Title:	Comments on N2626, Proposal on IPA Extensions & Combining Diacritic Marks for ISO/IEC 10646 in BMP
Doc. Type:	Expert contribution
Source:	Peter Constable, Microsoft
Date:	October 8, 2003
Action:	For consideration by JTC1/SC2/WG2, UTC
References:	WG2/N2626 (=L2/03-317), N2307 (= L2/00-421), N2312 (= L2/01-025), N2623 (= L2/03-326)
Distribution:	WG2 members, UTC members

Background

In document N2626, the Chinese National Body has proposed the addition of 320 characters for phonetic symbols. This repertoire is intended to complement the phonetic symbols already in the UCS in order to provide complete coverage of symbols used by linguists in China.

The proposed set of characters is quite mixed in terms of justification for encoding. There are characters that definitely should be added, though N2626 does not provide any documentation demonstrating attestation and usage; there are characters that definitely should not be encoded as the text elements already can be represented in the UCS; and there are characters that are difficult to evaluate in the absence of further documentation regarding intended usage.

This document will attempt to evaluate the different characters proposed in N2626.

Use of combining marks versus non-combining letters or symbols

Many of the characters proposed are combining marks. Five of these are clearly non-spacing marks, as shown in Figure 1:[1]

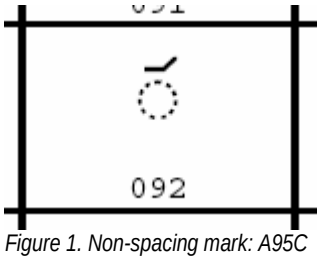


Figure 1. Non-spacing mark: A95C

Many are clearly spacing marks, as shown in Figure 2:

[1] In this document, I will cite characters from N2626 using the code positions in the range U+A900..U+AA3F as used in N2626.

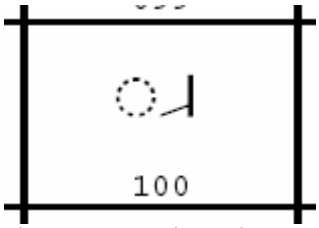


Figure 2. Non-spacing mark: A964

Some appear to be spacing marks, though in the absence of further documentation on usage by Chinese linguists, it is not certain that their positioning relative to the base symbol is not such that they might be considered non-spacing:

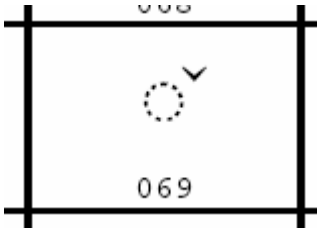


Figure 3. Combining mark, probably spacing: A945

It seems very unlikely that these should be considered non-spacing, and I will assume that all of these are, in fact, spacing, with one exception, A940, which will be discussed below.

There is an important issue that relates to the spacing marks, but not the non-spacing marks. (The non-spacing marks will be discussed in the next section.) The proposed spacing combining marks include the following:

> A90A..A911, A913..A940, A942..A95B, A964..A96D, A978..A981, A98C..A9A9, A9C8..A9E5

N2626 has proposed these characters as spacing combining marks, and has not discussed the possibility that these might alternately be non-combining letters or symbols. There are many reasons why most, and likely all, of these should be modifier letters rather than spacing combining marks. Note the following observations:

- Several of the proposed spacing marks are already encoded in the UCS as non-combining letters or symbols (details below).
- The only spacing combining marks currently in the UCS are marks used in Indic scripts for syllable-rhymes or other syllable modifications, and a small number musical symbols.
- N2626 itself is inconsistent in its handling of Chao tone letters, using combining marks for right-stemmed tone letters but non-combining letters for the right- and left-stemmed tone letter pairs used in indicated tone sandhi.[2]

[2] See, however, footnote 13.

Table 1 lists many of the proposed spacing marks that appear to be already encoded in the UCS as non-combining letters or symbols; this table shows the code position of the proposed character along with what appears to be (in the absence of additional information) an equivalent character in the UCS. This list does not include certain symbols that will be discussed further below, or Chao tone letters, which are discussed in a separate section.

