

REMARKS ON THE USE OF ZWJ & ZWNJ IN THE BRAHMI AND PERSO-ARABIC FAMILIES

The use of ZWJ/ZWNJ is at two levels:

1. EMBEDDED WITHIN THE OPEN TYPE FONT

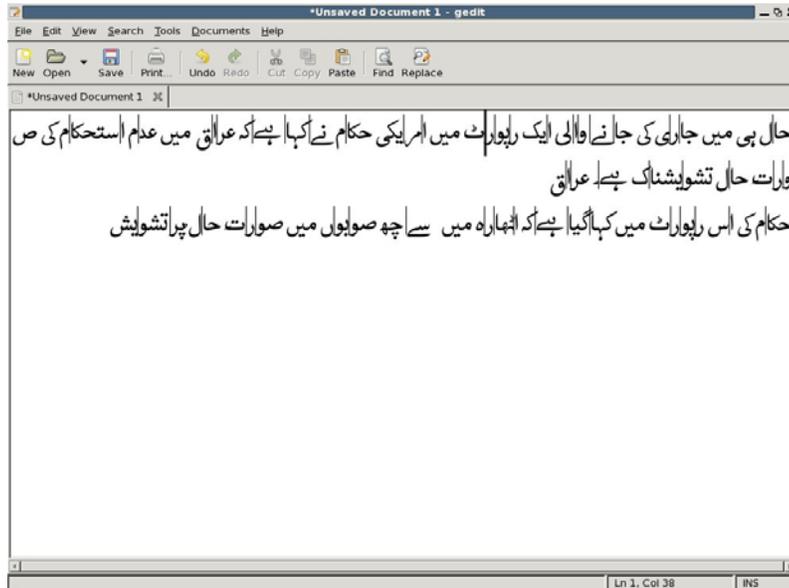
In the creation of Open Type Fonts where the ZWJ / ZWNJ are embedded within the font rules. This is especially used in the case of Urdu nastaaliq

نستعلیق

where the need has been felt to join the “tarsimey” or glyph slices.

نستعلیق

It should be noted that when the font is used under Windows XP, it renders perfectly. However under LINUX and FEDORA the ZWJ shows clear slices.



This is because QT/ICU does not support ZWJ/ZWNJ. IBM's typographer responsible for ICU has been contacted but a new update will take time. Removing ZWJ/ZWNJ from the font gives relatively good results on LINUX and FEDORA.

MAC has no problems with ZWJ/ZWNJ, embedded within the font..

This case is truly not germane to PRI 96, but is mentioned here to show cross-platform compatibility issues.

2. ZWJ/ZWNJ AT THE EXPLICIT LEVEL

2.1. A Historical Overview

Insofar as Brahmi based scripts are concerned ZWJ/ZWNJ seem to be the a take off on the INVISIBLE CONSONANT INV, as well as on the notions of Explicit Halanta¹ and Soft Halanta. See Appendix below

INV as well as Explicit and Soft halanta had two functions in ISCII.

A. In the first place they acted as display variants.

INV allowed for the creation of composite characters which require a consonant base but where the consonant itself ought to be invisible. Thus typing

क + HALANTA+INV yields the same result as क + HALANTA+ZWJ: क्

The Explicit halanta ् was also used for the formation of display variants. Basically it prevented the Consonant from joining to the next consonant.

क्त् yields the same result as क + HALANTA+ZWNJ+त: क्त

The Soft halanta formed by typing the Nukta Character after a Halanta allowed retention of the “half form” for the preceding consonant and prevents the form from combining with the following consonant.

श् ः व yields the same result as श+ ः+ ZWJ+ व= श्व

This is practically the same as using the INV

B. In the second instance they had a grammatical function. In ISCII this grammatical function was carried out by two types of modifiers:

EXPLICIT HALANTA

In Languages such as Bangla, Assamese, Oriya, the final implicit vowel is pronounced. However in loan words coming from English such as TABLE. The final implicit vowel is not pronounced. This is shown by putting a Halanta at the end of the word. The example is from Oriya

टेबल्

However if one appends to the word a suffixal element कु, the Root and Suffix join together: टेबल्कु, which is deemed as incorrect. To ensure that the two are clearly demarcated an Explicit Halanta is put after टेबल्+् which ensures the correct display form: टेबल्कु

¹ In keeping with Indian usage, the term Halanta will be treated as the Vowel Omission Sign co-terminous with the Western Indologist's use of the term VIRAMA which in Indian usage is nothing more than the Danda or the full stop.

