

**COMMISSION  
DES COMMUNAUTÉS  
EUROPÉENNES**

*Bruxelles le, 30 juillet 1992*  
**VERSION 4**

**SERVICE DE TRADUCTION**  
Informatique

SdT-02 (92) D/466



# MULTILINGUALISM

*The needs of the Institutions of the European Community*

this page is left blanc

**TABLE OF CONTENTS**

0. INTRODUCTION

1. LANGUAGES

2. CHARACTER RÉPERTOIRE

3. ORDERING

4. CODING

5. KEYBOARDS

**ANNEXES**

0. DEFINITIONS

1. LANGUAGES

2. CHARACTER RÉPERTOIRE

3. ADDITIONAL INFORMATION CONCERNING ORDERING

4. LIST OF KEYBOARDS

**REFERENCES**

this page is left blanc

## **0. INTRODUCTION**

The Institutions of the European Community produce documents in all 9 official languages of the Community (French, English, German, Italian, Dutch, Danish, Greek, Spanish and Portuguese). The need to handle all these languages at the same time is a political obligation which stems from the Treaties and cannot be questioned.

The creation of the European Economic Space which links the European Economic Community with the countries of the European Free Trade Association (EFTA) together with the continuing improvement in collaboration with the countries of Central and Eastern Europe oblige the European Institutions to plan for the regular production of documents in European languages other than the 9 official ones on a medium-term basis (i.e. 1 to 2 years). Some Services of the Commission (DG I, Translation Service, Office for Official Publications etc.) as well as several other Institutions (Council of Ministers, Court of Justice of the European Communities) have already been called upon to produce documents in these languages.

Furthermore the Community is especially interested in the development of its regions (the Maastricht Treaty even provides for a new Institution called "Comité des Régions"). It follows that languages used by large populations within European States should not be left aside, especially when they happen to be official languages at a regional level.

It is evident that massive production, transmission, interchange, processing, archiving, storage and retrieval of multilingual documents, glossaries and dictionaries, cannot be undertaken without the extensive use of powerful and flexible computer tools. The Institutions must therefore, face the challenge of introducing "extended" multilingualism in their computer systems as soon as possible. But they must first define what needs to be done in order to achieve this goal.

The present document comprises:

- inventories of the **languages** that must be considered for support in the computer systems of the European Institutions, as well as the priorities according to which these languages are to be supported;
- an inventory of the **character repertoire** needed to support these languages;
- the principles that must be taken into account in order to ensure the (alphabetical) **ordering** of the characters of the repertoire, in a language-dependent way;
- the (digital or electronic) **coding** of the characters of the repertoire;
- the **keyboard configurations** for each language/country according to the habits of (national and international) users.

In general this document is to be considered as the user specification of extended multilingualism for the European Institutions. It is hoped that the ongoing standardisation work in the context of the European Standardisation bodies concerned (CEN/CENELEC, ETSI and EWOS), will cover the subjects treated here in an exhaustive and complete way.

The present document was discussed and approved in the context of a special working group created by the Comité de Coordination de la Standardisation (CCS - previously known as the Standardisation Implementation Committee - SIC), following a request by the Translation Service of the Commission. The working group was chaired by Mr. J. Piette (Informatics Directorate, Commission). All other Institutions were invited to participate, and some actively did so. All the Institutions were kept informed about the group's work. P. Alevantis (Translation Service, Commission) acted as rapporteur.



this page left blanc

## 1. LANGUAGES

Languages to be considered for support in the computer systems of the European Institutions can be classified for practical reasons in four main categories:

- a. official languages of the European Communities (12 Member States);
- b. official languages of the 12 Member States but not of the Communities;
- c. official languages of the States with which the Communities are engaged in developing close relations [states of the European Free Trade Association (EFTA), states of Central and Eastern Europe, etc.];
- d. other languages currently used in the states mentioned in a, b, and c, above.

Based on the above-mentioned classification as well as on information gathered (see TABLES 1A and 1B in Annex 1), an exhaustive inventory of languages has been compiled (see TABLE 1C in Annex 1). After exclusion of languages spoken by less than 1% of the total population of the country concerned, the final inventory of languages to be considered for support in the context of extended multilingualism was established. Thus, languages to be considered for support figure in alphabetical order in the TABLE 1D of Annex 1 with their corresponding two-letter codes according to ISO 639 (1988).

Furthermore, in order to best plan for the introduction of extended multilingualism in the European Institutions, priorities have been set according to the actual needs of the Services concerned. Priority 1 covers official languages of the EEC, of EFTA countries, of certain Eastern and Central European countries and of Russia. Priority 2 covers the rest of the languages of TABLE 1D.

Experience has proved that in practice, no single user is expected to be able to treat simultaneously all the languages mentioned in TABLE 1D. However, some applications, like electronic dictionaries, word-processors, online databases, electronic mail etc. must be able to treat any combination of, as well as all, these languages at the same time.

In order to assure that a computer system (hardware/software) supports a given language in the context of the multilingual environment, the following points should be carefully verified:

- a. support of the **characters** (i.e the character répertoire) needed to correctly reproduce that language on any peripheral device connected to the system (screen, printer, etc.) (see also, Chapter 2 Character Répertoire),
- b. correct support, for the specific language, of any **special functions** associated with the application(s) used with the system (e.g. conversion of upper to lower case letters, cut and paste, search and replace),
- c. availability of algorithms ensuring correct and non-ambiguous alphabetical **ordering** of letters, words or expressions written in the language (see also, Chapter 3 Ordering),
- d. availability of algorithms assuring correct **hyphenation** of words or expressions written in the language,
- e. availability of (sub) programs assuring correct **control of the spelling** of words or expressions written in the language.
- f. possibility to identify different languages in the context of multilingual environments through the use of ISO 8879-type markers (i.e. SGML-type markers); in that case the ISO 639 codes should be used for **language identification**.





## **2. CHARACTER RÉPERTOIRE**

In order to define the repertoire of the characters needed, the languages to be supported have been classified as follows:

- a. latin-script languages i.e. languages using the latin alphabet plus diacritical marks;
- b. greek-script languages;
- c. cyrillic-script languages;
- d. armenian-, hebrew-, arabic-, georgian-script languages.
- e. special characters (standard phonetic alphabet, numbers, punctuation marks, semi-graphics, mathematical and other symbols).

The character répertoire has been defined on the basis of the Universal Multiple-Octet Coded Character Set (UCS) (ISO 10646, see ref. 5)<sup>1</sup>.

TABLE 2A of Annex 2 contains an exhaustive inventory of all characters and character combinations for latin-script languages. On the basis of this table the necessary Rows of ISO 10646 have been selected to compile TABLE 2D and TABLE 2E.

The basic layout of the characters of the répertoire is the one given in the Tables. It is to be implemented in the context of specific fonts, in the devices visually representing the characters i.e. display units and printers. Special attention should be paid to avoid ambiguities in the representation of similar-looking letters of different scripts; this can be done either by differentiating the layout of similar-looking characters or by indicating the script and/or the language context in which the user is working.

In order to assure that a computer system supports the character répertoire, the following points should be carefully verified:

- a. support of all the characters of the repertoire at least for the following fonts (see ref. 11):
  - Monospace (Regular, Bold, Italic, Bold Italic) for font sizes 8, 10, 12, and 18 (Courier);
  - SanSerif (Regular, Bold, Italic, Bold Italic) for font sizes 6, 8, 10, 12, 14, 18, 20 and 24 (Helvetica);
  - Serif (Regular, Bold, Italic, Bold Italic) for font sizes 6, 8, 10, 12, 14, 18, 20 and 24 (TimesRoman);
- b. in the context of the same font, characters belonging to different scripts, should be graphically compatible from an aesthetic point of view (i.e. all scripts can harmoniously co-exist in the same document).
- c. in the context of the same font, similar-looking characters belonging to different scripts should be differentiated by being easily identified (i.e. possibility to recognise similar-looking characters as belonging to the latin, the greek or the cyrillic script).

---

<sup>1</sup> the ISO/IEC JTC1/SC2/WG2 introduced several modifications in the DIS 10646 during its last meeting in Seoul (29 June - 3 July 1992); as these modifications concern the Tables and Rows mentioned in the present document, the contents of TABLE 2D and TABLE 2E are to be considered as provisional

### 3. ORDERING

The simultaneous support of many languages, alphabets and scripts, makes it necessary to provide for specific arrangements for the ordering of characters. Thus, in several cases (multilingual/multiscript electronic glossaries, data bases, telephone directories, etc.) it should be possible to implement an ordering scheme that is more exhaustive and sophisticated compared to the one based on the simple collating sequence of the digital equivalents of the characters concerned.

The ordering scheme proposed in the present document draws inspiration from a method developed by Rolf Gavare (see ref. 7) which is also the basis of a corresponding Swedish national standard (see ref. 8); a similar method is implemented in the context of a Canadian national standard (see ref. 9). The main principle of the proposed scheme is that ordering of distinct entries shall be predictable and unambiguous - this will imply that identical results are to be obtained if the same material is ordered over again and no complication arises in the case of merging groups of entries from different sources.

Before ordering takes place, the list of **entries** (i.e. single words or groups of words) to be ordered according to the present specification, should undergo an application-dependent preparatory treatment which consists of two steps as follows:

- a. for each entry establish the **ranking word(s)**, (i.e. the element(s) of the entry that should be considered for ordering) as well as the order they are to be taken into account (see Annex 3 for hints and examples concerning that first preparatory step);
- b. for each ranking word construct the corresponding **ordinal form** (i.e. the explicit character sequences to be compared when the actual ordering is to be performed) (see Annex 3 for conventions applying to that second preparatory step).

Ordering should then proceed in steps as follows:

- Step 1 define the way the ordering is to be performed i.e. whether it will be a "**word-by-word**" ordering or a "**character-by-character**" ordering; in that respect, a parameter A can be defined accordingly (e.g. A=W or C); the default value of parameter A can be W for some languages and/or applications<sup>1</sup>;
- Step 2 define the basic ordering sequence of scripts i.e. the sequence of the following groups of characters as whole entities:

S for special characters;  
L for latin-script characters;  
G for greek-script characters;  
C for cyrillic-script languages;  
A for armenian, H for hebrew, R for arabic, O for georgian characters.

A parameter B, made up of an eight-letter combination representing the order of the corresponding scripts may be defined. In that case it is assumed that no transliteration has taken place; the default value of parameter B should be SLGCAHRO;

---

<sup>1</sup> word-by-word ordering may prove very useful especially for some languages like german, or swedish.

- Step 3 define the alphabet of the language according to which the ordering is to be performed within the latin-, cyrillic- or arabic-etc.-scripts. In that respect a parameter C can be set equal to the ISO 639 two-letter code for the language whose alphabet is to be used as the basis of the ordering; C can also be set equal to 00 for "**full multilingual ordering**" (according to the general rules described in the present document).
- Step 4 for each ordinal form define the "**sorting keys**" (i.e. continuous and delimited chains of coded numbers corresponding to the characters of the ordinal form), that can be (re)arranged on the basis of their sole numerical (or digital) value; sorting keys are produced on the basis of ordinal forms according to the following principles:
- a. for word-by-word ordering, when the sorting keys of the ordinal forms of two first-ranking words are identical, the sorting keys of the ordinal forms of the second ranking words are compared and so on; the boundaries of ranking words are accordingly respected in the word-by-word ordering; if two ordinal forms differ only as regards to the characters found after a common initial part (i.e. if one form is an initial substring of the other) the shortest one shall be ordered first;
  - b. although parameter B can be defined in a way that special characters can be ordered after any script, they are to be ordered first in order not to be neglected by the user.
  - c. when parameter C has a value different from 00, the basis of the ordering of the corresponding script should be the one defined in Annex 3 for the different languages (e.g. french, swedish, german); in that way some characters that are otherwise considered as combinations of letters and diacritical marks are considered as unique characters (and not as combinations e.g. æ, ä for swedish, ch for spanish etc.).
  - d. the ordering should be realized in such a way that only letters are taken into account; only when the compared ordinal forms cannot be ordered, are the other non-alphabetical characters i.e. punctuation marks, special signs and digits to be considered;
  - e. the ordering of letters is carried out in four levels: first the normalized basic letters of the alphabet are compared according to the alphabet defined on the basis of parameter C; then - if the compared normalised forms are equivalent - attention is to be paid, for each normalized form in turn, to other attributes in the following order:
    - Variant letters,
    - Diacritical marks,
    - Capitals,
    - eventually Non-Alphabetical characters.
- Tables accompanying Annex 3 contain the order to be applied for different alphabets (basic letters), as far as variant letters are concerned; diacritics and non-alphabetic characters are always ordered according to their (natural) collating sequence, while capital letters are always ordered after the corresponding small ones.
- f. sorting keys should be constituted in a way that words with the greatest literal similarity (also as far as diacritics, capitals etc. are concerned), reckoning from the beginning of words, will be arranged most closely together.

Annex 3 contains a benchmark for controlling the algorithm used for ordering; this benchmark consists of two lists. A first list of randomly-presented entries is to be introduced according to the order given (and at least once in a different order) and the algorithm is to produce in all cases an ordering according to a second list.

#### 4. CODING

The characters of the répertoire are to be coded according to the ISO 10646 international standard (see ref. 5)<sup>1</sup>. It is anticipated that the basic multilingual plane of this standard (16-bit bytes) is more than sufficient to cover the needs of the European Institutions for the years to come. The Tables needed to support the répertoire are given below for Priority 1 & 2 Languages.