Proposed character	Existing UCS character
A90A	207F SUPERSCRIPT LATIN SMALL LETTER N
A90B	02E1 MODIFIER LETTER SMALL L
A90C	02B1 MODIFIER LETTER SMALL W
A90D	02B2 MODIFIER LETTER SMALL J
A90E	02E0 MODIFIER LETTER SMALL GAMMA
A90F	02E4 MODIFIER LETTER SMALL REVERSED GLOTTAL STOP
A910	02B6 MODIFIER LETTER SMALL TURNED R
A911	02B0 MODIFIER LETTER SMALL H
A918	02ED MODIFIER LETTER UNASPIRATED
A91A	02DC SMALL TILDE
A91C	02C8 MODIFIER LETTER VERTICAL LINE
A91D	02CC MODIFIER LETTER LOW VERTICAL LINE
A92C	02BB MODIFIER LETTER TURNED COMMA
A92D	02BD MODIFIER LETTER REVERSED COMMA
A93C	02D3 MODIFIER LETTER CENTRED LEFT HALF RING
A93D	02D2 MODIFIER LETTER CENTRED RIGHT HALF RING
A946	02C6 MODIFIER LETTER CIRCUMFLEX ACCENT
A949	02D4 MODIFIER LETTER UP TACK
A94A	02D5 MODIFIER LETTER DOWN TACK
A94D	207A SUPERSCRIPT PLUS SIGN
A94E	02D6 MODIFIER LETTER PLUS SIGN
A94F	02E3 MODIFIER LETTER SMALL X
A950	02DE MODIFIER LETTER RHOTIC HOOK

Table 1. Proposed spacing marks already encoded as non-combining letters or symbols

One proposed character that might be construed as a spacing mark requires special discussion: A940 COMBINING LEFT ANGLE ABOVE RIGHT. This character is, in fact, already encoded in the UCS as U+031A COMBINING LEFT ANGLE ABOVE.[3] Hence, a new character for this symbol is not required.

The character A92F is very similar to U+003A COLON; treated as a non-combining symbol, it is identical, at least in appearance. The authors of this proposal indicated in the proposal summary form (C.10) that none of the proposed characters can be considered

[3] It is understandable that the authors might propose a new character in this case as the code charts for Unicode 3.0 and Unicode 4.0 use a representative glyph for this character that has incorrect positioning, with the mark directly above rather than above right. This issue was addressed in an erratum on the Unicode Web site, at http://www.unicode.org/errata/index.html (date of erratum: 2003-5-23).

similar to an existing UCS character, but clearly that is not the case. There may be valid grounds for proposing a distinct character (e.g. based on the effects of character properties on text processes such as line breaking), but the proposal should, at least, mention the similarity of A92F with U+003A, and provide at least some justification for encoding a distinct character.

Several of the spacing marks in this proposal come in pairs positioned on the right and left of the base character and that are non-mirroring (the same outline is used on both sides), as illustrated in Figure 4:

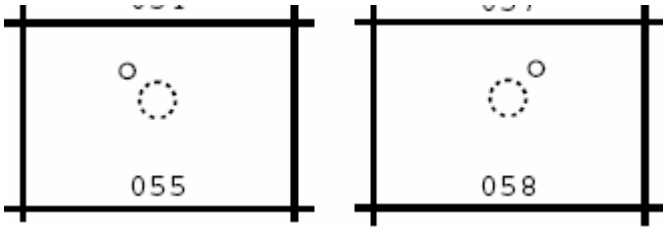


Figure 4. Non-mirroring left and right spacing mark pairs: A937, A93A"

If these are, indeed spacing marks, then if re-analyzed as non-combining symbols, they would be indistinguishable, and so could be unified into a single character.[4]

Table 2 shows the pairs of non- mirroring left and right marks that, it appears (in the absence of information indicating otherwise), could be unified if treated as non-combining letters or symbols. In most of these cases,[5] it appears that the unified character already exists in the UCS, and this is also shown in Table 2.