In Sanskrit and Vedic texts a similar need is felt to show that the two consonants of which the first is a Half form and the second a full form, do NOT constitute a ligature. The use of the Explicit Halant enables this:

भक्ति भक्ति

SOFT HALANTA

The Soft halanta is used in Malayalam to derive the chilli aksharam in the case of the following 5 consonants:

न ण र ल ळ

Thus the presence and absence of the Chilli Aksharam demarcates the presence and absence of the dative.

2.2. This Historical Overview is important since the use of ZWJ and ZWNJ closely parallels that of the EXPLICIT and SOFT Halanta and INV in the 8bit ISCII Code

ZWJ and ZWNJ fulfill the same functions.

2.2.1. The case of Brahmi-based scripts will be taken first:

A. **As display variants**, ZWJ and ZWNJ are used to generate out across all Indian scripts either the Half form of the first Consonant in a consonant pair joined by a Halanta or show the first Consonant followed by the Halanta and then the Full consonant:

क + HALANTA+ZWJ: क्

क + HALANTA+ZWNJ+तः क्त

This has a pedagogical value and can be used in teaching half forms to learners. In E-Language Learning programs it also has the value of creating a useful teaching tool for distinguishing between Half and Full forms

B. As grammatical modifiers, in scripts such as Malayalam, Devanagari (Sanskrit), Oriya, Bangla, Assamese, ZWJ/ZWNJ perform a true grammatical function.

-In the case of Malayalam the insertion of a ZWJ generates out the Chilli Aksharam. This will disappear once the Chilli Aksharam are accepted. But insofar as the keyboard is concerned, they will continue, since the Keyboard has no place for the 5 characters (which to our knowledge were historically variants of a Final consonant followed by an overtly present Halant)

-In the case of Sanskrit/Vedic the use of ZWNJ will be needed for special purposes where the need is felt to show that the two consonants of which the first is a Half form and the second a full form, do NOT constitute a ligature.

The use of the ZWNJ enables this:

भक्ति भक्ति

-In the case of Assamese Bangla and Oriya , the ZWNJ ensures that the root word to which a Halanta is added (which is a loan coming from English or another European language where the final consonant is not pronounced) does not create a conjunct when a grammatical suffix is added:

टेबल्कु (INCORRECT) vs टेबल्कु

-Similar use of ZWNJ/ZWJ is noticed in Kannada

NOTE: the use of ZWJ to generate out the eyelash ra र्ं ZWJ: ँ as deployed by Microsoft in its Devanagari keyboard is incorrect, since Unicode prescribes the use of र् and not र्ं.

2.2.B The case of Perso-Arabic scripts

These are used for writing Urdu, Kashmiri and Sindhi and the case requires close attention. The example quoted for Farsi in PRI 96 (Figure I) where the absence & presence of a ZWJ permits demarcation of a plural marker ALEF+YE from the name of the letter YE (written as ALEF+YE) is theoretically possible in Urdu and Kashmiri.

As a test case, Urdu will be analysed since this is a more prevalent script and Kashmiri closely follows the same “imlaaa” or composition norms.

Urdu does not use this device since it will resort to two methods as well as a third possible method

- A. Inserting a space between the final HEH and the combination ALEF+YE
- B. Treating the HEH goal as an ezafat as in NAMAḤ ASIN i.e. Asin’s History NUN+ALEF+MIM+ALEF+HEH GOAL ALIF+AIN+SIN+NUN
- C. A third possibility exists where the final vowel of a root word is separated from the next vowel by the use of the HAMZA which basically acts as a Glide.

To sum up: The Use of ZWJ/ZWNJ does not seem to have any acceptance in Urdu and Sindhi and Kashmiri. At best it could be introduced as a separator.

3. ADVANTAGES AND DISADVANTAGES

The uses of ZWJ as Grammatical and Display adjuncts have been enumerated above. They are needed and required in Indic scripts where such forms are a necessity.

However unless specific and stringent guidelines are laid down as to the “legal” use of ZWJ/ZWNJ and all other uses are deemed as illegal and shown as such by the use of a display character; two worst case scenarios may occur

1. Developers may use these to enable the keying in of characters: the case of eyelash ra is a good example of such an eventuality. This will lead to illegal storage and create problems for routines such as searching, sorting and also in applications such as Spell-Checkers, Lemmatisers etc.

2. In the of IDN this may result in spoofing and pfishing which could have societal and economic implications.
Thus inserting a ZWNJ where it is not needed creates a clone of a domain name:
बडौडा बडौडा Both are Baroda as in the Bank of Baroda with the second string having a ZWJ after the ब +ZWJ

4. CONCLUSION

While encouraging the use of ZWJ and ZWNJ, this document also warns that unless Unicode lays down stringent guidelines as to its use, the larger chances of misuse or illegal storage loom on the horizon, a possibility best averted.