It is evident however, that the ISO 10646 standard will need some time before it is fully implemented into real-life products. Thus, intermediate solutions have to be based on existing 8-bit standards (ISO 8859/.. series, ISO 6937/.. series) and extension techniques (ISO 2022, ISO 4873), while waiting for ISO 10646-compatible products. In the context of the implementation of such solutions the migration towards the ISO 10646 environment should be seriously considered.

##### **Priority 1 Languages**

Screen/Printer Fonts: for tables in italics and while waiting ISO 10646-compatible products, conformance with 8-bit coding standards (in parentheses)<sup>2</sup>

##### **Latin-script**

*TABLE 1-Row 00: ISO-646 IRV, TABLE 2-Row 00: LATIN-1 SUPPLEMENT (ISO 8859/1)*

*TABLE 3-Row 01: EXTENDED LATIN-A (ISO 8859/2,3,4,9 or SET with REGISTRATION No 154)*

TABLE 7-Row 02: SPACING MODIFIER LETTERS (ch)

##### **Greek-script**

*TABLE 9-Row 03: GREEK (ISO 8859/7)*

##### **Cyrillic-script**

*TABLE 10-Row 04: CYRILLIC (ISO 8859/5)*

##### **Other characters**

TABLE 33-Row 20: GENERAL PUNCTUATION

TABLE 34-Row 20: SUPERSCRIPTS AND SUBSCRIPTS

TABLE 35-Row 20: CURR. SYMBOLS, COMB. DIACR. MARKS FOR SYMBOLS

TABLE 38-Row 22: MATHEMATICAL OPERATORS

TABLE 39-Row 22: MATHEMATICAL OPERATORS

*TABLE 43-Row 25: BOX DRAWING (SET with REGISTRATION No 155)*

TABLE 44-Row 25: BLOCK ELEMENTS, GEOMETRIC SHAPES

Internal representation: while waiting for ISO 10646-compatible products, non-ambiguous internal representation.

##### **Priority 2 Languages**

##### **Latin-script**

TABLE 4-Row 01: EXTENDED LATIN-B, TABLE 5-Row 02: EXTENDED LATIN-B

##### **Greek-script**

TABLE 31-Row 1F: GREEK EXTENSIONS, TABLE 32-Row 1F: GREEK EXTENSIONS

##### **Cyrillic-script**

---

<sup>1</sup> the ISO/IEC JTC1/SC2/WG2 introduced several modifications in the DIS 10646 during its last meeting in Seoul (29 June - 3 July 1992); as these modifications concern the Tables and Rows mentioned in the present document, the contents of TABLE 2D and TABLE 2E are to be considered as provisional.

<sup>2</sup> i.e. support of SIC S11 plus special latin characters (Set 154) plus russian (ISO 8859/5).

TABLE 11-Row 04: CYRILLIC

**Armenian-script**

TABLE 12-Row 05: ARMENIAN

**Hebrew-script**

TABLE 13-Row 05: HEBREW

**Arabic-script**

TABLE 014-Row 06: ARABIC, TABLE 015-Row 06: ARABIC

**Georgian-script**

TABLE 28-Row 10: GEORGIAN

**Standard phonetic**

TABLE 6-Row 02: IPA EXTENSIONS

**Other characters**

TABLE 36-Row 21: LETTERLIKE SYMBOLS, NUMBER FORMS

TABLE 37-Row 21: ARROWS

TABLE 40-Row 23: MISCELLANEOUS TECHNICAL

TABLE 41-Row 24: CONTROL PICTURES, OPTICAL CHARACTER RECOGNITION

TABLE 42-Row 24: ENCLOSED ALPHANUMERICS

TABLE 45-Row 26: MISCELLANEOUS DINGBATS

TABLE 46-Row 27: DINGBATS, TABLE 47-Row 27: DINGBATS

In general, the European Institutions should be able to receive and transmit all the ISO 10646 characters, but the following considerations should be taken into account:

- a. to avoid the complications associated with the possibility offered by ISO 10646 to code accented characters in two different ways (single accented characters vs. combinations of characters and non-spacing accents), it has been decided not to implement TABLE 8-Row 03 COMBINING MODIFIER LETTERS in the environment of the European Institutions. Because outside organisations may decide to implement this Table for different reasons and in order to avoid problems in the internal environment (e.g. extra confusion/complexity in computation algorithms), all incoming occurrences of combinations involving characters belonging to the above-mentioned Table, should be converted to their non-combined equivalents before being considered for internal processing.
- b. characters belonging to the ISO 10646 standard but not contained in the Tables mentioned above, should not be «destroyed» when entering the internal environment; users should be informed of their existence through appropriate substitutions. Such substitutions should be implemented on the basis of established standards (e.g. various transliteration schemes) or should be considered application-dependent (e.g. introduction of fall-back characters like \*).

## **5. KEYBOARDS**

It has been recognised in the past that users of information technology systems in the European Institutions are entitled to use their national keyboard. However, in several occasions a fully multilingual keyboard was necessary. Two solutions have been implemented to solve the problem:

- the solution adopted by the Commission is based on the ISO 8884 international standard (see ref. 12) enlarged with the ELOT 1000 standard for greek (SIC S6 guide, see ref. 13);
- the other Institutions adopted the EUROPA keyboard (AZERTY and QWERTY versions), initially developed at the Commission.

The solution developed and recommended in the SIC S6 document, provides for the production of all Latin-script languages as the ISO 8884 standard contains all diacritics used in combined latin characters. Thus for the SIC S6 keyboard (which already supports greek), the support of extended multilingualism necessitates the addition of layers to produce the additional scripts of the répertoire (for Priority 1 Languages a layer corresponding to the cyrillic keyboard, for Priority 2 Languages layers corresponding to the armenian, hebrew, arabic, and georgian keyboards).

In order to assure that the keyboard used with a computer system supports multilingualism, the following points should be carefully verified:

- a. possibility for the user to define as the primary keyboard layer one of the national keyboards of the TABLE 5A (see Annex 5) or the ISO 8884 primary latin keyboard;
- b. possibility to switch in a user-friendly and ergonomically correct way to supplementary keyboard layers in order to produce the special latin characters of the repertoire that cannot be reproduced on the primary level, as well as the greek and cyrillic characters of the repertoire;
- c. possibility to identify at any given moment the active layer used (i.e. the script produced).

To implement the keyboard the following methods/technologies may be used: key engraving using different colors, LCD-keyboards, keyboardless interfaces (e.g. the user can design his/her own keyboard on a touch-sensitive screen, possibility for handwriting recognition).

this page left blanc



**ANNEXES**

this page is left blanc

**ANNEX 0 DEFINITIONS**

**Multilingualism** : the possibility to treat many languages at a time. In the case of the European Community the possibility to treat correctly and simultaneously the 9 official and working languages provided for in the Treaties.

**Restricted Multilingualism** : the possibility to treat only some of the 9 Community languages, e.g. only the latin languages (including special and accented latin characters) or only greek and english (not special latin characters).

**Extended Multilingualism** : the possibility to treat more than the 9 Community languages at a time, e.g. the 9 official languages simultaneously with the latin languages of Eastern Europe, and/or with languages using cyrillic characters etc.

**Character** : an elementary mark used to represent data, in the form of a graphic spatial arrangement of connected or adjacent strokes, such as a letter, a numeral, a punctuation mark, or a mathematical symbol.

**Repertoire** : a specified set of characters.

**Script** : A distinctive and complete set of characters used for the orthography of a particular group of languages.

**Alphabet** : a set of characters used to produce the written form of a language with a certain customary and language specific order, i.e. an ordered set of characters.

**Font** : a collection of character images having the same basic design (e.g. Helvetica Italic) together with information about these images both individually and collectively (descriptive and/or shape information, metrics, special attributes associated with each character etc.).

**Ordering** : the arrangement of characters or character combinations according to a specified sequence (e.g. in the case of words or word combinations according to an alphabetical ordering, in the case of numbers according to ascending or descending order, etc.).

this page is left blanc

**ANNEX 1 LANGUAGES**

**TABLE 1A: OFFICIAL LANGUAGES** (see ref. 1)

<b>1. <i>European Community</i></b>	
France	French
Germany	German
Italy	Italian
The Netherlands	Dutch
Belgium	French, Dutch, German
United Kingdom	English
Danemark	Danish
Ireland	English, Irish (Erse or Gaelic)
Luxembourg	French, German
Greece	Greek Modern
Spain	Spanish (Castillian)
Portugal	Portuguese
<b>2. <i>European Free Trade Association (EFTA) Countries</i></b>	
Norway	Norwegian
Sweden	Swedish
Switzerland	French, German, Italian
Austria	German
Finland	Finnish, Swedish
Iceland	Icelandic
<b>3. <i>East &amp; Central European Countries</i></b>	
Albania	Albanian
Bulgaria	Bulgarian
Czechoslovakia	Czech, Slovak
Hungary	Hungarian
Poland	Polish
Romania	Romanian
ex-Yugoslavia	Serbo-croatian, Macedonian <sup>1</sup> , Slovene
CIS (Commonwealth of Independent States)	Russian, Armenian, Azerbaijani, Estonian, Georgian, Kazakh, Kirgiz, Latvian, Lithuanian, Moldavian, Tadjik, Turkmenian, Ukrainian, Uzbek;
Belorussian,	
<b>4. <i>Other Countries</i></b>	
Malta	English, Maltese
Cyprus	Greek, Turkish
Turkey	Turkish
Andorra	Catalan
Israel	Arabic, Hebrew
Liechtenstein	German

---

<sup>1</sup> in its conclusions the European Council in Lisbon (26-27 June 1992) expressed its readiness to recognise the former Yougoslav Republic of Macedonia «under a name which does not include the term Macedonia»; as a result the term «macedonian» is used in the present document on a provisional basis (see ref. 14).

Multilingualism V.4 - page 22

Monaco	French
Greenland	Danish, Greenlandic
Faeroe Islands	Danish, Faeroese

**TABLE 1B: OTHER LANGUAGES<sup>1</sup>**

	<b>1. <i>European Community</i></b>
France	Basque*, Breton, Catalan* (Rousillonais), Corsican*, Alsatian, Occitan
Italy	Rhaetian, Sardinian
United Kingdom	Scots-Gaelic*, Welsh
Luxembourg	Luxembourgish
Netherlands	Frisian
Greece	Greek Classical <sup>2</sup>
Spain	Basque, Catalan, Galician
	<b>2. <i>European Free Trade Association (EFTA) Countries</i></b>
Switzerland	Romansh
Sweden, Norway, Finland	Lapp
	<b>3. <i>East &amp; Central European Countries</i></b>
Bulgaria	Romany
Czechoslovakia	Romany, Ukrainian*
Poland	Belorussian*, Ukrainian*
Romania	Romany*, Hebrew*, Tatar*, Ukrainian*
ex-Yugoslavia	Romany*, Vlach*
CIS	Bashkir* Buryat* Chechen* Chuvash* Dargin* Gagauz* Hebrew* Ingush* Kadardinian* Kara-Kalpak* Komi* Korean* Kumyk* Lezgian* Mari* Mordovian* Ossetian* Tatar Tuvinian* Udmurt* Ulghur* Yakut
	<b>4. <i>Other Countries</i></b>
Turkey	Arabic, Kurdish
Israel	Yiddish

---

<sup>1</sup> official languages of the same or other state are not mentioned if they are already mentioned in TABLE 1A; languages spoken by less than 1 % of the total population are marked with an asterisk\*

<sup>2</sup> for legislation before 1975 and for classical studies or literature.

**TABLE 1C: ALPHABETICAL INVENTORY OF LANGUAGES**(languages spoken by less than 1 % of the population are marked with an asterisk \*)<sup>1</sup>

<i>Language</i>	<i>ISO 639 (1988)</i>		KUMYK*	
ALBANIAN	sq		KURDISH	ku
ALSATIAN <sup>2</sup>			LAPP	
ARABIC		ar	LATVIAN	lv
ARMENIAN	hy		LEZGIAN*	
AZERBAIJANI		az	LITHUANIAN	lt
BASHKIR*		ba	LUXEMBOURGISH	
BASQUE	eu		MACEDONIAN <sup>3</sup>	mk
BELORUSSIAN		be	MALTESE	mt
BRETON	br		MARI*	
BULGARIAN		bg	MOLDAVIAN	mo
BURYAT*			MORDOVINIAN*	
CATALAN		ca	NORWEGIAN	no
CHECHEN*			OCCITAN	oc
CHUVASH*			OSSETIAN*	
CORSICAN		co	POLISH	pl
CZECH		cs	PORTUGUESE	pt
DANISH	da		RHAETIAN-ROMANSH	rm
DARGIN*			ROMANIAN	ro
DUTCH		nl	ROMANY	
ENGLISH	en		RUSSIAN	ru
ESTONIAN		et	SARDINIAN <sup>2</sup>	
FAEROESE			SCOTS-GAELIC	gd
FINNISH	fi		SERBO-CROATIAN	sh
FRENCH	fr		SLOVAK	sk
FRISIAN			SLOVENE	sl
GAGAUZ*			SPANISH	es
GALICIAN		gl	SWEDISH	sv
GEORGIAN	ka		TADZHIK	tg
GERMAN	de		TATAR	tt
GREEK CLASSICAL			TURKISH	tr
GREEK MODERN	el		TURKMENIAN	tk
GREENLANDIC		kl	TUVINIAN*	
HEBREW	iw		UDMURT*	
HUNGARIAN		hu	UKRAINIAN	uk
ICELANDIC	is		ULGHUR*	
INGUSH*			UZBEK	uz
IRISH	ga		VLACH*	
ITALIAN	it		WELSH	cy
KADARDINIAN*			YAKUT	
KARA-KALPAK*			YIDDISH	ji
KAZAKH	kk			
KIRGIZ		ky		
KOMI*				
<i>Language</i>	<i>ISO 639 (1988)</i>			
KOREAN*		ko		

<sup>1</sup> for several languages ISO codes have not yet been established.<sup>2</sup> Alsatian = repertoire for German, Sardinian = repertoire for Rhaetian-Romansh<sup>3</sup> term used on a provisional basis (see TABLE 1A and ref. 14).