Non-reflective spacing mark pair from N2626	Suggested equivalent non-combining character in UCS
A913, A916	02C9 MODIFIER LETTER MACRON
A914, A917	02D7 MODIFIER LETTER MINUS SIGN
A930, A936	02D9 DOT ABOVE
A937, A93A	02DA RING ABOVE
A939, A93B	02F3 MODIFIER LETTER LOW RING
A942, A945	02C7 CARON
A931, A932	

Table 2. Proposed non- mirroring left/right paired spacing marks to be unified as non-combining symbols

The characters A920..A928 may also include other such pairs, but I am unfamiliar with their usage in Chinese linguistics, and it is unclear what the functions of these characters are. Hence, these characters cannot be evaluated until further information is available, including illustrations of their usage in actual documents.

[4] The only grounds that might exist for treating these non-mirroring paired characters as combining marks rather than non-combining letters or symbols would be if their position relative to the base symbol is such that they would *not* be indistinguishable if treated as non-combining letters and symbols.

[5] I would certainly contend that A931 and A932 should be unified with one another. Without further information, it is difficult to know whether they could be represented by an existing UCS character, such as U+002E FULL STOP.

Assuming that the analysis of A913 and A914 in Table 2 is correct, it then appears acceptable to conclude that A915 is already represented in the UCS as U+02CD MODIFIER LETTER LOW MACRON.

The proposal also includes some *mirroring* left/right pairs of spacing marks, as illustrated in Figure 5:

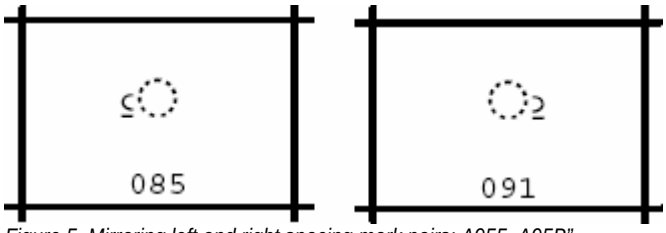
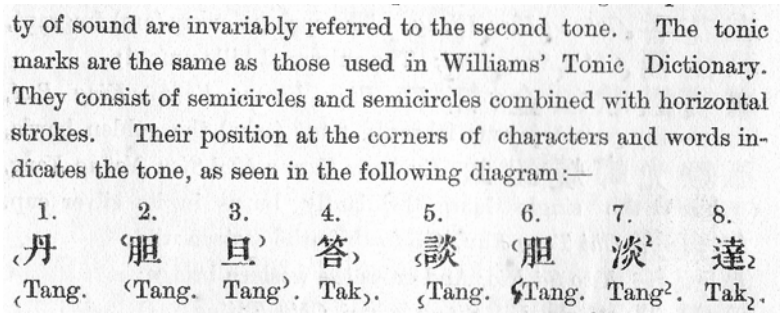


Figure 5. Mirroring left and right spacing mark pairs: A955, A95B"

There are four of these pairs (eight characters in all): A954..A95B. Clearly, these cannot be unified. They can, however, be treated as non-combining symbols. In fact, there is a very significant reason why it would be preferable *not* to treat these as combining marks: to do so would result in re-ordering combining marks (they appear to the left of the base character in left-to-right text), which would necessitate complex rendering processing to support these characters, as well as the usability issues associated with editing text that has a visual sequencing different from the logical sequencing. These issues would not arise if the characters were not combining marks.

The need for these characters is demonstrated by attestations such as the following:[6]


ty of sound are invariably referred to the second tone. The tonic marks are the same as those used in Williams' Tonic Dictionary. They consist of semicircles and semicircles combined with horizontal strokes. Their position at the corners of characters and words indicates the tone, as seen in the following diagram:—

1. 2. 3. 4. 5. 6. 7. 8.
˛丹 ˚胆 旦˙ 答˙ ˛談 ˚胆 淡˛ 達˛
˛Tang. ˚Tang. Tang˙. Tak˙. ˛Tang. ˚Tang. Tang˛. Tak˛.

