirmed. The Project Leader will review the order of scripts for accuracy.**

Comment 3. Principle of assigning values of numeric code elements [page 3] We had difficulty finding any principle of assigning values of numeric code elements within each range.

Suggestion: If there is no principle of assigning values of numeric code elements within each range, set up a principle, state this principle in the standard, and follow this principle in assigning numeric code elements.

**Accepted in principle. A general principle (not a strongly specific principle) was followed and text explaining it will be added to the draft.**

Comment 4. The first paragraph of subsection 4.3. [page 3] We have difficulty understanding this paragraph: “The two-letter ... Guj/Gu).”

Suggestion: Clarify this paragraph. (“and where” may need to be changed to “where” (?))

**Accepted in principle. The paragraph will be reviewed for clarity.**

Comment 5. Script names in French [Table 1, page 6 and in other tables] Script names in French starts with a lower-case letter. We wonder if there is a reason for this. One person I contacted told me that proper nouns would start with an upper-case letter whereas an adjective would start with a lower-case letter.

Suggestion: Review if the first letter in French could be written in lower-case letter.

**Accepted. We have reviewed it and determined that the Korean comment is incorrect. Lower-case is used for (most of) these names in French.**

Comment 6. Script name Hangeul vs. Hangul [page 5, section 4.3; Table 1, page 6 and other tables]

Since we have ISO TR 11941 for transliteration of Hangeul, we would like to transliterate Hangeul into Latin using that TR. We clearly mentioned this fact to the editor

(?) of this CD at the London meeting in 1998.

We have already used the transliteration in ISO TR 11941 for ISO/IEC 10646 and ISO 3166-2.

**Hangul is used as the name of the script in ISO/IEC 10646.**

In those standards actual transliteration is performed. The list of names here is the usual name in English and French, such as *Cyrillic* and *cyrillique* not a transliteration like *kirillica* etc.

Suggestion. Change “Hangul” to “Hangeul” in English.

**Accepted in principle. The normal, ordinary name for the script in English is “Hangul” or “Hangŭl” and this will be used. The names of scripts are not words being transliterated; they are names. However in light of the Korean comment the transliteration will be given in parentheses as it may serve some users of the standard. See also comments on French below.**

Comment 7. Script name hangŭl in French [page 5, section 4.3; Table 1, page 6 and other tables] Since u with a circumflex above is not used in French words, ordinary French people will have difficulty pronouncing this word. Furthermore, since there is no French phoneme corresponding to eu as in Hangeul, it is not possible to write down Hangeul ‘correctly’ in French.

**“Correctly”? Anyway it is done conventionally, in a long-established tradition.**

Considering this fact, we would like to use Hangeul both as script names both in English and in French. (Of course, we understand that some French people would pronounce hangeul as something like hanjeul.)

Suggestion. Change “hangul (with a circumflex above u)” to “hangeul” in French.

**Rejected. The traditional name for the script in French is “hangŭl”. All French people (not some of them) would pronounce “hangeul” as “hanjul”. The transliterated name can be added in parentheses as for English.**

Comment 8. Hangeul is not a syllabic script [Table 1, page 6 and other tables] The numeric code element for Hangeul is 420, which indicates that Hangeul is a syllabic script. Hangeul symbols can be decomposed into letters corresponding to phonemes. As we understand it, therefore, Hangeul IS an alphabetic script, NOT a syllabic script.

**It is a pity that it is encoded as a syllabary in ISO/IEC 10646, isn’t it? That is the reason this classification was chosen. Of course, Hangŭl is an alphabet. It is one of the world’s best alphabets.**

As we said in item 2 above, classification of scripts is controversial and we can see one concrete example here.

**This is the only “controversial” example shown to date, and it is not controversial, it is just an error.**

In contrast, Hiragana (410), Katakana (411), and Yi (460) are good examples of syllable scripts, since symbols in these scripts cannot be further decomposed into letters corresponding to phonemes.

**Correct, of course.**

Suggestion: 1) Follow the suggestion mentioned in item 2 above and do not try to classify scripts.

**Rejected.**

2) In case classification of scripts remains in this standard, change the category of Hangeul from syllabic scripts (400-499) to left-to-right alphabetic scripts (200-299).

**Accepted.**

Comment 9. alias [Table 1, page 6 and other tables] We could not find the definition of alias in this CD.

Suggestion: Discuss whether we need to include aliases in this standard and, if we need it, then include its definition.

**Accepted. A definition will be added.**

Comment 10. Kh, Khn - alias for Hangeul + Han [Table 1, page 6 and other tables]

We do not seem to have proposed this alias. Furthermore, if we should include Han, we may have to include Latin additionally.

Suggestion: 1) Delete “Kh, Khn - alias for Hangeul + Han” in Tables 1 through 5.

**Accepted. The alias for Hangŭl (Hangeul) + Han will be deleted.**

2) If alias for Hangeul + Han should remain in the standard, change “Kh, Khn” to “Hh, Hgh”.

**Not necessary.**

Comment 11. Gugyeol - Annex A: scripts under consideration for future editions of ISO 15924 [page 16]. Gugyeol has been under investigation for inclusion in ISO/IEC 10646, although no conclusion has been reached yet.

Suggestion: Add “Gugyeol” to Annex A.

**Accepted.**

Comment 12. [page 2, left column, bottom line]

Suggestion: In NOTE, change “revariding” to “regarding”.

**Accepted.**