**TABLE 1D: LANGUAGES TO BE CONSIDERED**

(with the exception of languages spoken by less than 1 % of the population )

(alphabetically)

<i>Language</i>	<i>ISO 639 (1988)</i>		<i>Language</i>	<i>ISO 639 (1988)</i>	
ALBANIAN	sq	L	KIRGIZ	ky	A
ALSATIAN		L	KURDISH	ku	A
ARABIC	ar	A	LATVIAN	lv	L
ARMENIAN	hy	A	LITHUANIAN	lt	L
AZERBAIJANI	az	C	LUXEMBOURGISH	L	
BASQUE	eu	L	MACEDONIAN <sup>1</sup>	mk	C
BELORUSSIAN	be	C	MALTESE	mt	L
BRETON	br	L	MOLDAVIAN	mo	C
BULGARIAN	bg	C	NORWEGIAN	no	L
CATALAN	ca	L	OCCITAN	oc	L
CORSICAN <sup>c</sup>	co	L	POLISH	pl	L
CZECH	cs	L	PORTUGUESE	pt	L
DANISH	da	L	RHAETIAN-ROMANSH	rm	L
DUTCH	nl	L	ROMANIAN	ro	L
ENGLISH	en	L	ROMANY		L ?
ESTONIAN	et	L	RUSSIAN	ru	C
FAEROESE			SARDINIAN	L	
FINNISH	fi	L	SCOTS-GAELIC	gd	L
FRENCH	fr	L	SERBO-CROATIAN	sh	L C
GALICIAN	gl	L	SLOVAK	sk	L
GEORGIAN	ka	A	SLOVENE	sl	L
GERMAN	de	L	SPANISH	es	L
GREEK CLASSICAL		G	SWEDISH	sv	L
GREEK MODERN	el	G	TADZHIK	tg	C
GREENLANDIC	kl	L	TATAR	tt	C
HEBREW	iw	A	TURKISH	tr	L
HUNGARIAN	hu	L	TURKMENIAN	tk	C
ICELANDIC	is	L	UKRAINIAN	uk	C
IRISH	ga	L	UZBEK	uz	C
ITALIAN	it	L	WELSH	cy	L
KAZAKH	kk	A	YAKUT		C
			YIDDISH	ji	A

L=Latin-script, G=Greek-script, C=Cyrillic-script, A=Arabic, Hebrew, Armenian, Georgian

<sup>1</sup> term used on a provisional basis (see TABLE 1A and ref. 14).

**TABLE 1E: LANGUAGES TO BE SUPPORTED-PRIORITIES**

(on the basis of actual needs)

**PRIORITY 1 LANGUAGES**

(EEC, EFTA, part of Central &amp; Eastern Europe, Russia)

<i>Language</i>	<i>ISO 639 (1988)</i>			<i>Language</i>	<i>ISO 639 (1988)</i>		
<b>Supported now</b>				<b>Supported now</b>			
ALBANIAN	sq	L	y	IRISH	ga	L	y
BULGARIAN		bg	C	ITALIAN	it	L	y
n				NORWEGIAN		no	L
CZECH		cs	L	y			
n				POLISH		pl	L
DANISH	da	L	y	n			
DUTCH		nl	L	PORTUGUESE		pt	L
y*				y			
ENGLISH	en	L	y	ROMANIAN	ro	L	n
FINNISH	fi	L	n	RUSSIAN	ru	C	n
FRENCH	fr	L	y*	SERBO-CROATIAN	sh	L C	n
GERMAN	de	L	y	SLOVAK	sk	L	n
GREEK MODERN	el	G	y	SLOVENE		sl	L
HUNGARIAN		hu	L	n			
n				SPANISH	es	L	n*
ICELANDIC	is	L	y	SWEDISH		sv	L
				y			

**PRIORITY 2 LANGUAGES**

<i>Language</i>	<i>ISO 639 (1988)</i>					
<b>Supported now</b>				KURDISH	ku	A
ALSATIAN			L	LATVIAN	lv	L
y						
ARABIC		ar	A			
ARMENIAN	hy	A				
AZERBAIJANI		az	C			
BASQUE	eu	L	y			
BELORUSSIAN		be	C			
BRETON	br	L	y			
CATALAN		ca	L			
y*						
CORSICAN		co	L			
y						
ESTONIAN		et	L			
FAEROESE						
GALICIAN		gl	L			
y						
GEORGIAN	ka	A				
GREEK CLASSICAL			G			
GREENLANDIC		kl	L			
HEBREW	iw	A				
KAZAKH	kk	A				
KIRGIZ		ky	A			

<i>Language</i>	<i>ISO 639 (1988)</i>		
<i>Supported now</i>			
LITHUANIAN	lt		L
LUXEMBOURGISH	L		y
MACEDONIAN <sup>1</sup>	mk		C
MALTESE	mt		L
MOLDAVIAN	mo		C
OCCITAN	oc		L
y			
RHAETIAN-ROMANSH	rm		L
y			
ROMANY			L ?
SARDINIAN	L		y
SCOTS-GAELIC	gd		L
y			
TADZHIK	tg		C
TATAR	tt		C
TURKISH	tr	L	
TURKMENIAN	tk		C
UKRAINIAN	uk		C
UZBEK	uz		C
WELSH	cy		L
y			
YAKUT			C
YIDDISH	ji	A	

L=Latin-script, G=Greek-script, C=Cyrillic-script, A=Arabic, Hebrew, Armenian, Georgian

<sup>1</sup> term used on a provisional basis (see TABLE 1A and ref. 14).

\* however some characters for these languages are not currently supported (e.g. ij for dutch, œ for french, ð for spanish).

**ANNEX 2 CHARACTER REPERTOIRE**

**TABLE 2A: Character répertoire for Latin-based European languages (by character) (see ref 2,3) (= cannot be reproduced)**

LANGUAGES	A	B	B	C	C	D	D	E	E	F	F	G	G	G	H	I	I	I	L	L	M	N	O	P	P	R	R	S	S	S	S	S	S	T	W	F	F				
Characters	L	A	R	A	Z	A	U	N	S	I	R	A	E	R	U	C	R	I	A	L	A	O	C	P	P	H	R	S	S	S	S	S	P	W	T	U	F	F			
	N	Q	E	T	E	N	T	G	T	N	N	L	R	E	N	E	I	A	T	I	T	R	C	L	R	A	M	O	S	E	L	L	A	E	R	E	R	I	S		
	U	O	L	H	S	H	I	N	I	C	C	A	N	A	A	A	H	I	I	V	U	E	E	I	T	E	N	S	O	A	V	O	V	E	I	I	H	O	I		
a A - z Z	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
acute á Â	.	.	.	.	*	.	*	.	.	.	.	*	.	*	*	*	*	.	.	.	.	.	*	.	*	.	.	*	.	.	.	.	.	*	.	.	*	.	.	*	
grave à Æ	.	.	.	.	*	.	.	*	.	.	.	*	.	*	.	.	.	*	.	.	.	.	*	.	*	.	.	*	.	.	.	.	.	.	.	*	.	.	*		
circu â Æ	.	.	.	.	.	.	.	*	.	.	*	.	.	.	.	.	.	.	.	.	.	.	.	*	.	*	.	.	.	.	.	.	.	.	.	*	.	.	*		
diære ä Ä	.	.	.	.	.	.	*	.	*	*	.	*	.	.	.	.	.	.	.	.	.	.	.	.	*	.	.	.	*	.	.	.	.	.	*	.	.	*			
tilde ã Ã	.	.	.	.	.	.	.	.	.	.	.	.	*	.	.	.	.	.	.	.	.	.	.	*	.	*	.	.	.	.	.	.	.	.	.	.	.	.	.		
breve ă Ǻ	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	*	.	.	.	.	.	.	.	.	.	.	.	.	.	.		
ring å Å	.	.	.	.	.	*	.	.	.	.	.	.	*	.	.	.	.	.	.	*	.	.	.	.	.	.	.	.	.	.	.	.	.	.	*	.	.	.	.		
macron ā Ā	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	*	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		
ogonek ą Ą	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	*	.	.	.	.	*	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		
a+e æ Æ	.	.	.	.	.	*	.	*	.	.	.	.	*	.	.	*	.	.	.	.	.	.	*	.	.	.	.	.	.	.	.	.	.	.	.	.	.	*	.	.	
acute ć Ć	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	*	.	.	.	*	.	.	.	.	.	*	.	.	.	.	.	.	.	.	.	.		
caron č Č	.	.	.	.	*	.	.	.	.	.	.	.	.	.	.	.	.	*	*	.	.	.	.	.	.	.	.	*	*	*	*	.	.	.	.	.	.	.	.		
dot ab ċ Ć	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	*	.	.	*	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		
cedill ç Ç	*	.	.	*	.	.	*	.	.	.	*	.	.	.	.	.	.	.	.	.	.	.	*	.	*	.	.	.	.	.	.	.	.	.	.	*	.	.	.		
c+h C+H =	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	*	.	.	.	.		
caron đ Ď	.	.	.	.	*	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	*	*	.	.	.	.	.	.	.	.	.	.	
stroke đ Đ	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	*	.	.	.	.	.	.	.	.	.	.	.	
eth ð Ð	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	*	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	*	.	.	





**TABLE 2A: Continued**

LANGUAGES	A	B	B	C	C	D	D	E	E	F	F	G	G	G	H	I	I	I	L	L	M	N	O	P	P	R	R	S	S	S	S	S	S	I	W	F	F		
Characters	L	A	R	A	Z	A	U	N	S	I	R	A	E	R	U	C	I	A	L	I	A	O	C	P	P	H	O	S	S	S	S	S	I	W	F	F			
	N	Q	E	T	E	N	T	G	T	N	N	M	R	E	N	E	I	A	T	H	L	C	C	L	T	A	M	O	S	O	V	L	E	R	E	R	I		
acute	š Š	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	*	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
caron	š Š	.	.	.	*	.	.	.	*	*	.	.	.	.	.	.	.	.	*	*	.	.	.	.	.	.	.	*	*	*	.	.	.	.	.	.	.	.	.
cedill	ş Ş	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	*	.	.	.	.	.	.	.	.	.	.	.	.
sharp s	ß ß	.	.	.	.	.	.	.	.	.	.	.	*	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
caron	ť ť	.	.	.	.	*	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	*	.	.	.	.	.	.	.	.
cedill	ț Ț	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	*	.	.	.	.	.	.	.	.	.	.	.	.	.
thorn	þ Þ	.	.	.	.	.	.	.	.	.	.	.	.	.	.	*	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
acute	ú Ú	.	.	.	*	*	.	.	.	.	.	*	.	*	*	*	*	*	.	.	.	.	*	.	*	.	.	.	*	.	.	.	.	.	*	*	.	*	
grave	ù Ù	.	.	*	.	.	.	.	.	.	.	.	.	*	.	.	.	*	.	.	.	.	.	*	.	*	.	.	.	.	.	.	.	.	*	*	.	*	
circum	û Û	.	.	.	.	.	.	.	.	*	*	.	.	*	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	*	*	.	*	
diæres	ü Ü	*	*	*	.	.	*	.	*	*	*	*	*	*	.	.	.	.	.	.	.	.	*	*	*	.	.	.	.	.	*	.	*	*	*	*	*		
tilde	ũ Õ	.	.	.	.	.	.	.	.	.	.	.	.	*	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
dou ac	ű Ű	.	.	.	.	.	.	.	.	.	.	.	.	.	*	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ring	ũ Ű	.	.	.	.	*	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
macron	ū Ū	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	*	*	*	*	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
ogonek	ų Ų	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	*	*	*	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
circum	ŵ Ŵ	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
acute	ý Ý	.	.	.	.	*	.	.	.	.	.	.	.	.	.	*	.	.	.	.	.	.	.	.	.	.	.	.	*	.	.	.	.	.	.	.	.	*	
circum	ÿ Ŷ	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
diæres	ÿ Ŷ	.	.	.	.	.	.	.	.	.	*	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	*	*	.	.
acute	ž Ž	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	*	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
caron	ž Ž	.	.	.	*	.	.	.	*	*	.	.	.	.	.	.	.	*	*	*	*	.	.	.	.	.	.	*	*	*	.	.	.	.	.	.	.	.	
dot ab	ž Ž	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	*	*	*	*	*	*	*	.	.	.	.	.	.	.	.	.	.	.	.	.	

Character combinations marked with = cannot be reproduced with the word-processing system used to compile the present document.

**TABLE 2B: Possible additions for other languages**

for ESPERANTO		for AFRIKAANS	
circumflex c	ĉ Ĉ	apostrophe n	'n =
circumflex g	ĝ Ĝ		
circumflex h	ĥ Ĥ	for LAP	
circumflex j	ĵ Ĵ		
circumflex s	ŝ Ŝ	eng	ŋ Ņ
breve u	ŭ Ŭ	stroke t	t T

**TABLE 2C: Accents and diacritical marks necessary for Latin-script European Languages**<sup>ref 2,3</sup>

´	Acute accent
`	Grave accent
ˆ	Circumflex accent
¨	Diæresis/umlaut mark
˜	Tilde
˘	Caron
˙	Breve
˚	Double acute accent
◦	Ring
·	Dot above
ˉ	Macron
¸	Cedilla
˛	Ogonek



**TABLE 2D: CHARACTER REPERTOIRE (TABLES OF ISO 10646)**

**IMPORTANT NOTICE :**

The ISO/IEC JTC1/SC2/WG2 introduced several modifications in the DIS 10646 during its last meeting in Seoul (29 June - 3 July 1992); as these modifications concern the Tables and Rows mentioned in the present document, the contents of this TABLE are to be considered as provisional.