Figure 6. Yin/yang tone modifiers (Baldwin 1871, p. 4)

Although this sample is not recent, these symbols are still in use by linguists that study Chinese languages.[7] Hence, I support the proposal to add these characters to the UCS, though as non-combining symbols rather than as spacing combining marks. Also, the names proposed for these characters reflects their geometric form, but there are names for these symbols that are in reasonably wide usage. Thus, I suggest the following names for these characters:

A954 ˛ MODIFIER LETTER CHINESE TONE YIN PING

[6] Note that there is a typographic error in this text: the text is specifically intending to introduce all eight symbols, but the sixth symbol is mistakenly typeset using the second symbol.

[7] Lon Diehl, personal communication.

A955 ˂ MODIFIER LETTER CHINESE TONE YIN SHANG

A956 ˃ MODIFIER LETTER CHINESE TONE YIN QU

A957 ˩ MODIFIER LETTER CHINESE TONE YIN RU

A958 ˧ MODIFIER LETTER CHINESE TONE YANG PING

A959 ˅ MODIFIER LETTER CHINESE TONE YANG SHANG

A95A ˀ MODIFIER LETTER CHINESE TONE YANG QU

A95B ˨ MODIFIER LETTER CHINESE TONE YANG RU

The remaining spacing marks that I have not discussed are the following:

A919, A91B, A91E, A91F, A929..A92B, A92E, A932..A935, A938, A93E, A943, A944, A947, A948, A94B, A94C, A951..A953.

These have no particular issues to mention, apart from the fact that the authors should provide additional information regarding these characters, including function and illustrative samples—as should be done for all characters in the proposal. Attestation for A91E, A91F and A931 (unified with A932—see above) is illustrated in Figure 7:[8]

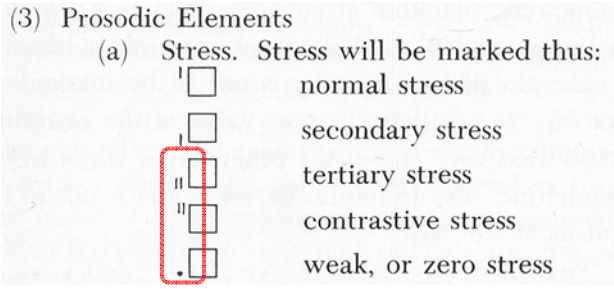


Figure 7. Various modifiers: A91E, A91F, A931/A932 (Chao 1968, p. xxiii)

I do not have sources to illustrate attestation of the other characters at this time, however.

Proposed non-spacing combining marks

There are five non-spacing combining marks proposed in N2626:

A912, A941, A95C..A95E[9]

Without further information about usage by Chinese linguists, I can only speculate that A912 is used to indicate labialization. The existing character U+032B COMBINING INVERTED DOUBLE ARCH BELOW "◌̪" is used to indicate labialization, and as it has a similar appearance, it could be suggested that these are glyph variants. I consider there to be enough difference between these shapes to warrant a separate character. The proposal should, however, discuss this issue and provide some justification for encoding a distinct character.

---

[8] In this sample, the author is indicating the position in which these symbols occur, to the left of the beginning of a syllable. To achieve this visual presentation does not require or present a case for these characters being treated as combining marks, however.

[9] See also the discussion of A940 in the previous section.

There is a problem with the character A941, COMBINING DOUBLE INVERTED BREVE BELOW in that the name does not match the representative glyph shown in the chart:

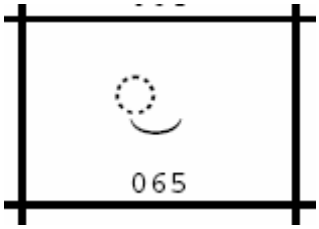


Figure 8. A941: "COMBINING DOUBLE INVERTED BREVE BELOW"

The glyph shown has the regular, upright orientation for breve, not the inverted orientation. This raises a question as to which was intended: inverted or non-inverted. There is a known need for the latter, but I am not aware of any need for the former. Thus, I assume that the glyph is correct and that the name should be COMBINING DOUBLE BREVE BELOW.

Given that correction, this character has been approved by UTC for addition to Unicode at U+035C.[10] This character is included in N2623.

