INTERNATIONAL STANDARD

ISO 3602

First edition
1989-09-01

Documentation — Romanization of Japanese (*kana* script)

Documentation — Romanisation du japonais (écriture en kana*)*

Reference number
ISO 3602 : 1989 (E)

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

Draft International Standards adopted by the technical committees are circulated to the member bodies for approval before their acceptance as International Standards by the ISO Council. They are approved in accordance with ISO procedures requiring at least 75 % approval by the member bodies voting.

International Standard ISO 3602 was prepared by Technical Committee ISO/TC 46, *Documentation*.

International Organization for Standardization
Case postale 56 • CH-1211 Genève 20 • Switzerland

Printed in Switzerland

Introduction

Standards on conversion of systems of writing

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

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

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

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

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

General principles of conversion of writing systems

Definitions and methods

The words in a language, which are written according to a given script (the converted system), sometimes have to be rendered according to a different system (the conversion system), normally used for a different language. This procedure is often used for historical or geographical texts, cartographical documents and, in particular, bibliographical work where characters must be converted from different writing systems into a single alphabet to allow for alphabetical intercalation in bibliographies, catalogues, indexes, toponymic lists, etc.

It is indispensable in that it permits the univocal transmission of a written message between two countries using different writing systems or exchanging a message the writing of which is different from their own.

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

The two basic methods of conversion of a system of writing are transliteration and transcription.

Transliteration is the process which consists of representing the characters[1] of an alphabetical or syllabic writing by the characters of a conversion alphabet.

In principle, this conversion should be made character by character: each character of the converted graphical system is rendered by only one character of the conversion alphabet, this being the easiest way to ensure the complete and unambiguous reversibility of the conversion alphabet in the converted system.

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

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

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

Transcription is the process whereby the pronunciation of a given language is noted by the system of signs of a conversion language. A transcription system is of necessity based on the orthographical conversions of the conversion language. Transcription is not strictly reversible.

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

To carry out **romanization**, the conversion of non-Latin writing systems to the Latin alphabet, either transliteration or transcription or a combination of the two may be used depending on the nature of the converted system.

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

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

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

---

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

When romanizing a script which has no upper case characters, it is usual to capitalize some words, following national practice.

Principles of conversion for syllabaries

In syllabic systems of writing, the syllabic character, corresponding to the notation of a syllable within a given language, is the significant graphical unit.

A syllabary is the set of syllabic characters in use to write down a language with a syllabic writing system.

A syllabic character can be comprised of only one elementary sign, or of two or more signs which are combined or juxtaposed. A system in which a given elementary sign regularly receives the same value in any character where it appears (exception being taken from the phonetic evolution of the language) allows for a reversible transliteration.

A syllabic character has to be globally converted, taking account of the function of each elementary sign in the whole character, but not sign by sign. So an elementary sign may receive in the conversion system different equivalences, depending on the category of the syllabic character to which it pertains. The transliteration table of a syllabary assigns a biunivocal equivalent to each character, so ensuring a complete reversibility.

If a language using a syllabic system of writing is usually written without rules governing the division between characters and/or words, the conversion system must include such rules taking account of the morphological and grammatical structure of the language.

Documentation — Romanization of Japanese (*kana* script)

1 Scope

This International Standard establishes a system for the romanization of the present-day Japanese written language. Unrestricted application for the system requires that the romanizer possess a detailed knowledge of the language in its present-day written form.

2 Explanations and definitions

Japanese writing is composed of Chinese characters, *kanzi*, and syllabic Japanese script, *kana*. Although *kana* can express every syllable in Japanese, according to the *kanazukai* rule, common Japanese documents mix Chinese characters and *kana*. The way of sharing the task to express a certain idea by *kanzi* and *kana* is governed by the *onkunhyô* table and the *okurigana* rule.

There are two types of *kana*: *hiragana* and *katakana*. Most Japanese words expressed by *kana* employ *hiragana*, and *katakana* is used only for non-Chinese loan words, onomatopoeia and in certain special cases where it is necessary to stress the word. There is a one-to-one correspondence between *hiragana* and *katakana*.

This International Standard refers only to the transcription of *kana* into the Latin alphabet. It gives no direct way to transcribe either *kanzi* or the mixture of *kanzi* and *kana* into the Latin alphabet. Romanizers are expected to know the rules governing the relations between *kanzi* and *kana*.

3 System employed

3.1 The system of romanization empoyed shall be that generally known as *kunreisiki*, as it appears in table 1, table 2, table 3a and table 3b. Owing to some characteristics of the script, this system of conversion is not strictly reversible.