***PRIORITY 1 LANGUAGES***

**Characters for latin-script languages**

TABLE 1-Row 00: ISO-646 IRV

TABLE 2-Row 00: LATIN-1 SUPPLEMENT

TABLE 3-Row 01: EXTENDED LATIN-A (ij, œ, š, Ÿ, ž)<sup>1</sup>

TABLE 7-Row 02: SPACING MODIFIER LETTERS (ch, accents)

**Characters for greek-script languages**

TABLE 9-Row 03: GREEK

**Characters for cyrillic-script languages**

TABLE 10-Row 04: CYRILLIC<sup>2</sup>

**Other characters**

TABLE 33-Row 20: GENERAL PUNCTUATION

TABLE 34-Row 20: SUPERSCRIPTS AND SUBSCRIPTS

TABLE 35-Row 20: CURRENCY SYMBOLS, COMBINING DIACRITICAL MARKS FOR SYMBOLS

TABLE 38-Row 22: MATHEMATICAL OPERATORS

TABLE 39-Row 22: MATHEMATICAL OPERATORS

TABLE 43-Row 25: BOX DRAWING

TABLE 44-Row 25: BLOCK ELEMENTS, GEOMETRIC SHAPES

***PRIORITY 2 LANGUAGES***

**Additional characters for latin-script languages**

TABLE 4-Row 01: EXTENDED LATIN-B<sup>3</sup>

TABLE 5-Row 02: EXTENDED LATIN-B

**Additional characters for greek-script languages**

TABLE 31-Row 1F: GREEK EXTENSIONS

TABLE 32-Row 1F: GREEK EXTENSIONS

**Additional characters for cyrillic-script languages**

TABLE 11-Row 04: CYRILLIC

**Characters for armenian-script languages**

---

<sup>1</sup> the characters in this Table also cover all the latin-script PRIORITY 2 Languages.

<sup>2</sup> the characters of these Tables also cover all the cyrillic-script PRIORITY 2 Languages.

<sup>3</sup> to cover several latin characters for PRIORITY 2 Languages written with the latin and the cyrillic scripts.

TABLE 12-Row 05: ARMENIAN

**Characters for hebrew-script languages**

TABLE 13-Row 05: HEBREW

**Characters for arabic-script languages**

TABLE 014-Row 06: ARABIC

TABLE 015-Row 06: ARABIC

**Characters for georgian-script languages**

TABLE 28-Row 10: GEORGIAN

**Characters for standard phonetic**

TABLE 6-Row 02: IPA EXTENSIONS

**Other characters**

TABLE 36-Row 21: LETTERLIKE SYMBOLS, NUMBER FORMS

TABLE 37-Row 21: ARROWS

TABLE 40-Row 23: MISCELLANEOUS TECHNICAL

TABLE 41-Row 24: CONTROL PICTURES, OPTICAL CHARACTER RECOGNITION

TABLE 42-Row 24: ENCLOSED ALPHANUMERICS

TABLE 45-Row 26: MISCELLANEOUS DINGBATS

TABLE 46-Row 27: DINGBATS

TABLE 47-Row 27: DINGBATS

**TABLE 2E: CHARACTERS**

## IMPORTANT NOTICE :

The ISO/IEC JTC1/SC2/WG2 introduced several modifications in the DIS 10646 during its last meeting in Seoul (29 June - 3 July 1992); as these modifications concern the Tables and Rows mentioned in the present document, the contents of this TABLE are to be considered as provisional.

<>: Spacing Characters, []: Non-spacing characters

**CO CONTROL CODES**

<NUL>	/d000/d000	NULL
⋯ <IS1> <US>	⋯ /d000/d031	⋯ INFORMATION SEPARATOR ONE (UNIT SEPARATOR)

**TABLE 1-ROW 00: ISO-646 IRV (NAMES ACCORDING TO ISO 10646)**

<SP>	/d000/d032	SPACE
<!>	/d000/d033	EXCLAMATION MARK
<">	/d000/d034	QUOTATION MARK
<#>	/d000/d035	NUMBER SIGN
<\$>	/d000/d036	DOLLAR SIGN
<%>	/d000/d037	PERCENT SIGN
<&>	/d000/d038	AMPERSAND
<'>	/d000/d039	APOSTROPHE
<( >	/d000/d040	LEFT PARENTHESIS
<)>	/d000/d041	RIGHT PARENTHESIS
<*>	/d000/d042	ASTERISK
<+>	/d000/d043	PLUS SIGN
<, >	/d000/d044	COMMA
<- >	/d000/d045	HYPHEN-MINUS
<. >	/d000/d046	FULL STOP
</ >	/d000/d047	SOLIDUS
<0 >	/d000/d048	DIGIT ZERO
<1 >	/d000/d049	DIGIT ONE
<2 >	/d000/d050	DIGIT TWO
<3 >	/d000/d051	DIGIT THREE
<4 >	/d000/d052	DIGIT FOUR
<5 >	/d000/d053	DIGIT FIVE
<6 >	/d000/d054	DIGIT SIX
<7 >	/d000/d055	DIGIT SEVEN
<8 >	/d000/d056	DIGIT EIGHT
<9 >	/d000/d057	DIGIT NINE
<: >	/d000/d058	COLON
<; >	/d000/d059	SEMICOLON
<< >	/d000/d060	LESS-THAN SIGN
<=>	/d000/d061	EQUALS SIGN
<> >	/d000/d062	GREATER-THAN SIGN
<? >	/d000/d063	QUESTION MARK
<@ >	/d000/d064	COMMERCIAL AT
<A >	/d000/d065	LATIN CAPITAL LETTER A
<B >	/d000/d066	LATIN CAPITAL LETTER B
<C >	/d000/d067	LATIN CAPITAL LETTER C

<D>	/d000/d068	LATIN CAPITAL LETTER D
<E>	/d000/d069	LATIN CAPITAL LETTER E
<F>	/d000/d070	LATIN CAPITAL LETTER F
<G>	/d000/d071	LATIN CAPITAL LETTER G
<H>	/d000/d072	LATIN CAPITAL LETTER H
<I>	/d000/d073	LATIN CAPITAL LETTER I
<J>	/d000/d074	LATIN CAPITAL LETTER J
<K>	/d000/d075	LATIN CAPITAL LETTER K
<L>	/d000/d076	LATIN CAPITAL LETTER L
<M>	/d000/d077	LATIN CAPITAL LETTER M
<N>	/d000/d078	LATIN CAPITAL LETTER N
<O>	/d000/d079	LATIN CAPITAL LETTER O
<P>	/d000/d080	LATIN CAPITAL LETTER P
<Q>	/d000/d081	LATIN CAPITAL LETTER Q
<R>	/d000/d082	LATIN CAPITAL LETTER R
<S>	/d000/d083	LATIN CAPITAL LETTER S
<T>	/d000/d084	LATIN CAPITAL LETTER T
<U>	/d000/d085	LATIN CAPITAL LETTER U
<V>	/d000/d086	LATIN CAPITAL LETTER V
<W>	/d000/d087	LATIN CAPITAL LETTER W
<X>	/d000/d088	LATIN CAPITAL LETTER X
<Y>	/d000/d089	LATIN CAPITAL LETTER Y
<Z>	/d000/d090	LATIN CAPITAL LETTER Z
<[>	/d000/d091	LEFT SQUARE BRACKET
<\>	/d000/d092	REVERSE SOLIDUS
<]>	/d000/d093	RIGHT SQUARE BRACKET
<^>	/d000/d094	CIRCUMFLEX ACCENT
<_>	/d000/d095	LOW LINE
<`>	/d000/d096	GRAVE ACCENT
<a>	/d000/d097	LATIN SMALL LETTER A
<b>	/d000/d098	LATIN SMALL LETTER B
<c>	/d000/d099	LATIN SMALL LETTER C
<d>	/d000/d100	LATIN SMALL LETTER D
<e>	/d000/d101	LATIN SMALL LETTER E
<f>	/d000/d102	LATIN SMALL LETTER F
<g>	/d000/d173	LATIN SMALL LETTER G
<h>	/d000/d104	LATIN SMALL LETTER H
<i>	/d000/d105	LATIN SMALL LETTER I
<j>	/d000/d106	LATIN SMALL LETTER J
<k>	/d000/d107	LATIN SMALL LETTER K
<l>	/d000/d108	LATIN SMALL LETTER L
<m>	/d000/d109	LATIN SMALL LETTER M
<n>	/d000/d110	LATIN SMALL LETTER N
<o>	/d000/d111	LATIN SMALL LETTER O
<p>	/d000/d112	LATIN SMALL LETTER P
<q>	/d000/d113	LATIN SMALL LETTER Q
<r>	/d000/d114	LATIN SMALL LETTER R
<s>	/d000/d115	LATIN SMALL LETTER S
<t>	/d000/d116	LATIN SMALL LETTER T
<u>	/d000/d117	LATIN SMALL LETTER U
<v>	/d000/d118	LATIN SMALL LETTER V
<w>	/d000/d119	LATIN SMALL LETTER W
<x>	/d000/d120	LATIN SMALL LETTER X
<y>	/d000/d121	LATIN SMALL LETTER Y
<z>	/d000/d122	LATIN SMALL LETTER Z
<{>	/d000/d123	LEFT CURLY BRACKET

< >	/d000/d124	VERTICAL LINE
<}>	/d000/d125	RIGHT CURLY BRACKET
<~>	/d000/d126	TILDE
<DEL>	/d000/d127	DELETE

**C1 CONTROL CODES (TO BE ADDED LATER)**

<>	/d000/d128	...
..	...	...
<>	/d000/d159	...

**TABLE 2-ROW 00: LATIN-1 Supplement (NAMES ACCORDING TO ISO 10646)**

<NBSP>	/d000/d160	NON-BREAK SPACE
<¡>	/d000/d161	INVERTED EXCLAMATION MARK
<¢>	/d000/d162	CENT SIGN
<£>	/d000/d163	POUND SIGN
<¤>	/d000/d164	CURRENCY SIGN
<¥>	/d000/d165	YEN SIGN
< >	/d000/d166	BROKEN BAR
<§>	/d000/d167	SECTION SIGN
<¨>	/d000/d168	DIAERESIS
<©>	/d000/d169	COPYRIGHT SIGN
<ª>	/d000/d170	FEMININE ORDINAL INDICATOR
<«>	/d000/d171	LEFT-POINTING DOUBLE ANGLE QUOTATION MARK
<¬>	/d000/d172	NOT SIGN
<SHY>	/d000/d173	SOFT HYPHEN
<®>	/d000/d174	REGISTERED SIGN
<¯>	/d000/d175	MACRON
<°>	/d000/d176	DEGREE SIGN
<±>	/d000/d177	PLUS-MINUS SIGN
<²>	/d000/d178	SUPERSCRIP TWO
<³>	/d000/d179	SUPERSCRIP THREE
<´>	/d000/d180	ACUTE ACCENT
<µ>	/d000/d181	MICRO SIGN
<¶>	/d000/d182	PILCROW SIGN
<·>	/d000/d183	MIDDLE DOT
<¸>	/d000/d184	CEDILLA
<¹>	/d000/d185	SUPERSCRIP ONE
<º>	/d000/d186	MASCULINE ORDINAL INDICATOR
<»>	/d000/d187	RIGHT-POINTING DOUBLE ANGLE QUOTATION MARK
<¼>	/d000/d188	VULGAR FRACTION ONE QUARTER
<½>	/d000/d189	VULGAR FRACTION ONE HALF
<¾>	/d000/d190	VULGAR FRACTION THREE QUARTERS
<¿>	/d000/d191	INVERTED QUESTION MARK
<À>	/d000/d192	LATIN CAPITAL LETTER A WITH GRAVE
<Á>	/d000/d193	LATIN CAPITAL LETTER A WITH ACUTE
<Â>	/d000/d194	LATIN CAPITAL LETTER A WITH CIRCUMFLEX
<Ã>	/d000/d195	LATIN CAPITAL LETTER A WITH TILDE
<Ä>	/d000/d196	LATIN CAPITAL LETTER A WITH DIAERESIS
<Å>	/d000/d197	LATIN CAPITAL LETTER A WITH RING ABOVE
<Æ>	/d000/d198	LATIN CAPITAL LETTER AE
<Ç>	/d000/d199	LATIN CAPITAL LETTER C WITH CEDILLA