The characters A95C..A95E are IPA-approved diacritics used as an alternate means to indicate contour tones. The need for these characters can be easily verified, as shown in the following sample from the IPA Handbook (IPA 1999):

◌̌	Wedge; háček	Rising contour	ă	524	030C
◌̂	Circumflex	Falling contour	â	525	0302
◌̋	Macron plus acute accent	High rising contour	á	526	------
◌̏	Grave accent plus macron	Low rising contour	à	527	------
◌᷄	Grave plus acute plus grave accent	Rising-falling contour	a᷄	528	------

Figure 9. IPA-approved tone diacritics (IPA 1999, p. 184)"

It might be suggested by some, particularly given the descriptions shown in Figure 9, that these can be considered ligatures of the combining marks acute, grave and macron. That would not be a good encoding model, however, as it would create a need to introduce mechanisms to control whether combinations of these are displayed as ligatures or as stacking diacritics, and given that such control mechanisms would likely involve ZERO WIDTH JOINER or ZERO WIDTH NON JOINER, there would be problems involved in using these control characters with combining marks. As a result, I support the proposal to add these three characters, as presented in N2626, to the UCS.

While these three tone diacritics are the only three cited in the IPA Handbook, there are three additional tone diacritics of this sort that are also used by linguists: macron-grave "◌̄̀", acute-macron "◌̄́" and acute-grave-acute "◌̀́". I am aware of the latter being used, but

[10] This is documented on the Unicode Web site at http://www.unicode.org/alloc/Pipeline.html.

do not have samples to verify this at this time. The following figures demonstrate attestation of the other two, however:

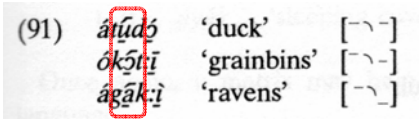


Figure 10. Non-IPA tone diacritics: acute-macron (Gilley 1992, p. 49)"

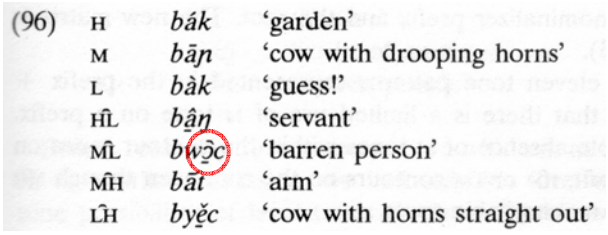


Figure 11. Non-IPA tone diacritics: macron-grave (Gilley 1992, p. 51)"

This raises an issue with regard to the names proposed for the tone diacritic characters in N2626:

A95C combining right dull angle above

A95D combining left dull angle above

A95E combining inverted tidle

The names "right dull angle" and "left dull angle" are insufficiently explicit to accommodate the characters illustrated in Figure 10 and Figure 11 as well as those proposed in N2626. Also, the suggested name for A95E is inappropriate as the character is not an inverted tilde: the angular shape is distinct from that of tilde. Thus, I suggest the following names for the tone diacritics proposed in N2626, and for the other three tone diacritic symbols that should also be included in a proposal:

"◌͒"A95C COMBINING MACRON-ACUTE

"◌͘"A95D COMBINING GRAVE-MACRON

"◌͗"A95E COMBINING GRAVE-ACUTE-GRAVE

"◌̀" COMBINING MACRON-GRAVE

"◌́" COMBINING ACUTE-MACRON

"◌̃" COMBINING ACUTE-GRAVE-ACUTE

Chao tone letters

The majority of characters in this proposal—225 out of 320—are for symbols known as *Chao tone letters*, after the Chinese linguist, Chao Yuen Ren, who first introduced them. Five Chao tone letters are already encoded in the UCS:

U+ 02E5 ˥ MODIFIER LETTER EXTRA-HIGH TONE BAR

U+02E6 ˦ MODIFIER LETTER HIGH TONE BAR

U+02E7 ˧ MODIFIER LETTER MID TONE BAR

U+02E8 ˨ MODIFIER LETTER LOW TONE BAR

U+02E9 ˩ MODIFIER LETTER EXTRA-LOW TONE BAR

Just as N2626 inappropriately proposes spacing combining marks that duplicate existing non-combining letters and symbols, as discussed above, it similarly proposes five combining marks, AA04..AA08, that would duplicate the five tone letters already in the UCS. As in the previous cases, there is no valid reason for introducing duplicate characters as combining marks.