3.2 These tables exclude some special signs expressing dialect and foreign sounds in *kana*.

4 Morpheme boundaries

In certain exceptional cases, two *kana* scripts can be regarded as either forming a digraph denoting one syllable or representing two independent syllables. A train of three *kana* scripts こうし, for example, containing a digraph こう and し, can be interpreted as representing the word *"kôsi"*, meaning "lattice", or *"kousi"*, meaning "calf". In Japanese dictionaries, the separation of a digraph is shown by some mark, e.g. a dot or a hyphen. Thus the above example may be shown by こ・うし for *"kousi"*, and こ うし for *"kôsi"*.

5 General rules of application

5.1 Word division

In all Japanese documents, a sentence in *kanzi* and *kana* is spelt in a sequence without divisions by words, in romanized Japanese texts separation into words is necessary.

5.2 Capitalization

Initial capital letters are used at the beginning of a sentence and for all proper nouns, following national practice.

5.3 Letter "*n*" at the end of a syllable

When preceding a vowel or "*y*" in the same word, an "*n*" (*kana* ん or ン) ending a syllable is followed by an apostrophe; for example, *kan'ô* ("cherry-blossom viewing"), *kin'yû* ("finance"). When the "*n*" initiates a syllable, it is written without an apostrophe; e.g. *kinyû* ("entry"), *kanô* ("possible").

5.4 Doubled consonants

If small-sized っ (character 72 of table 1) is used before a syllable beginning with a consonant (e.g. こ = *ko*), this sign is written slightly to the right of centre (or slightly lower when writing sideways); it is then transcribed by the duplication of that consonant, e.g. がっこう = *gakkô*.

5.5 Long vowels

In *kana* spelling, long vowels are represented by certain digraphs (see table 3a) or trigraphs (see table 3b). There are, however, exceptional cases in *kana* spelling where digraphs do not represent real digraphs but two independent syllables for the reasons given in clause 4. Whenever doubtful, it is recommended to consult a dictionary.

In romanization, long vowels are shown by the addition of a circumflex to the vowel, e.g. a long *o* becomes *ô*.

In borrowed words shown in *katakana*, a lengthening bar (—) is used after the *kana* script, e.g. カー (not カア) = *kâ*, ビール (not ビイル) = *bîru*, and ソース (not ソオス nor ソウス) = *sôsu*.

These bars are always transcribed by a circumflex.

6 Punctuation

Usual Japanese punctuation marks are transcribed as follows:

Japanese marks	Latin marks
。	. (Full stop)
、	, (Comma)
・	- (Hyphen or space)
「	" (Left quotation mark)
」	" (Right quotation mark)
（	((Left parenthesis)
）	) (Right parenthesis)

NOTE — A scheme for stringent transliteration would differ from this transcription system on the following items:

Table 1, characters 26 and 29 would be romanized always as *ha* and *he* respectively.

Table 1, character 45 would be written as *wo*.

Table 1, characters 58 and 59 would be written as *di* and *du* respectively.

Table 2, characters 28, 29 and 30 would be written as *dya*, *dyu* and *dyo* respectively.

In 5.5, the lengthening bar would be transliterated by a macron on the preceding vowel, e.g. *bīru*.

Table 1 — Simple *kana* signs representing non-palatalized syllables

No.	H	K	R	No.	H	K	R	No.	H	K	R	No.	H	K	R	No.	H	K	R
1	あ	ア	a	2	い	イ	i	3	う	ウ	u	4	え	エ	e	5	お	オ	o
6	か	カ	ka	7	き	キ	ki	8	く	ク	ku	9	け	ケ	ke	10	こ	コ	ko
11	さ	サ	sa	12	し	シ	si	13	す	ス	su	14	せ	セ	se	15	そ	ソ	so
16	た	タ	ta	17	ち	チ	ti	18	つ	ツ	tu	19	て	テ	te	20	と	ト	to
21	な	ナ	na	22	に	ニ	ni	23	ぬ	ヌ	nu	24	ね	ネ	ne	25	の	ノ	no
26	は	ハ	ha[1]	27	ひ	ヒ	hi	28	ふ	フ	hu	29	へ	ヘ	he[2]	30	ほ	ホ	ho
31	ま	マ	ma	32	み	ミ	mi	33	む	ム	mu	34	め	メ	me	35	も	モ	mo
36	や	ヤ	ya			—		37	ゆ	ユ	yu			—		38	よ	ヨ	yo
39	ら	ラ	ra	40	り	リ	ri	41	る	ル	ru	42	れ	レ	re	43	ろ	ロ	ro
44	わ	ワ	wa			—				—				—		45	を	ヲ	o[3]
																46	ん	ン	n
47	が	ガ	ga	48	ぎ	ギ	gi	49	ぐ	グ	gu	50	げ	ゲ	ge	51	ご	ゴ	go
52	ざ	ザ	za	53	じ	ジ	zi	54	ず	ズ	zu	55	ぜ	ゼ	ze	56	ぞ	ゾ	zo
57	だ	ダ	da	58	ぢ	ヂ	zi[4]	59	づ	ヅ	zu[4]	60	で	デ	de	61	ど	ド	do
62	ば	バ	ba	63	び	ビ	bi	64	ぶ	ブ	bu	65	べ	ベ	be	66	ぼ	ボ	bo
67	ぱ	パ	pa	68	ぴ	ピ	pi	69	ぷ	プ	pu	70	ぺ	ペ	pe	71	ぽ	ポ	po
								72	っ	ッ	–[5]								