<È>	/d000/d200	LATIN CAPITAL LETTER E WITH GRAVE
<É>	/d000/d201	LATIN CAPITAL LETTER E WITH ACUTE
<Ê>	/d000/d202	LATIN CAPITAL LETTER E WITH CIRCUMFLEX
<Ë>	/d000/d203	LATIN CAPITAL LETTER E WITH DIAERESIS
<Ì>	/d000/d204	LATIN CAPITAL LETTER I WITH GRAVE
<Í>	/d000/d205	LATIN CAPITAL LETTER I WITH ACUTE
<Î>	/d000/d206	LATIN CAPITAL LETTER I WITH CIRCUMFLEX
<Ï>	/d000/d207	LATIN CAPITAL LETTER I WITH DIAERESIS
<Ð>	/d000/d208	LATIN CAPITAL LETTER ETH (Icelandic)
<Ñ>	/d000/d209	LATIN CAPITAL LETTER N WITH TILDE
<Ò>	/d000/d210	LATIN CAPITAL LETTER O WITH GRAVE
<Ó>	/d000/d211	LATIN CAPITAL LETTER O WITH ACUTE
<Ô>	/d000/d212	LATIN CAPITAL LETTER O WITH CIRCUMFLEX
<Õ>	/d000/d213	LATIN CAPITAL LETTER O WITH TILDE
<Ö>	/d000/d214	LATIN CAPITAL LETTER O WITH DIAERESIS
<×>	/d000/d215	MULTIPLICATION SIGN
<Ø>	/d000/d216	LATIN CAPITAL LETTER O WITH STROKE
<Û>	/d000/d217	LATIN CAPITAL LETTER U WITH GRAVE
<Ü>	/d000/d218	LATIN CAPITAL LETTER U WITH ACUTE
<Û>	/d000/d219	LATIN CAPITAL LETTER U WITH CIRCUMFLEX
<Û>	/d000/d220	LATIN CAPITAL LETTER U WITH DIAERESIS
<Ý>	/d000/d221	LATIN CAPITAL LETTER Y WITH ACUTE
<Þ>	/d000/d222	LATIN CAPITAL LETTER THORN (Icelandic)
<ß>	/d000/d223	LATIN SMALL GERMAN LETTER SHARP S (German)
<à>	/d000/d224	LATIN SMALL LETTER A WITH GRAVE
<á>	/d000/d225	LATIN SMALL LETTER A WITH ACUTE
<â>	/d000/d226	LATIN SMALL LETTER A WITH CIRCUMFLEX
<ã>	/d000/d227	LATIN SMALL LETTER A WITH TILDE
<ä>	/d000/d228	LATIN SMALL LETTER A WITH DIAERESIS
<å>	/d000/d229	LATIN SMALL LETTER A WITH RING ABOVE
<æ>	/d000/d230	LATIN SMALL LETTER AE
<ç>	/d000/d231	LATIN SMALL LETTER C WITH CEDILLA
<è>	/d000/d232	LATIN SMALL LETTER E WITH GRAVE
<é>	/d000/d233	LATIN SMALL LETTER E WITH ACUTE
<ê>	/d000/d234	LATIN SMALL LETTER E WITH CIRCUMFLEX
<ë>	/d000/d235	LATIN SMALL LETTER E WITH DIAERESIS
<ì>	/d000/d236	LATIN SMALL LETTER I WITH GRAVE
<í>	/d000/d237	LATIN SMALL LETTER I WITH ACUTE
<î>	/d000/d238	LATIN SMALL LETTER I WITH CIRCUMFLEX
<ï>	/d000/d239	LATIN SMALL LETTER I WITH DIAERESIS
<ð>	/d000/d240	LATIN SMALL LETTER ETH (Icelandic)
<ñ>	/d000/d241	LATIN SMALL LETTER N WITH TILDE
<ò>	/d000/d242	LATIN SMALL LETTER O WITH GRAVE
<ó>	/d000/d243	LATIN SMALL LETTER O WITH ACUTE
<ô>	/d000/d244	LATIN SMALL LETTER O WITH CIRCUMFLEX
<õ>	/d000/d245	LATIN SMALL LETTER O WITH TILDE
<ö>	/d000/d246	LATIN SMALL LETTER O WITH DIAERESIS
<÷>	/d000/d247	DIVISION SIGN
<ø>	/d000/d248	LATIN SMALL LETTER O WITH STROKE
<ù>	/d000/d249	LATIN SMALL LETTER U WITH GRAVE
<ú>	/d000/d250	LATIN SMALL LETTER U WITH ACUTE
<û>	/d000/d251	LATIN SMALL LETTER U WITH CIRCUMFLEX
<ü>	/d000/d252	LATIN SMALL LETTER U WITH DIAERESIS
<ý>	/d000/d253	LATIN SMALL LETTER Y WITH ACUTE
<þ>	/d000/d254	LATIN SMALL LETTER THORN (Icelandic)
<ÿ>	/d000/d255	LATIN SMALL LETTER Y WITH DIAERESIS

**TABLE 3-ROW 01-EXTENDED LATIN-A (NAMES ACCORDING TO ISO 10646)**

<Ā>	/d001/d000	LATIN CAPITAL LETTER A WITH MACRON
<ā>	/d001/d001	LATIN SMALL LETTER A WITH MACRON
<Ă>	/d001/d002	LATIN CAPITAL LETTER A WITH BREVE
<ă>	/d001/d003	LATIN SMALL LETTER A WITH BREVE
<Ą>	/d001/d004	LATIN CAPITAL LETTER A WITH OGONEK
<ą>	/d001/d005	LATIN SMALL LETTER A WITH OGONEK
<Ć>	/d001/d006	LATIN CAPITAL LETTER C WITH ACUTE
<ć>	/d001/d007	LATIN SMALL LETTER C WITH ACUTE
<Ĉ>	/d001/d008	LATIN CAPITAL LETTER C WITH CIRCUMFLEX
<ĉ>	/d001/d009	LATIN SMALL LETTER C WITH CIRCUMFLEX
<Ċ>	/d001/d010	LATIN CAPITAL LETTER C WITH DOT ABOVE
<ċ>	/d001/d011	LATIN SMALL LETTER C WITH DOT ABOVE
<Č>	/d001/d012	LATIN CAPITAL LETTER C WITH CARON
<č>	/d001/d013	LATIN SMALL LETTER C WITH CARON
<Ď>	/d001/d014	LATIN CAPITAL LETTER D WITH CARON
<Ǳ>	/d001/d015	LATIN SMALL LETTER D WITH CARON
<Ð>	/d001/d016	LATIN CAPITAL LETTER D WITH STROKE
<ð>	/d001/d017	LATIN SMALL LETTER D WITH STROKE
<Ē>	/d001/d018	LATIN CAPITAL LETTER E WITH MACRON
<ē>	/d001/d019	LATIN SMALL LETTER E WITH MACRON
<Ĕ>	/d001/d020	LATIN CAPITAL LETTER E WITH BREVE
<ĕ>	/d001/d021	LATIN SMALL LETTER E WITH BREVE
<Ė>	/d001/d022	LATIN CAPITAL LETTER E WITH DOT ABOVE
<ė>	/d001/d023	LATIN SMALL LETTER E WITH DOT ABOVE
<Ę>	/d001/d024	LATIN CAPITAL LETTER E WITH OGONEK
<ę>	/d001/d025	LATIN SMALL LETTER E WITH OGONEK
<Ě>	/d001/d026	LATIN CAPITAL LETTER E WITH CARON
<ě>	/d001/d027	LATIN SMALL LETTER E WITH CARON
<Ĝ>	/d001/d028	LATIN CAPITAL LETTER G WITH CIRCUMFLEX
<ĝ>	/d001/d029	LATIN SMALL LETTER G WITH CIRCUMFLEX
<Ġ>	/d001/d030	LATIN CAPITAL LETTER G WITH BREVE
<ġ>	/d001/d031	LATIN SMALL LETTER G WITH BREVE
<Ģ>	/d001/d032	LATIN CAPITAL LETTER G WITH DOT ABOVE
<ģ>	/d001/d033	LATIN SMALL LETTER G WITH DOT ABOVE
<Ġ>	/d001/d034	LATIN CAPITAL LETTER G WITH CEDILLA
<ġ>	/d001/d035	LATIN SMALL LETTER G WITH CEDILLA
<Ĥ>	/d001/d036	LATIN CAPITAL LETTER H WITH CIRCUMFLEX
<ĥ>	/d001/d037	LATIN SMALL LETTER H WITH CIRCUMFLEX
<Ħ>	/d001/d038	LATIN CAPITAL LETTER H WITH STROKE
<ħ>	/d001/d039	LATIN SMALL LETTER H WITH STROKE
<İ>	/d001/d040	LATIN CAPITAL LETTER I WITH TILDE
<î>	/d001/d041	LATIN SMALL LETTER I WITH TILDE

<İ>	/d001/d042	LATIN CAPITAL LETTER I WITH MACRON
<i>	/d001/d043	LATIN SMALL LETTER I WITH MACRON
<ı̇>	/d001/d044	LATIN CAPITAL LETTER I WITH BREVE
<ı̈>	/d001/d045	LATIN SMALL LETTER I WITH BREVE
<İ>	/d001/d046	LATIN CAPITAL LETTER I WITH OGONEK
<i̇>	/d001/d047	LATIN SMALL LETTER I WITH OGONEK
<İ̇>	/d001/d048	LATIN CAPITAL LETTER I WITH DOT ABOVE
<i̇>	/d001/d049	LATIN SMALL LETTER DOTLESS I
<IJ>	/d001/d050	LATIN CAPITAL LIGATURE IJ
<ij>	/d001/d051	LATIN SMALL LIGATURE IJ
<Ĳ>	/d001/d052	LATIN CAPITAL LETTER J WITH CIRCUMFLEX
<ĵ>	/d001/d053	LATIN SMALL LETTER J WITH CIRCUMFLEX
<Ķ>	/d001/d054	LATIN CAPITAL LETTER K WITH CEDILLA
<k̃>	/d001/d055	LATIN SMALL LETTER K WITH CEDILLA
<ƙ>	/d001/d056	LATIN SMALL LETTER KRA (Greenlandic)
<Ĺ>	/d001/d057	LATIN CAPITAL LETTER L WITH ACUTE
<l̇>	/d001/d058	LATIN SMALL LETTER L WITH ACUTE
<Ľ>	/d001/d059	LATIN CAPITAL LETTER L WITH CEDILLA
<l̃>	/d001/d060	LATIN SMALL LETTER L WITH CEDILLA
<Ľ̇>	/d001/d061	LATIN CAPITAL LETTER L WITH CARON
<ĺ̇>	/d001/d062	LATIN SMALL LETTER L WITH CARON
<Ł>	/d001/d063	LATIN CAPITAL LETTER L WITH MIDDLE DOT
<ł>	/d001/d064	LATIN SMALL LETTER L WITH MIDDLE DOT
<Ł̇>	/d001/d065	LATIN CAPITAL LETTER L WITH STROKE
<ł̇>	/d001/d066	LATIN SMALL LETTER L WITH STROKE
<Ń>	/d001/d067	LATIN CAPITAL LETTER N WITH ACUTE
<ṅ>	/d001/d068	LATIN SMALL LETTER N WITH ACUTE
<Ñ>	/d001/d069	LATIN CAPITAL LETTER N WITH CEDILLA
<ñ>	/d001/d070	LATIN SMALL LETTER N WITH CEDILLA
<Ñ̇>	/d001/d071	LATIN CAPITAL LETTER N WITH CARON
<ṅ̇>	/d001/d072	LATIN SMALL LETTER N WITH CARON
<ñ̇>	/d001/d073	LATIN SMALL LETTER N PRECEDED BY APOSTROPHE
<Ŋ>	/d001/d074	LATIN CAPITAL LETTER ENG (Lappish)
<ŋ>	/d001/d075	LATIN SMALL LETTER ENG (Lappish)
<Ō>	/d001/d076	LATIN CAPITAL LETTER O WITH MACRON
<ō>	/d001/d077	LATIN SMALL LETTER O WITH MACRON
<ö̇>	/d001/d078	LATIN CAPITAL LETTER O WITH BREVE
<ö̈>	/d001/d079	LATIN SMALL LETTER O WITH BREVE
<Ŏ>	/d001/d080	LATIN CAPITAL LETTER O WITH DOUBLE ACUTE
<ö̈̈>	/d001/d081	LATIN SMALL LETTER O WITH DOUBLE ACUTE
<Œ>	/d001/d082	LATIN CAPITAL LIGATURE OE
<œ>	/d001/d083	LATIN SMALL LIGATURE OE
<Ŕ>	/d001/d084	LATIN CAPITAL LETTER R WITH ACUTE
<ṙ>	/d001/d085	LATIN SMALL LETTER R WITH ACUTE
<Ŗ>	/d001/d086	LATIN CAPITAL LETTER R WITH CEDILLA
<r̃>	/d001/d087	LATIN SMALL LETTER R WITH CEDILLA