Of the 225 proposed tone letters, 100 are for contour tones:

A964..A96D, A978..A981, A98C..A9A9, A9C8..A9E5, AA0E..AA17, AA22..AA2B

These are proposed as combining marks, and the arguments for treating these as non-combining symbols rather than combining marks apply here as above. The more serious issue for these characters, however, is that these should all be considered presentation forms: these are ligatures, and can be represented as sequences using various combinations of the five UCS characters listed above. This is illustrated by the examples in Table 3:

Contour tone letter	Proposed character	UCS character sequence
ꕀ	A966	02E9 ˩ + 02E6 ˦
ꖁ	A9A1	02E5 ˥ + 02E9 ˩ + 02E7 ˧
ꗛ	A9DB	02E7 ˧ + 02E6 ˦ + 02E9 ˩
꘤	AA24	02E7 ˧ + 02E8 ˨

Table 3. Representation of contour tone letters as sequences of UCS characters

This representation for contour Chao tone letters has been described in *The Unicode Standard* since at least version 2.0,[11] and is discussed in greater length in documents N2307 and N2312. Thus, these 100 proposed contour-tone characters should be rejected.

The proposal also includes 110 (non-combining) characters for tone letters used in indicating tone sandhi.[12] In Chinese linguistics, utterances in which tone sandhi occurs are sometimes transcribed using paired tone letters: one right-stemmed tone letter on the left, indicating the underlying tone, and a left-stemmed tone letter on the right, indicating the surface "sandhi" tone. This is illustrated in Figure 12:

[11] See *The Unicode Standard, Version 2.0*, p. 6-13; or, *The Unicode Standard, Version 3.0*, p. 178; or *The Unicode Standard, Version 4.0*, p. 185.

[12] *Sandhi* is a linguistics term referring to the modification of a speech sound when juxtaposed with other sounds. Thus, *tone sandhi* refers to a change in tone, and is typically conditioned by the tone or stress of surrounding syllables.

Here tone markers on the left of the vertical bar indicate inherent tones; those on the right indicate tones in sandhi, which can be represented as ˩˥˧˨˩ . It can be seen that the tones of the second and the tones of the second and third syllables are a little bit higher than that of the first syllable. However, Xu et al. (1982:128) also presented the tone contour of EXY compounds as ˩˥˩ , for example,

六月天 ɦlo? ˨˩ yɪ? ˨˩ t'i ˥˩ 'Weather typical of June'

白木耳 ba? ˨˩ mo? ˨˧ əl ˨˩ 'tremella'.

Figure 12. Words with tone sandhi transcribed using right- and left-stemmed tone letters (Shunde 1986, p. 26)"

The proposal in N2626 is to treat each of these right- and left-stemmed tone letter pairs as an atomic character:

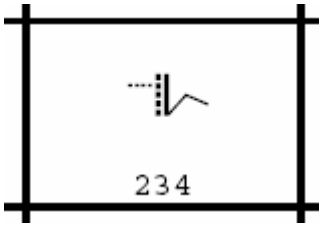


Figure 13. Proposed digraph characters for sandhi tone: A9EA

Two details should be noted in relation to these characters. First, the left-hand component in each sandhi character uses dotted strokes. The sample in Figure 12 demonstrates that linguists do not always use dotted strokes. That is not a significant issue, however, as the dotted versus solid distinction can be considered a font variation—I am not aware of any semantic significance applied to the dotted-stroke forms.[13]

Secondly, all of the characters in the proposal use exactly the same left-hand component, using the shape that, on its own, would be used to indicate level high tone. I find it surprising that this would be the case as the available evidence indicates that linguists use various forms, including contour forms, for the left-hand component of a tone-sandhi transcription, as can be seen in Figure 12 and Figure 14.