Legend: H = *hiragana*
K = *katakana*
R = romanized

1) Romanized *wa* when used for the grammatical particle so pronounced.

2) Romanized *e* when used for the grammatical particle so pronounced.

3) を (ヲ) is used only for the grammatical particle denoting direct object complement, but romanized *o*.

4) ぢ (ヂ) and づ (ヅ) are used as voiced ち (チ) and つ (ツ) when adjacent to ち (チ) and つ (ツ) respectively or in certain compound words but they are romanized into *zi* and *zu* respectively.

5) Used to double a consonant; see 5.4.

Table 2 — *Kana* digraphs[1] representing palatalized syllables

No.	H	K	R	No.	H	K	R	No.	H	K	R
1	きゃ	キャ	kya	2	きゅ	キュ	kyu	3	きょ	キョ	kyo
4	しゃ	シャ	sya	5	しゅ	シュ	syu	6	しょ	ショ	syo
7	ちゃ	チャ	tya	8	ちゅ	チュ	tyu	9	ちょ	チョ	tyo
10	にゃ	ニャ	nya	11	にゅ	ニュ	nyu	12	にょ	ニョ	nyo
13	ひゃ	ヒャ	hya	14	ひゅ	ヒュ	hyu	15	ひょ	ヒョ	hyo
16	みゃ	ミャ	mya	17	みゅ	ミュ	myu	18	みょ	ミョ	myo
19	りゃ	リャ	rya	20	りゅ	リュ	ryu	21	りょ	リョ	ryo
22	ぎゃ	ギャ	gya	23	ぎゅ	ギュ	ˏyu	24	ぎょ	ギョ	gyo
25	じゃ	ジャ	zya	26	じゅ	ジュ	zyu	27	じょ	ジョ	zyo
28	ぢゃ	ヂャ	zya[2]	29	ぢゅ	ヂュ	zyu[2]	30	ぢょ	ヂョ	zyo[2]
31	びゃ	ビャ	bya	32	びゅ	ビュ	byu	33	びょ	ビョ	byo
34	ぴゃ	ピャ	pya	35	ぴゅ	ピュ	pyu	36	ぴょ	ピョ	pyo

Legend: H = *hiragana*
K = *katakana*
R = romanized

1) These digraphs consist of a normal script and a succeeding smaller や (ヤ), ゆ (ユ), よ (ヨ) written slightly to the right of centre (or slightly lower when writing sideways).

2) ぢゃ (ヂャ), ぢゅ (ヂュ) and ぢょ (ヂョ) are used as voiced ちゃ (チャ), ちゅ (チュ) and ちょ (チョ) respectively in certain compound words but they are romanized *zya*, *zyu* and *zyo* respectively.

Table 3a — *Kana* digraphs representing long-vowel syllables[1)]