<Ř>	/d001/d088	LATIN CAPITAL LETTER R WITH CARON
<ř>	/d001/d089	LATIN SMALL LETTER R WITH CARON
<Š>	/d001/d090	LATIN CAPITAL LETTER S WITH ACUTE
<š>	/d001/d091	LATIN SMALL LETTER S WITH ACUTE
<Ŝ>	/d001/d092	LATIN CAPITAL LETTER S WITH CIRCUMFLEX
<ŝ>	/d001/d093	LATIN SMALL LETTER S WITH CIRCUMFLEX
<Ş>	/d001/d094	LATIN CAPITAL LETTER S WITH CEDILLA
<ş>	/d001/d095	LATIN SMALL LETTER S WITH CEDILLA
<Ṣ̌>	/d001/d096	LATIN CAPITAL LETTER S WITH CARON
<ṣ̌>	/d001/d097	LATIN SMALL LETTER S WITH CARON
<Ť>	/d001/d098	LATIN CAPITAL LETTER T WITH CEDILLA
<ť>	/d001/d099	LATIN SMALL LETTER T WITH CEDILLA
<Ṭ̌>	/d001/d100	LATIN CAPITAL LETTER T WITH CARON
<ṭ̌>	/d001/d101	LATIN SMALL LETTER T WITH CARON
<Ƨ>	/d001/d102	LATIN CAPITAL LETTER T WITH STROKE
<Ƨ̣>	/d001/d103	LATIN SMALL LETTER T WITH STROKE
<Ū>	/d001/d104	LATIN CAPITAL LETTER U WITH TILDE
<ũ>	/d001/d105	LATIN SMALL LETTER U WITH TILDE
<Ū̄>	/d001/d106	LATIN CAPITAL LETTER U WITH MACRON
<ū̄>	/d001/d107	LATIN SMALL LETTER U WITH MACRON
<Ū̆>	/d001/d108	LATIN CAPITAL LETTER U WITH BREVE
<ū̆>	/d001/d109	LATIN SMALL LETTER U WITH BREVE
<Ū̇>	/d001/d110	LATIN CAPITAL LETTER U WITH RING ABOVE
<ū̇>	/d001/d111	LATIN SMALL LETTER U WITH RING ABOVE
<Ṻ>	/d001/d112	LATIN CAPITAL LETTER U WITH DOUBLE ACUTE
<ṻ>	/d001/d113	LATIN SMALL LETTER U WITH DOUBLE ACUTE
<Ū̊>	/d001/d114	LATIN CAPITAL LETTER U WITH OGONEK
<ū̊>	/d001/d115	LATIN SMALL LETTER U WITH OGONEK
<Ŵ>	/d001/d116	LATIN CAPITAL LETTER W WITH CIRCUMFLEX
<ŵ>	/d001/d117	LATIN SMALL LETTER W WITH CIRCUMFLEX
<Ỳ>	/d001/d118	LATIN CAPITAL LETTER Y WITH CIRCUMFLEX
<ỳ>	/d001/d119	LATIN SMALL LETTER Y WITH CIRCUMFLEX
<ÿ>	/d001/d120	LATIN CAPITAL LETTER Y WITH DIAERESIS
<Ẑ>	/d001/d121	LATIN CAPITAL LETTER Z WITH ACUTE
<ẑ>	/d001/d122	LATIN SMALL LETTER Z WITH ACUTE
<Ẓ>	/d001/d123	LATIN CAPITAL LETTER Z WITH DOT ABOVE
<ẓ>	/d001/d124	LATIN SMALL LETTER Z WITH DOT ABOVE
<Ẕ>	/d001/d125	LATIN CAPITAL LETTER Z WITH CARON
<ẕ>	/d001/d126	LATIN SMALL LETTER Z WITH CARON
<Ϟ>	/d001/d127	(This position shall not be used)

**TABLE 4-ROW 01: EXTENDED LATIN-B (Priority 2)**

< Ƨ̣ >	/d001/d128	LATIN SMALL LETTER B WITH STROKE
...	...	...
< ȯ >	/d001/d255	LATIN SMALL LETTER O WITH STROKE AND ACUTE

**TABLE 5-ROW 02: EXTENDED LATIN-B (Priority 2)**

< Å >	/d002/d000	LATIN CAPITAL LETTER A WITH DOUBLE GRAVE
...	...	...
< Ū >	/d002/d023	LATIN SMALL LETTER U WITH INVERTED BREVE
◊	/d002/d024	(This position shall not be used)
...	...	...
◊	/d002/d079	(This position shall not be used)

**TABLE 6-ROW 02: IPA EXTENSIONS (Priority 2)**

< Ǟ >	/d002/d080	LATIN SMALL LETTER TURNED A
...	...	...
< <sup>t</sup> ç >	/d002/d168	LATIN SMALL LETTER TC WITH CURL
◊	/d002/d169	(This position shall not be used)
...	...	...
◊	/d002/d175	(This position shall not be used)

**TABLE 7-ROW 02-SPACING MODIFIER LETTERS (NAMES ACCORDING TO ISO 10646)**

<h>	/d002/d176	MODIFIER LETTER SMALL H
<h̃>	/d002/d177	MODIFIER LETTER SMALL H WITH HOOK
<j>	/d002/d178	MODIFIER LETTER SMALL J
<r>	/d002/d179	MODIFIER LETTER SMALL R
< ɽ >	/d002/d180	MODIFIER LETTER SMALL TURNED R
< ɽ̃ >	/d002/d181	MODIFIER LETTER SMALL TURNED R WITH HOOK
< ɹ̥ >	/d002/d182	MODIFIER LETTER SMALL CAPITAL INVERTED R
<w>	/d002/d183	MODIFIER LETTER SMALL W
<y>	/d002/d184	MODIFIER LETTER SMALL Y
<˘>	/d002/d185	MODIFIER LETTER PRIME
<˝>	/d002/d186	MODIFIER LETTER DOUBLE PRIME
<ˆ>	/d002/d187	MODIFIER LETTER TURNED COMMA
<˘˘>	/d002/d188	GREEK PSILI PNEUMATA (smooth)
<˘˘˘>	/d002/d189	GREEK DASIA PNEUMATA (rough)
< ˆ ) >	/d002/d190	MODIFIER LETTER RIGHT HALF RING
< ˆ ( >	/d002/d191	MODIFIER LETTER LEFT HALF RING
< ʔ >	/d002/d192	MODIFIER LETTER GLOTTAL STOP
< ʔ̥ >	/d002/d193	MODIFIER LETTER REVERSED GLOTTAL STOP
<<>	/d002/d194	MODIFIER LETTER LEFT ARROWHEAD
<>>	/d002/d195	MODIFIER LETTER RIGHT ARROWHEAD

<^>	/d002/d196	MODIFIER LETTER UP ARROWHEAD
<v>	/d002/d197	MODIFIER LETTER DOWN ARROWHEAD
<^>	/d002/d198	MODIFIER LETTER CIRCUMFLEX ACCENT
<˘>	/d002/d199	CARON
< >	/d002/d200	MODIFIER LETTER VERTICAL LINE
<->	/d002/d201	MODIFIER LETTER MACRON
<ˆ>	/d002/d202	MODIFIER LETTER ACCUTE ACCENT
<◊>	/d002/d203	MODIFIER LETTER GRAVE ACCENT
< _>	/d002/d204	MODIFIER LETTER LOW VERTICAL LINE
<_>	/d002/d205	MODIFIER LETTER LOW MACRON
<˘_>	/d002/d206	MODIFIER LETTER LOW GRAVE ACCENT
<ˆ_>	/d002/d207	MODIFIER LETTER LOW ACUTE ACCENT
<▲>	/d002/d208	MODIFIER LETTER TRIANGULAR COLON
<▼>	/d002/d209	MODIFIER LETTER HALF TRIANGULAR COLON
< )>	/d002/d210	MODIFIER LETTER CENTRED RIGHT HALF RING
<( >	/d002/d211	MODIFIER LETTER CENTRED LEFT HALF RING
<⊥>	/d002/d212	MODIFIER LETTER UP TACK
<⊥>	/d002/d213	MODIFIER LETTER DOWN TACK
<+>	/d002/d214	MODIFIER LETTER PLUS SIGN
<->	/d002/d215	MODIFIER LETTER MINUS SIGN
<˘>	/d002/d216	BREVE
<˙>	/d002/d217	DOT ABOVE
<◌◦>	/d002/d218	RING ABOVE
<◌◌>	/d002/d219	OGONEK
<˘˘>	/d002/d220	SMALL TILDE
<˘˘˘>	/d002/d221	DOUBLE ACUTE ACCENT
<◌◌>	/d002/d222	MODIFIER LETTER RHOTIC HOOK
<◊>	/d002/d223	(This position shall not be used)
<◌γ>	/d002/d224	MODIFIER LETTER SMALL GAMMA
<◌l>	/d002/d225	MODIFIER LETTER SMALL L
<◌s>	/d002/d226	MODIFIER LETTER SMALL S
<◌x>	/d002/d227	MODIFIER LETTER SMALL X
<◌◌>	/d002/d228	MODIFIER LETTER SMALL REVERSED GLOTTAL STOP
<  >	/d002/d229	MODIFIER LETTER EXTRA-HIGH TONE BAR
<  >	/d002/d230	MODIFIER LETTER HIGH TONE BAR
<  >	/d002/d231	MODIFIER LETTER MID TONE BAR
<  >	/d002/d232	MODIFIER LETTER LOW TONE BAR
<  >	/d002/d233	MODIFIER LETTER EXTRA-LOW TONE BAR
<◊>	/d002/d234	(This position shall not be used)
...	...	...

◊ /d002/d255 (This position shall not be used)

**TABLE 8-ROW 03: COMBINING DIACRITICAL MARKS (This table is not to be used in order to avoid extra confusion/complexity in computation algorithms)**

[ ]	/d003/d000	COMBINING GRAVE ACCENT
...	...	...
[ ]	/d003/d065	COMBINING ACUTE TONE MARK (Vietnamese)
◊	/d003/d066	(This position shall not be used)
...	...	...
◊	/d003/d111	(This position shall not be used)

**TABLE 9-ROW 03: GREEK (NAMES ACCORDING TO ISO 10646)**

[ ]	/d003/d112	GREEK COMBINING IOTA BELOW
[ ]	/d003/d113	GREEK COMBINING DASIA PNEUMATA
[ ]	/d003/d114	GREEK COMBINING PSILI PNEUMATA
◊	/d003/d115	(This position shall not be used)
...	...	...
◊	/d003/d131	(This position shall not be used)
[ ]	/d003/d132	GREEK COMBINING TONOS
[ ]	/d003/d133	GREEK COMBINING DIAERESIS WITH TONOS
<A>	/d003/d134	GREEK CAPITAL LETTER ALPHA WITH ACUTE
◊	/d003/d135	(This position shall not be used)
<E>	/d003/d136	GREEK CAPITAL LETTER EPSILON WITH ACUTE
<H>	/d003/d137	GREEK CAPITAL LETTER ETA WITH ACUTE
<I>	/d003/d138	GREEK CAPITAL LETTER IOTA WITH ACUTE
◊	/d003/d139	(This position shall not be used)
<O>	/d003/d140	GREEK CAPITAL LETTER OMIKRON WITH ACUTE
◊	/d003/d141	(This position shall not be used)
<Y>	/d003/d142	GREEK CAPITAL LETTER UPSILON WITH ACUTE
<Ω>	/d003/d143	GREEK CAPITAL LETTER OMEGA WITH ACUTE
<ι>	/d003/d144	GREEK SMALL LETTER IOTA WITH ACUTE AND DIAERESIS
<A>	/d003/d145	GREEK CAPITAL LETTER ALPHA
<B>	/d003/d146	GREEK CAPITAL LETTER BETA
<Γ>	/d003/d147	GREEK CAPITAL LETTER GAMMA
<Δ>	/d003/d148	GREEK CAPITAL LETTER DELTA
<E>	/d003/d149	GREEK CAPITAL LETTER EPSILON
<Z>	/d003/d150	GREEK CAPITAL LETTER ZETA
<H>	/d003/d151	GREEK CAPITAL LETTER ETA
<Θ>	/d003/d152	GREEK CAPITAL LETTER THETA
<Ι>	/d003/d153	GREEK CAPITAL LETTER IOTA
<K>	/d003/d154	GREEK CAPITAL LETTER KAPPA
<Λ>	/d003/d155	GREEK CAPITAL LETTER LAMDA
<M>	/d003/d156	GREEK CAPITAL LETTER MU
<N>	/d003/d157	GREEK CAPITAL LETTER NU
<Ξ>	/d003/d158	GREEK CAPITAL LETTER XI
<O>	/d003/d159	GREEK CAPITAL LETTER OMICRON
<Π>	/d003/d160	GREEK CAPITAL LETTER PI
<P>	/d003/d161	GREEK CAPITAL LETTER RHO
◊	/d003/d162	(This position shall not be used)
<Σ>	/d003/d163	GREEK CAPITAL LETTER SIGMA
<T>	/d003/d164	GREEK CAPITAL LETTER TAU