[13] After completing the initial draft of this document, it occurred to me that the dotted-stroke forms may have been used by the authors to indicate that these left-stem tone letters are combining marks, and that they only combine with the rightt-stemmed tone characters, proposed in N2626 also as combining marks. If that is the case, portions of my comments are not applicable (though other comments would be warranted: the need for two levels of spacing combining marks only highlights the fact that these should really be treated as non-combining characters). My conclusions, however, still apply: I recommend encoding all Chao tone letters as non-combining characters, and that contour letters be considered ligatures encoded as sequences using combinations of five right-stemmed or five left-stemmed level tone letters.

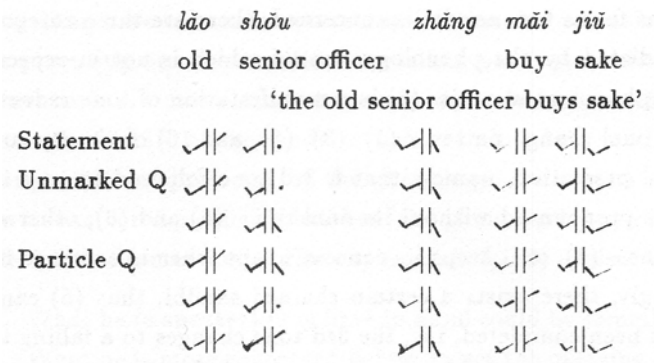


Figure 14. Transcription of sandhi tone (Shen 1989, p. 51)

Thus, it seems appropriate to assume a need for the left-hand component to take any level or contour shape that might otherwise be required when used for non-sandhi tones. Encoding sandhi combinations as atomic characters, though, this would require over 10,000 characters.

It has been shown, however, that any of the non-sandhi tones, level or contour, can be represented using merely the five existing right-stemmed tone letter characters in the UCS. These can be used to represent the left-hand component of sandhi tone transcriptions. Similarly, the right-hand components could also be represented using five left-stemmed tone letters.

Thus, I concur with the need to represent sandhi tone combinations, but I propose, rather than using the 110 atomic sandhi characters proposed in N2626, that this be done using sequences of the five existing right-stemmed tone letter characters in the UCS in combination with sequences of five left-stemmed tone letter characters. Accordingly, I recommend that the proposal be revised to replace those 110 characters with the following:

- ˥ MODIFIER LETTER EXTRA-HIGH LEFT-STEM TONE BAR
- ˦ MODIFIER LETTER HIGH LEFT-STEM TONE BAR
- ˧ MODIFIER LETTER MID LEFT-STEM TONE BAR
- ˨ MODIFIER LETTER LOW LEFT-STEM TONE BAR
- ˩ MODIFIER LETTER EXTRA-LOW LEFT-STEM TONE BAR

Sandhi combinations would then be represented using sequences as illustrated in the following examples:

Sandhi combination	Proposed character	UCS character sequence
˥˦	AA1F	02E6 ˥ + ˦ + ˦
˥˧˩	A9B0	02E6 ˥ + ˦ + ˩ + ˦
˧˥	—	02E7 ˧ + 02E6 ˥ + 02E9 ˩ + ˩ + ˦
˧˩	—	02E7 ˧ + 02E8 ˨ + ˨ + ˩ + ˨

Table 4. Representation of tone sandhi as sequences of level tone characters

Not only will encoding sandhi combinations as right-stemmed and left-stemmed sequences accommodate various forms for the left-hand component in a sandhi combination, it will also simplify supporting sandhi combinations involving dot tone letters, which I will now discuss.

The remaining ten Chao tone characters in this proposal are for dot tone letters: five levels with both right- and left-stemmed forms.