HINDI

Words with ZWNJ:

Improper appearance in Unicode	Proper appearance required as per language convention
अद्भुत	अद्भुत
आश्चर्य	आश्चर्य
प्रश्न	प्रश्न
आश्वासन	आश्वासन
पत्नी	पत्नी
शक्ति	शक्ति
आह्लादित	आह्लादित

With non-joiner	Code points	Names abbreviated
अद्भुत - अद्भुत	0905+0926+094D+200C+092D+0941+0924	A+DA+SIGN VIRAMA+ZWNJ+BHA+SGN U+TA
आश्चर्य - आश्चर्य	0906+0936+094D+200C+091A+0930+094D+092F	AA+SHA+SIGN VIRAMA+ZWNJ+CA+RA+094D+YA
प्रश्न - प्रश्न	092A+094D+200C +0930+0936+094D+0928	PA+SIGN VIRAM+RA+SHA+SIGN VIRAMA+ZWNJ +NA
आश्वासन- आश्वासन	0906+0936+094D+200C +0935+093E+0938+0928	AA+SHA+SIGN VIRAMA+ ZWNJ +VA+SIGN AA+SA+NA
पत्नी- पत्नी	092A+0924+094D+200C+0928+0940	PA+TA+SIGN VIRAMA+ ZWNJ +NA+ SIGNII
शक्ति- शक्ति	0936+0915+094D+200C+0924+093F	SHA+KA+SIGN VIRAMA+ ZWNJ +TA+SIGNI
आह्लादित- आह्लादित	0906+0939+094D+200C+0932+093E+0926+093F+0924	AA+HA+SIGN VIRANA+ ZWNJ +LA+SIGNAA+DA+SIGNI+TA

KONKANI

Zero Width Joiner

वाऱ्यार – वाऱ्यार (Mangal font)

वाऱ्यार – वाऱ्यार (Arial Unicode font)

Examples in Mangal font in the table.

Konkani Words	Code points	Names abbreviated
वाऱ्यार (incorrect)	0935+093E+0931+094D+092F+093E+0930	VA+AA+RRA+SIGN VIKRAMA+ZWJ
वाऱ्यार (correct)	0935+093E+0930+094D+200D+092F+093E+0930	(200D)+ YA+AA+RA

Zero Width Non Joiner

अर्द्यार - अर्द्यार (Mangal font)

अर्द्यार - अर्द्यार (Arial Unicode font)

Konkani Words	Code points	Names abbreviated
अर्द्यार (incorrect)	0905+0930+094D+0926+094D+092F+093E+0930	A+RA+SIGN VIKRAMA+DA+SIGN
अर्द्यार (correct)	0905+0930+094D+0926+094D+200C+092F+093E+0930	VIKRAMA+ZWNJ (200C)+YA+AA+RA

QUESTION

Konkani terms like “उद्देग” & “उद्देश” gets inflicted with both the joiners and shows correct spelling. (Zero width joiner and Zero width non-joiner)

Eg : उद्देश – उद्देश (Mangal font).

(Note: The example which is given in QUESTION column does not appear correctly in PDF format. So please check the file in Mangal font).

MALAYALAM

Word forms where without the ZWNJ is incorrect.

Malayalam words	Code Points	Names (abbreviated)
വാക്തരക്കം	0D35 + 0D3E + 0D15 + 0D4D + 200C + 0D24 + 0D6A + 0D15 + 0D15 + 0D02	VA + SIGN AA + KA + VIRAMA + ZWNJ + TA + R + KA + VIRAMA + KA + SIGN ANUSVARA
വാക്തരക്കം	0D35 + 0D3E + 0D15 + 0D4D + 0D24 + 0D6A + 0D15 + 0D15 + 0D02	VA + SIGN AA + KA + VIRAMA + TA + R + KA + VIRAMA + KA + SIGN ANUSVARA
വാക്സരണി	0D35 + 0D3E + 0D15 + 0D4D + 200C + OD38 + OD30 + OD23 + 0D3F	VA + SIGN AA + KA + VIRAMA + ZWNJ + SA + RA + NNA + SIGN I
	0D35 + 0D3E + 0D15 + 0D4D + OD38 + OD30 + OD23 + 0D3F	VA + SIGN AA + KA + VIRAMA + SA + RA + NNA + SIGN I
വാക്സാമരത്ഥ്യം	0D35 + 0D3E + 0D15 + 0D4D + 200C + OD38 + 0D4D + 0D2E + OD6A + 0D24 + 0D4D + 0D25 + 0D24 + 0D2F + 0D24 + 0D02	VA + SIGN AA + KA + VIRAMA + ZWNJ + SA + SIGN AA + MA + R + TA + VIRAMA + THA + VIRAMA + YA + SIGN ANUSVARA
	0D35 + 0D3E + 0D15 + 0D4D + OD38 + 0D4D + 0D2E + OD6A + 0D24 + 0D4D + 0D25 + 0D24 + 0D2F + 0D24 + 0D02	VA + SIGN AA + KA + VIRAMA + SA + SIGN AA + MA + R + TA + VIRAMA + THA + VIRAMA + YA + SIGN ANUSVARA
ദിഗ്വിജയം	0D26 + 0D3F + 0D17 + 0D4D + 200C + 0D35 + 0D3F + 0D1C + 0D2F + 0D02	DA + SIGN I + GA + VIRAMA + ZWNJ + VA + SIGN I + JA + YA + SIGN ANUSVARA
ദിഗ്വിജയം	0D26 + 0D3F + 0D17 + 0D4D + 0D35 + 0D3F + 0D1C + 0D2F + 0D02	DA + SIGN I + GA + VIRAMA + VA + SIGN I + JA + YA + SIGN ANUSVARA
ദിക്വിജയം	0D26 + 0D3F + 0D15 + 0D4D + 200C + 0D35 + 0D3F + 0D1C + 0D2F + 0D02	DA + SIGN I + KA + VIRAMA + ZWNJ + VA + SIGN I + JA + YA + SIGN ANUSVARA
ദിക്വിജയം	0D26 + 0D3F + 0D15 + 0D4D + 0D35 + 0D3F + 0D1C + 0D2F + 0D02	DA + SIGN I + KA + VIRAMA + VA + SIGN I + JA + YA + SIGN ANUSVARA