<Υ>	/d003/d165	GREEK CAPITAL LETTER UPSILON
<Φ>	/d003/d166	GREEK CAPITAL LETTER PHI
<Χ>	/d003/d167	GREEK CAPITAL LETTER CHI
<Ψ>	/d003/d168	GREEK CAPITAL LETTER PSI
<Ω>	/d003/d169	GREEK CAPITAL LETTER OMEGA
<ϊ>	/d003/d170	GREEK CAPITAL LETTER IOTA WITH DIAERESIS
<ÿ>	/d003/d171	GREEK CAPITAL LETTER UPSILON WITH DIAERESIS
<ά>	/d003/d172	GREEK SMALL LETTER ALPHA WITH ACUTE
<έ>	/d003/d173	GREEK SMALL LETTER EPSILON WITH ACUTE
<ή>	/d003/d174	GREEK SMALL LETTER ETA WITH ACUTE
<ί>	/d003/d175	GREEK SMALL LETTER IOTA WITH ACUTE
<Ϸ>	/d003/d176	GREEK SMALL LETTER UPSILON WITH ACUTE AND DIAERESIS
<α>	/d003/d177	GREEK SMALL LETTER ALPHA
<β>	/d003/d178	GREEK SMALL LETTER BETA
<γ>	/d003/d179	GREEK SMALL LETTER GAMMA
<δ>	/d003/d180	GREEK SMALL LETTER DELTA
<ε>	/d003/d181	GREEK SMALL LETTER EPSILON
<ζ>	/d003/d182	GREEK SMALL LETTER ZETA
<η>	/d003/d183	GREEK SMALL LETTER ETA
<θ>	/d003/d184	GREEK SMALL LETTER THETA
<ι>	/d003/d185	GREEK SMALL LETTER IOTA
<κ>	/d003/d186	GREEK SMALL LETTER KAPPA
<λ>	/d003/d187	GREEK SMALL LETTER LAMDA
<μ>	/d003/d188	GREEK SMALL LETTER MU
<ν>	/d003/d189	GREEK SMALL LETTER NU
<ξ>	/d003/d190	GREEK SMALL LETTER XI
<ο>	/d003/d191	GREEK SMALL LETTER OMICRON
<π>	/d003/d192	GREEK SMALL LETTER PI
<ρ>	/d003/d193	GREEK SMALL LETTER RHO
<ς>	/d003/d194	GREEK SMALL LETTER FINAL SIGMA
<σ>	/d003/d195	GREEK SMALL LETTER SIGMA
<τ>	/d003/d196	GREEK SMALL LETTER TAU
<υ>	/d003/d197	GREEK SMALL LETTER UPSILON
<φ>	/d003/d198	GREEK SMALL LETTER PHI
<χ>	/d003/d199	GREEK SMALL LETTER CHI
<ψ>	/d003/d200	GREEK SMALL LETTER PSI
<ω>	/d003/d201	GREEK SMALL LETTER OMEGA
<ϊ>	/d003/d202	GREEK SMALL LETTER IOTA WITH DIAERESIS
<Ϸ>	/d003/d203	GREEK SMALL LETTER UPSILON WITH DIAERESIS
<ό>	/d003/d204	GREEK SMALL LETTER OMICRON WITH ACUTE
<ύ>	/d003/d205	GREEK SMALL LETTER UPSILON WITH ACUTE
<ώ>	/d003/d206	GREEK SMALL LETTER OMEGA WITH ACUTE
<>	/d003/d207	(This position shall not be used)
<β>	/d003/d208	GREEK SMALL LETTER BETA INITIAL FORM
<ϑ>	/d003/d209	SCRIPT THETA SYMBOL
<Υ̣>	/d003/d210	GREEK CAPITAL LETTER UPSILON WITH HOOK
<Υ̇>	/d003/d211	GREEK CAPITAL LETTER UPSILON WITH ACUTE AND HOOK
<ỵ̈>	/d003/d212	GREEK CAPITAL LETTER UPSILON WITH DIAERESIS AND HOOK
<ϕ>	/d003/d213	SCRIPT PHI SYMBOL
<ω>	/d003/d214	SCRIPT OMEGA SYMBOL
<?>	/d003/d215	GREEK QUESTION MARK

<◊>	/d003/d216	GREEK NUMERAL SIGN
<ι>	/d003/d217	GREEK LOWER NUMERAL SIGN
<Σ>	/d003/d218	GREEK CAPITAL LETTER STIGMA
<σ>	/d003/d219	GREEK SMALL LETTER STIGMA
<Ϝ>	/d003/d220	GREEK CAPITAL LETTER DIGAMMA
<ϝ>	/d003/d221	GREEK SMALL LETTER DIGAMMA
<Ϟ>	/d003/d222	GREEK CAPITAL LETTER KOPPA
<ϟ>	/d003/d223	GREEK SMALL LETTER KOPPA
<Ϡ>	/d003/d224	GREEK CAPITAL LETTER SAMPI
<ϡ>	/d003/d225	GREEK SMALL LETTER SAMPI
<◊>	/d003/d226	GREEK CAPITAL LETTER SHEI
<◊>	/d003/d227	GREEK SMALL LETTER SHEI
<◊>	/d003/d228	GREEK CAPITAL LETTER FEI
<◊>	/d003/d229	GREEK SMALL LETTER FEI
<◊>	/d003/d230	GREEK CAPITAL LETTER KHEI
<◊>	/d003/d231	GREEK SMALL LETTER KHEI
<◊>	/d003/d232	GREEK CAPITAL LETTER HORI
<◊>	/d003/d233	GREEK SMALL LETTER HORI
<◊>	/d003/d234	GREEK CAPITAL LETTER GANGIA
<◊>	/d003/d235	GREEK SMALL LETTER GANGIA
<◊>	/d003/d236	GREEK CAPITAL LETTER SHIMA
<◊>	/d003/d237	GREEK SMALL LETTER SHIMA
<◊>	/d003/d238	GREEK CAPITAL LETTER DEI
<◊>	/d003/d239	GREEK SMALL LETTER DEI
<κ>	/d003/d240	SCRIPT KAPPA SYMBOL
<ρ>	/d003/d241	GREEK SMALL LETTER SCRIPT RHO
<ς>	/d003/d242	GREEK SMALL LETTER LUNATE SIGMA
<◊>	/d003/d243	GREEK TONOS
<◊>	/d003/d244	ACUTE ACCENT AND DIAERESIS (Tonos and Diallytika)
<◊>	/d003/d245	GREEK IOTA BELOW
<◊>	/d003/d246	(This position shall not be used)
...	...	...
<◊>	/d003/d255	(This position shall not be used)

**TABLE 10-ROW 04: CYRILLIC (NAMES ACCORDING TO ISO 10646)**

<◊>	/d004/d000	(This position shall not be used)
<Ё>	/d004/d001	CYRILLIC CAPITAL LETTER IO
<Ђ>	/d004/d002	CYRILLIC CAPITAL LETTER DJE
<Ѓ>	/d004/d003	CYRILLIC CAPITAL LETTER GJE
<Є>	/d004/d004	CYRILLIC CAPITAL LETTER UKRAINIAN IE
<Є>	/d004/d005	CYRILLIC CAPITAL LETTER DZE
<І>	/d004/d006	CYRILLIC CAPITAL LETTER BYELORUSSIAN-UKRAINIAN I
<Ѡ>	/d004/d007	CYRILLIC CAPITAL LETTER YI
<Ј>	/d004/d008	CYRILLIC CAPITAL LETTER JE
<Јб>	/d004/d009	CYRILLIC CAPITAL LETTER LJE
<Нб>	/d004/d010	CYRILLIC CAPITAL LETTER NJE
<Һ>	/d004/d011	CYRILLIC CAPITAL LETTER TSHE

<К>	/d004/d012	CYRILLIC CAPITAL LETTER KJE
<∅>	/d004/d013	(This position shall not be used)
<Ӳ>	/d004/d014	CYRILLIC CAPITAL LETTER SHORT U
<Ц>	/d004/d015	CYRILLIC CAPITAL LETTER DZHE
<А>	/d004/d016	CYRILLIC CAPITAL LETTER A
<Б>	/d004/d017	CYRILLIC CAPITAL LETTER BE
<В>	/d004/d018	CYRILLIC CAPITAL LETTER VE
<Г>	/d004/d019	CYRILLIC CAPITAL LETTER GHE
<Д>	/d004/d020	CYRILLIC CAPITAL LETTER DE
<Е>	/d004/d021	CYRILLIC CAPITAL LETTER IE
<Ж>	/d004/d022	CYRILLIC CAPITAL LETTER ZHE
<З>	/d004/d023	CYRILLIC CAPITAL LETTER ZE
<И>	/d004/d024	CYRILLIC CAPITAL LETTER I
<Ӣ>	/d004/d025	CYRILLIC CAPITAL LETTER SHORT I
<К>	/d004/d026	CYRILLIC CAPITAL LETTER KA
<Л>	/d004/d027	CYRILLIC CAPITAL LETTER EL
<М>	/d004/d028	CYRILLIC CAPITAL LETTER EM
<Н>	/d004/d029	CYRILLIC CAPITAL LETTER EN
<О>	/d004/d030	CYRILLIC CAPITAL LETTER O
<П>	/d004/d031	CYRILLIC CAPITAL LETTER PE
<Р>	/d004/d032	CYRILLIC CAPITAL LETTER ER
<С>	/d004/d033	CYRILLIC CAPITAL LETTER ES
<Т>	/d004/d034	CYRILLIC CAPITAL LETTER TE
<У>	/d004/d035	CYRILLIC CAPITAL LETTER U
<Ф>	/d004/d036	CYRILLIC CAPITAL LETTER EF
<Х>	/d004/d037	CYRILLIC CAPITAL LETTER HA
<Ц>	/d004/d038	CYRILLIC CAPITAL LETTER TSE
<Ч>	/d004/d039	CYRILLIC CAPITAL LETTER CHE
<Ш>	/d004/d040	CYRILLIC CAPITAL LETTER SHA
<Щ>	/d004/d041	CYRILLIC CAPITAL LETTER SHCHA
<Ъ>	/d004/d042	CYRILLIC CAPITAL LETTER HARD SIGN
<Ы>	/d004/d043	CYRILLIC CAPITAL LETTER YERU
<ь>	/d004/d044	CYRILLIC CAPITAL LETTER SOFT SIGN
<Э>	/d004/d045	CYRILLIC CAPITAL LETTER E
<Ю>	/d004/d046	CYRILLIC CAPITAL LETTER YU
<Я>	/d004/d047	CYRILLIC CAPITAL LETTER YA
<a>	/d004/d048	CYRILLIC SMALL LETTER A
<б>	/d004/d049	CYRILLIC SMALL LETTER BE
<в>	/d004/d050	CYRILLIC SMALL LETTER VE
<г>	/d004/d051	CYRILLIC SMALL LETTER GHE
<д>	/d004/d052	CYRILLIC SMALL LETTER DE
<е>	/d004/d053	CYRILLIC SMALL LETTER IE
<ж>	/d004/d054	CYRILLIC SMALL LETTER ZHE
<з>	/d004/d055	CYRILLIC SMALL LETTER ZE
<и>	/d004/d056	CYRILLIC SMALL LETTER I
<ӣ>	/d004/d057	CYRILLIC SMALL LETTER SHORT I
<к>	/d004/d058	CYRILLIC SMALL LETTER KA
<л>	/d004/d059	CYRILLIC SMALL LETTER EL
<м>	/d004/d060	CYRILLIC SMALL LETTER EM
<н>	/d004/d061	CYRILLIC SMALL LETTER EN
<о>	/d004/d062	CYRILLIC SMALL LETTER O
<п>	/d004/d063	CYRILLIC SMALL LETTER PE
<р>	/d004/d064	CYRILLIC SMALL LETTER ER
<с>	/d004/d065	CYRILLIC SMALL LETTER ES
<т>	/d004/d066	CYRILLIC SMALL LETTER TE
<у>	/d004/d067	CYRILLIC SMALL LETTER U

<Ѡ>	/d004/d068	CYRILLIC SMALL LETTER EF
<Ѳ>	/d004/d069	CYRILLIC SMALL LETTER HA
<Ѵ>	/d004/d070	CYRILLIC SMALL LETTER TSE
<Ѷ>	/d004/d071	CYRILLIC SMALL LETTER CHE
<Ѹ>	/d004/d072	CYRILLIC SMALL LETTER SHA
<Ѻ>	/d004/d073	CYRILLIC SMALL LETTER SHCHA
<Ѽ>	/d004/d074	CYRILLIC SMALL LETTER HARD SIGN
<Ѿ>	/d004/d075	CYRILLIC SMALL LETTER YERU
<Ѣ>	/d004/d076	CYRILLIC SMALL LETTER SOFT SIGN
<Ѥ>	/d004/d077	CYRILLIC SMALL LETTER E
<Ѧ>	/d004/d078	CYRILLIC SMALL LETTER YU
<Ѩ>	/d004/d079	CYRILLIC SMALL LETTER YA
⋄	/d004/d080	(This position shall not be used)
<Ѫ>	/d004/d081	CYRILLIC SMALL LETTER IO
<Ѭ>	/d004/d082	CYRILLIC SMALL LETTER DJE
<Ѯ>	/d004/d083	CYRILLIC SMALL LETTER GJE
<Ѱ>	/d004/d084	CYRILLIC SMALL LETTER UKRAINIAN IE
<Ѳ>	/d004/d085	CYRILLIC SMALL LETTER DZE
<Ѵ>	/d004/d086	CYRILLIC SMALL LETTER BYELORUSSIAN-UKRAINIAN I
<Ѷ>	/d004/d087	CYRILLIC SMALL LETTER YI
<Ѹ>	/d004/d088	CYRILLIC SMALL LETTER JE
<Ѻ>	/d004/d089	CYRILLIC SMALL LETTER LJE
<Ѽ>	/d004/d090	CYRILLIC SMALL LETTER NJE
<Ѿ>	/d004/d091	CYRILLIC SMALL LETTER TSHE
<Ѣ>	/d004/d092	CYRILLIC SMALL LETTER KJE
⋄	/d004/d093	(This position shall not be used)
<Ѧ>	/d004/d094	CYRILLIC SMALL LETTER SHORT U
<Ѩ>	/d004/d095	CYRILLIC SMALL LETTER DZHE
⋄	/d004/d096	CYRILLIC CAPITAL LETTER OMEGA
⋄	/d004/d097	CYRILLIC SMALL LETTER OMEGA
⋄	/d004/d098	CYRILLIC CAPITAL LETTER YAT
⋄	/d004/d099	CYRILLIC SMALL LETTER YAT
⋄	/d004/d100	CYRILLIC CAPITAL LETTER IOTIFIED E
⋄	/d004/d101	CYRILLIC SMALL LETTER IOTIFIED E
⋄	/d004/d102	CYRILLIC CAPITAL LETTER LITTLE YUS
⋄	/d004/d103	CYRILLIC SMALL LETTER LITTLE YUS
⋄	/d004/d104	CYRILLIC CAPITAL LETTER IOTIFIED LITTLE YUS
⋄	/d004/d105	CYRILLIC SMALL LETTER IOTIFIED LITTLE YUS
⋄	/d004/d106	CYRILLIC CAPITAL LETTER BIG YUS
⋄	/d004/d107	CYRILLIC SMALL LETTER BIG YUS
⋄	/d004/d108	CYRILLIC CAPITAL LETTER IOTIFIED BIG YUS
⋄	/d004/d109	CYRILLIC SMALL LETTER IOTIFIED BIG YUS
⋄	/d004/d110	CYRILLIC CAPITAL LETTER KSI
⋄	/d004/d111	CYRILLIC SMALL LETTER KSI
⋄	/d004/d112	CYRILLIC CAPITAL LETTER PSI
⋄	/d004/d113	CYRILLIC SMALL LETTER PSI
⋄	/d004/d114	CYRILLIC CAPITAL LETTER FITA
⋄	/d004/d115	CYRILLIC SMALL LETTER FITA
⋄	/d004/d116	CYRILLIC CAPITAL LETTER IZHITSA
⋄	/d004/d117	CYRILLIC SMALL LETTER IZHITSA
⋄	/d004/d118	CYRILLIC CAPITAL LETTER IZHITSA WITH DOUBLE GRAVE ACCENT
⋄	/d004/d119	CYRILLIC SMALL LETTER IZHITSA WITH DOUBLE GRAVE ACCENT