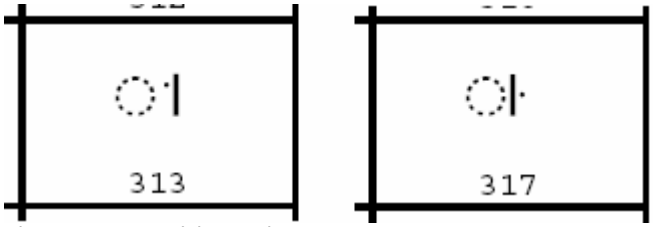


Figure 15. Proposed dot tone letters: AA39, AA3D

Dot tone letters are used in Chinese linguistics to indicate tones in certain weakly-stressed syllables having a less-distinct quality—there is little or no pitch variation, and the duration is short. These are often referred to in Chinese linguistics as “neutral tones”:

The pitch of the neutral tone is:

˩ half-low	after 1st Tone:	他的	*ta.de* ‘his’
˧ middle	after 2nd Tone:	黄的	*hwang.de* ‘yellow one’
˦ half-high	after 3rd Tone:	你的	*nii.de* ‘yours’
˩ low	after 4th Tone:	大的	*dah.de* ‘big one’

Figure 16. Dot tone letters for “neutral” tones (Chao 1968, p. 36)

Dot tone letters can occur on their own with right-stemmed forms, as seen in Figure 16 and also in Figure 17:

In the case of two or more consecutive neutral tones, Chao (1933) and Qi have proposed that the pitch of each depends on the preceding one. As a result of this chained dependence, from the 2nd neutral tone on, all tones are low. For example:

dǎ ˨˦ *ban* ˦ ‘to dress up’

dǎ ˨˦ *ban* ˦ *le* ˩ ‘to have dressed up’

dǎ ˨˦ *ban* ˦ *le* ˩ *mei* ˩ *you* ˩ ‘to have or have not dressed up’

Figure 17. Right-stemmed dot tone letters (Shen 1984, p. 40)

They can also occur in sandhi combinations with left-stemmed forms, as seen in Figure 14, above.

Hence, I support the proposal to add these ten dot tone letters to the UCS. I suggest, however, the following names in order to provide greater consistency with the names for bar tone letters:

˥ MODIFIER LETTER EXTRA-HIGH TONE DOT

˥ MODIFIER LETTER HIGH TONE DOT

˦ MODIFIER LETTER MID TONE DOT

˧ MODIFIER LETTER LOW TONE DOT

˨ MODIFIER LETTER EXTRA-LOW TONE DOT

˩ MODIFIER LETTER EXTRA-HIGH LEFT-STEM TONE DOT

˩ MODIFIER LETTER HIGH LEFT-STEM TONE DOT

˩ MODIFIER LETTER MID LEFT-STEM TONE DOT

˩ MODIFIER LETTER LOW LEFT-STEM TONE DOT

˩ MODIFIER LETTER EXTRA-LOW LEFT-STEM TONE DOT

Conclusion

The characters proposed in N2626 are quite varied in terms of their acceptability and justification for encoding in the UCS. Some are definitely valid proposals needed by existing user communities, although additional documentation is needed. Some of the characters should be encoded, though only after the proposal has been modified to treat them as non-combining letters or symbols rather than combining marks, only after certain unifications have taken place, or only after the proposed names have been revised. Again, additional information should be provided, particularly illustrative samples.

This proposal also includes many characters, particularly contour tone symbols, that can already be encoded in the UCS using existing characters.

Significant revision of this proposal is, therefore, recommended.

References

Baldwin, C.C. 1871. *Manual of the Foochow dialect.* Foochow: Methodist Episcopal Mission Press.

Chao, Yuen Ren. 1968. *A grammar of spoken Chinese.* Berkeley, CA: University of California Press.

Gilley, Leoma G. 1992. *An autosegmental approach to Shilluk phonology.* (Summer Institute of Linguistics and The University of Texas at Arlington publications in linguistics, 103.) Dallas: Summer Institute of Linguistics and University of Texas at Arlington.

International Phonetic Association. 1999. *Handbook of the International Phonetic Association: a guide to the use of the International Phonetic Alphabet.* Cambridge: University of Cambridge Press.

Shen, Xiao-nan Susan. 1989. *The prosody of Mandarin Chinese.* (University of California publications in linguistics, 118.) Berkeley, CA: University of California Press.

Shunde, Jin. 1986. *Shanghai morphotonemics*. Bloomington, IA: Indiana University Linguistics Club.