No.	H	K	R	No.	H	K	R	No.	H	K	R	No.	H	K	R	No.	H	K	R	No.	H	K	R
1	ああ	アア	â	2	いい	イイ	î	3	うう	ウウ	û	4	ええ	エエ	ê	5	おお	オオ	ô	6	おう	オウ	ô
7	かあ	カア	kâ	8	きい	キイ	kî	9	くう	クウ	kû	10	けえ	ケエ	kê	11	こお	コオ	kô	12	こう	コウ	kô
13	さあ	サア	sâ	14	しい	シイ	sî	15	すう	スウ	sû	16	せえ	セエ	sê	17	そお	ソオ	sô	18	そう	ソウ	sô
19	たあ	タア	tâ	20	ちい	チイ	tî	21	つう	ツウ	tû	22	てえ	テエ	tê	23	とお	トオ	tô	24	とう	トウ	tô
25	なあ	ナア	nâ	26	にい	ニイ	nî	27	ぬう	ヌウ	nû	28	ねえ	ネエ	nê	29	のお	ノオ	nô	30	のう	ノウ	nô
31	はあ	ハア	hâ	32	ひい	ヒイ	hî	33	ふう	フウ	hû	34	へえ	ヘエ	hê	35	ほお	ホオ	hô	36	ほう	ホウ	hô
37	まあ	マア	mâ	38	みい	ミイ	mî	39	むう	ムウ	mû	40	めえ	メエ	mê	41	もお	モオ	mô	42	もう	モウ	mô
43	やあ	ヤア	yâ					44	ゆう	ユウ	yû					45	よお	ヨオ	yô	46	よう	ヨウ	yô
47	らあ	ラア	râ	48	りい	リイ	rî	49	るう	ルウ	rû	50	れえ	レエ	rê	51	ろお	ロオ	rô	52	ろう	ロウ	rô
53	わあ	ワア	wâ																				
54	があ	ガア	gâ	55	ぎい	ギイ	gî	56	ぐう	グウ	gû	57	げえ	ゲエ	gê	58	ごお	ゴオ	gô	59	ごう	ゴウ	gô
60	ざあ	ザア	zâ	61	じい	ジイ	zî	62	ずう	ズウ	zû	63	ぜえ	ゼエ	zê	64	ぞお	ゾオ	zô	65	ぞう	ゾウ	zô
66	だあ	ダア	dâ	67	ぢい	ヂイ	zî	68	づう	ヅウ	zû	69	でえ	デエ	dê	70	どお	ドオ	dô	71	どう	ドウ	dô
72	ばあ	バア	bâ	73	びい	ビイ	bî	74	ぶう	ブウ	bû	75	べえ	ベエ	bê	76	ぼお	ボオ	bô	77	ぼう	ボウ	bô
78	ばあ	パア	pâ	79	びい	ピイ	pî	80	ぶう	プウ	pû	81	ぺえ	ペエ	pê	82	ぽお	ポオ	pô	83	ぽう	ポウ	pô

Legend: H = *hiragana*
K = *katakana*
R = romanized

1) When there is a morpheme boundary between vowels, they should not be rewritten in the form of a long vowel with a circumflex accent, even when they are similarly pronounced.

Examples:

ながあめ	nagaame	くう	kuu
いいん	iin	いう	iu
きいろ	kiiro	くされえん	kusareen
ほしい	hosii	おもう	omou

Table 3b — *Kana* trigraphs[2] representing palatalized long-vowel[1] syllables

No.	H	K	R	No.	H	K	R	No.	H	K	R
1	きゃあ	キャア	kyâ	2	きゅう	キュウ	kyû	3	きょう	キョウ	kyô
4	しゃあ	シャア	syâ	5	しゅう	シュウ	syû	6	しょう	ショウ	syô
7	ちゃあ	チャア	tyâ	8	ちゅう	チュウ	tyû	9	ちょう	チョウ	tyô
10	にゃあ	ニャア	nyâ	11	にゅう	ニュウ	nyû	12	にょう	ニョウ	nyô
13	ひゃあ	ヒャア	hyâ	14	ひゅう	ヒュウ	hyû	15	ひょう	ヒョウ	hyô
16	みゃあ	ミャア	myâ	17	みゅう	ミュウ	myû	18	みょう	ミョウ	myô
19	りゃあ	リャア	ryâ	20	りゅう	リュウ	ryû	21	りょう	リョウ	ryô
22	ぎゃあ	ギャア	gyâ	23	ぎゅう	ギュウ	gyû	24	ぎょう	ギョウ	gyô
25	じゃあ	ジャア	zyâ	26	じゅう	ジュウ	zyû	27	じょう	ジョウ	zyô
28	ぢゃあ	ヂャア	zyâ	29	ぢゅう	ヂュウ	zyû	30	ぢょう	ヂョウ	zyô
31	びゃあ	ビャア	byâ	32	びゅう	ビュウ	byû	33	びょう	ビョウ	byô
34	ぴゃあ	ピャア	pyâ	35	ぴゅう	ピュウ	pyû	36	ぴょう	ピョウ	pyô

Legend: H = *hiragana*
K = *katakana*
R = romanized

1) When there is a morpheme boundary between vowels, they should not be rewritten in the form of a long vowel with a circumflex accent, even when they are similarly pronounced.

Examples:				
	なが あめ	nagaame	く う	kuu
	いいん	iin	い う	iu
	きいろ	kiiro	くされえん	kusareen
	ほしい	hosii	おもう	omou

2) These *kana* trigraphs consist of a normal script and a subscript, followed by a third normal script.

Examples of romanization:

おかあさん	okâsan	おおきい	ôkii
くうき	kûki	おとうさん	otôsan
ねえさん	nêsan	こうぎょう	kôgyô

UDC 003.324.2.035

Descriptors : *documentation, alphabets, character sets, Japanese characters, transliteration.*

Price based on 6 pages