<◇>	/d004/d120	CYRILLIC CAPITAL LETTER UK
<◇>	/d004/d121	CYRILLIC SMALL LETTER UK
<◇>	/d004/d122	CYRILLIC CAPITAL LETTER ROUND OMEGA
<◇>	/d004/d123	CYRILLIC SMALL LETTER ROUND OMEGA
<◇>	/d004/d124	CYRILLIC CAPITAL LETTER OMEGA WITH TITLO
<◇>	/d004/d125	CYRILLIC SMALL LETTER OMEGA WITH TITLO
<◇>	/d004/d126	CYRILLIC CAPITAL LETTER OT
<◇>	/d004/d127	CYRILLIC SMALL LETTER OT

**TABLE 11-ROW 04: CYRILLIC (Priority 2)**

<◇>	/d004/d128	CYRILLIC CAPITAL LETTER KOPPA
...	...	...
<◇>	/d004/d255	(This position shall not be used)

**TABLE 12-ROW 05: ARMENIAN (Priority 2)**

**TABLE 13-ROW 05: HEBREW (Priority 2)**

**TABLE 14-ROW 06: ARABIC (Priority 2)**

**TABLE 15-ROW 06: ARABIC (Priority 2)**

**TABLE 16-ROW 09 to TABLE 27-ROW 10 ARE OF NO INTEREST FOR THE EEC**

**TABLE 28-ROW 10: GEORGIAN (Priority 2)**

**TABLE 29-ROW 1E & TABLE 30-ROW 1E: ADDITIONAL EXTENDED LATIN ARE OF NO INTEREST FOR THE EEC**

**TABLE 31-ROW 1F: GREEK EXTENSIONS (Priority 2)**

**TABLE 32-ROW 1F: GREEK EXTENSIONS (Priority 2)**

**TABLE 33-ROW 20: GENERAL PUNCTUATION (NAMES ACCORDING TO ISO 10646)**

<□>	/d023/d000	EN QUAD
...	...	...
<□>	/d023/d111	NOMINAL DIGIT SHAPES

**TABLE 34-ROW 20: SUPERSCRIPTS AND SUBSCRIPTS (NAMES ACCORDING TO ISO 10646)**

< <sup>0</sup> >	/d023/d112	SUPERSCRIPIT DIGIT ZERO
...	...	...
<◇>	/d023/d159	(This position shall not be used)

**TABLE 35-ROW 20: CURRENCY SYMBOLS, COMBINING DIACRITICAL MARKS FOR SYMBOLS (NAMES ACCORDING TO ISO 10646)**

<€>	/d023/d160	EURO-CURRENCY SIGN
...	...	...
<◇>	/d023/d255	(This position shall not be used)

**TABLE 36-ROW 21: LETTERLIKE SYMBOLS, NUMBER FORMS (Priority 2)**

**TABLE 37-ROW 21: ARROWS (Priority 2)**

**TABLE 38-ROW 22: MATHEMATICAL OPERATORS (NAMES ACCORDING TO ISO 10646)**

<∇>	/d025/d000	FOR ALL
∴		
<≥>	/d025/d127	SUCCEEDS OR EQUIVALENT TO

**TABLE 39-ROW 22: MATHEMATICAL OPERATORS (NAMES ACCORDING TO ISO 10646)**

<≠>	/d025/d128	DOES NOT PRECEDE
∴		
<∴>	/d025/d241	DOWN RIGHT DIAGONAL ELLIPSIS
◊	/d025/d242	(This position shall not be used)
∴		
◊	/d025/d255	(This position shall not be used)

**TABLE 40-ROW 23: MISCELLANEOUS TECHNICAL (Priority 2)**

**TABLE 41-ROW 23: CONTROL PICTURES, OPTICAL CHARACTER RECOGNIT. (Priority 2)**

**TABLE 42-ROW 24: ENCLOSED ALPHANUMERIC (Priority 2)**

**TABLE 43-ROW 25: BOX DRAWING (NAMES ACCORDING TO ISO 10646)**

<→>	/d028/d000	BOX DRAWINGS LIGHT HORIZONTAL
∴		
<⌋>	/d028/d127	BOX DRAWINGS HEAVY UP AND LIGHT DOWN

**TABLE 44-ROW 25: BLOCK ELEMENTS, GEOMETRIC SHAPES (NAMES ACCORDING TO ISO 10646)**

<■>	/d028/d128	UPPER HALF BLOCK
∴		
◊	/d028/d255	(This position shall not be used)

**TABLE 45-ROW 26: MISCELLANEOUS DINGBATS (NAMES ACCORDING TO ISO 10646)**

<☀>	/d029/d000	BLACK SUN WITH RAYS
∴		
<♯>	/d029/d111	MUSIC SHARP SIGN
◊	/d029/d112	(This position shall not be used)
∴		
◊	/d029/d127	(This position shall not be used)

**TABLE 46-ROW 27: DINGBATS (Priority 2)**

**TABLE 47-ROW 27: DINGBATS (Priority 2)**

**TABLE 48-ROW 30 to TABLE 105-ROW FA (CHINESE, JAPANESE, KOREAN) ARE OF NO INTEREST TO THE EEC**

**ANNEX 3 ADDITIONAL INFORMATION CONCERNING ORDERING****1. ESTABLISHING THE RANKING WORDS: PRINCIPLES, HINTS & EXAMPLES**

The ranking words are selected from the entry, most often word by word, but in certain applications special rules are to be applied concerning word priorities. In bibliographic filing for example the ampersand (&) is traditionally considered as equivalent to "and" (or "et", "og", "och", etc. according to the language of the entry); on the other hand name prefixes (like "Sir", "af", "von", "van", "van der") are often neglected as are articles or other attributes (e.g. "General", "International Dept. of", "Statens", "Kungl.", "AB", "Svenska", etc.) when at the beginning of the entry.

The sequence of ranking words within an entry may be established by inversion of word order (William Shakespeare → Shakespeare, William) or by using parentheses or different fonts (Braun, Wernher von → von Braun, Wernher). In that case the ranking words may be established on the basis of formal criteria, possible to implement in the context of computer programs (see ref. 7).

**2. CONVENTIONS APPLYING TO THE CONSTRUCTION OF ORDINAL FORMS**

Ordinal forms should normally be constructed on the basis of ranking words by application of the strict spelling maxim (also called "the literal principle"); it implies that the positioning of the entries in the ordering is entirely based upon spelling - not on pronunciation, meaning, normalized orthography, phonetic form or any other way requiring a manual generation of the corresponding ordinal form.

The compilation of the ordinal form may include some regular language or application-dependent character transliterations.

**TABLE 3A: POSSIBLE CHARACTER TRANSLITERATIONS**

Original character	Equivalent	Remarks
æ, Æ	ae, AE	for all languages except Danish, Norwegian, Icelandic
œ, Œ	oe, OE	for all languages except Norwegian
ch, CH	ch, CH	for all languages except Spanish
ll, Ll	ll, LL	for all languages except Spanish
ij, IJ	ij, IJ	for all languages except Dutch
ß	ss	
þ, Þ	th, TH	for all languages except Icelandic
&	_&_	_ means space
/	_/_	
-	-	
greek script	latin script	According to the ELOT 743 <sup>1</sup>
cyrillic script	latin script	According to ISO 9, 259, DP 2805
arabic script	latin script	According to ISO 233
hebrew script	latin script	According to ISO 259

---

<sup>1</sup> ISO 843.2 is not recommended

any script                      phonetic alphabet                      According to International Phonetic Alphabet<sup>1</sup>  
 greek/roman/arabic numbers      normal digits

### 3. THE ALPHABETS

#### a. Latin-script languages

ENGLISH, GERMAN, FRENCH, ITALIAN, PORTUGUESE

	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z
C	ä		ç		é				î						ö							ü				
	à				è				ï						ô							ù				
	â				ê				ì						ó							û				
	á				ë				í						ò							ú				

SPANISH

	a	b	c	ch	d	e	f	g	h	i	j	k	l	ll	m	n	ñ	o	p	q	r	s	t	u	v	w	x	y	z
C	á					é				í								ó						ú					
																								ü					

SWEDISH

	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	x	y	z	å	ä	ö
C	à				é																	w	ü			æ	ø	

DANISH

	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	x	y	z	æ	ø	å
C					é																	w						

NORWEGIAN

	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	æ	ø	å
C					é																								

ICELANDIC

	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	x	y	z	þ	æ	ö
C	á			ð	é				í						ó						ú	w		ý				
											k <sup>2</sup>				ŋ <sup>3</sup>													

C stands for combinations of characters and diacritical marks; underlined characters are to be considered as VARIANTS for the alphabets concerned.

#### b. Greek-script languages

	α	β	γ	Ϝ	δ	ε	ζ	η	θ	ι	κ	λ	μ	ν	ξ	ο	π	ρ	ς	σ	τ	υ	φ	χ	ψ	ω
C	ά				έ			ή		ί					ό							ύ				ώ
										ϊ												ϋ				
										ί												ϋ				

It may be desired to order latin-script characters in-between greek-script ones. Several schemes

<sup>1</sup> other possible schemes include the International African Alphabet (anglophone), the International Reference Alphabet (francophone) and the Library of Congress transliteration systems; see also ISO/TC 46/SC 4/WG 1 NWI 17.1, 37, 81, 96

<sup>2</sup> greenlandic kra

<sup>3</sup> lappish eng

have been proposed (see ref. 10) but they are to be considered as application-dependent.

c. Cyrillic-script languages To be added later

d. Arabic, etc-script languages To be added later

e. Benchmark To be added later

**ANNEX 5 LIST OF KEYBOARDS**

**Priority 1 Languages**

**SIC S6 (ISO 8884 plus greek) extended for RUSSIAN**

**QWERTY Variants**

DANISH, DUTCH, UK, ITALIAN, NORWEGIAN, PORTUGUESE, SPANISH,  
SWEDISH/FINNISH, EUROPA, US (often the default).

**AZERTY Variants**

FRENCH/BELGIAN, SWISS FRENCH, EUROPA.

**QWERTZ Variants**

GERMAN, SWISS GERMAN.

**GREEK  $\xi\epsilon\rho\tau\upsilon$ /QWERTY**

**RUSSIAN**

**Priority 2 Languages**

**ARMENIAN**

**HEBREW**

**ARABIC**

**GEORGIAN**

## **REFERENCES**

1. Britannica Book of the Year 1990 (Britannica World Data)  
  
"Linguistic composition of the Nations of the World", part 5. Europe and the USSR, by H. Kloss and G. D. McDonnell, Centre International de Recherche sur le Bilinguisme, Les Presses de l'Université Laval, Québec, 1984.
2. ISO 6937 Coded character sets for text communication (1983)
3. ISO DIS 10367 Repertoire of standardized coded graphic character sets for use in 8-bit codes (January 1990)
4. ISO 8859/7 8-bit single byte coded graphic character sets: Latin/greek alphabet (1987) (plus personal greek bibliography)
5. ISO/IEC DIS 10646-1.2 Information technology - Universal Multiple-Octet Coded Character Set (UCS)-Part 1: Architecture and Basic Multilingual Plane (26 December 1991, ISO/IEC JTC1/SC2/WG2 N783)
6. UNICODE Version 1.0; Draft Standard; Final Review Document; The Unicode Consortium; (December 1990)
7. Rolf GAVARE "Alphabetical Ordering in a Lexicographical Perspective; A computational model for exacting alphabetization of dictionary and encyclopedic material" in Studies in Computer-Aided Lexicology, Data linguistica 18. Almqvist & Wiksell International, Stockholm 1988
8. DRAFT Förslag STATSKONTORETS Tekniska NORM Nr 34:1, 1991-02-07
9. 243.4.1-1990 Norme canadienne de classement alphanumérique pour les jeux de caractères de la norme CSA Z243.4, Association Canadienne de Normalisation, version produite par Alain LaBonté, 22 février 1991.
10. POULAIN, J. "Le Grec, Langue Communautaire. Quelques Conséquences", Terminologie et Traduction N° 2-1987
11. EWOS EG ODA/92/3 rev. Implementation requirements for fonts in formatted form documents with attached the EWOS EG ODA/91/53 (P10-48): Font support "Recommended Practices" for ODA formatted form documents conforming to FOD26 or FOD 36.
12. ISO 8884: Information processing - Text and office systems - Keyboards for multiple Latin-alphabet languages - Layout and operation (1989).
13. CEC SIC S6 - Information Processing Standardisation Specification: Standard conformant multilingual Keyboard for the EC Latin-Greek Alphabet (Version 1990 April 24th).
14. European Council in Lisbon (26-27 June 1992) : Conclusions of the Presidency, Brussels 27/6/92, RAPID data-base ref. number DOC/92/3.