MANIPURI

The following words require ZWNJ so that they appear in the proper form:

Manipuri Word	Code Points	Names (abbreviated)
কাঙখল্ (Correct form)	0995+09BE+0999+200C+09CD+0996+09B2+09CD	KA+AA+NGA+ZWNJ+VIRAMA+KHA+LA+VIRAMA
কাঙ্বল্ (incorrect form)	0995+09BE+0999+09CD+09B2+09CD	KA+AA+NGA+VIRAMA+KHA+LA+VIRAMA
তপ্প (Correct form)	09A4+09AA+200C+09CD+09AA	TA+PA+ZWNJ+VIRAMA+PA
তপ্ব (Incorrect form)	09A4+09AA+09CD+09AA	TA+PA+VIRAMA+PA
ফম্ব (correct form)	09AB+09AE+200C+09CD+09AC	PHA+MA+ZWNJ+VIRAMA+BA
ফম্ব্ব (incorrect form)	09AB+09AE+09CD+09AC	PHA+MA+VIRAMA+BA
সঙ্গোম্	09B8+0999+200C+09CD+0997+09CB	SA+NGA+ZWNJ+VIRAMA

(correct form)	+09AE+09CD	+GA+O+MA+VIRAMA
সপ্তম (incorrect form)	09B8+0999+09CD+0997+09CB+09AE +09CD	SA+NGA+VIRAMA+GA +O+MA+VIRAMA
নম্বাক (correct form)	09A8+09AE+200C+09CD+09A5+09BE +0995+09CD	NA+MA+ZWNJ+VIRAMA +THA+AA+KA+VIRAMA
নম্বাক (incorrect form)	09A8+09AE+09CD+09A5+09BE +0995+09CD	NA+MA+VIRAMA+THA +AA+KA+VIRAMA

MARATHI

The following words require ZWNJ & ZWJ so that they appear in the proper form :-

Zero with Non-joiner

दुसर्था -- दुसर्था

Marathi words	Code points	Names abbreviated
दुसर्था (incorrect word) दुसर्था (correct word)	0926+0941+0931+094D+200C+092F+093E	DA+SIGN U+SA+RA+SIGN VIRAMA+ZWNJ+YA+SGN AA

Zero with joiner

अॅप्लेट -- ऐप्लेट

Marathi words	Code points	Names abbreviated
अॅप्लेट ऐप्लेट	0905+200D+0945+092A+094D+0932+0947+091F	A+SIGN CANDRA E+PA+SIGN VIRAMA+LA+SIGN E+TTA

ASSAMESE

The following words require ZWNJ so that they appear in the proper form:-

Assamese Word	Code Points	Names (abbreviated)
ৰ্যাম (correct form)	09F0+200C+09CD+09AF+09BE+09AE	RA+ZWNJ+VIRAMA+YA+SIGN AA+MA
ৰ্যাম (incorrect form)	09F0+09CD+09AF+09BE+09AE	RA+VIRAMA+YA+SIGN AA+MA
তহঁচিল (correct form)	09A4+09B9+09CD+200C+099A+09BF+09B2	TA+HA+VIRAMA+ZWNJ+CA+SIGN I+LA
তহঁচিল (incorrect form)	09A4+09B9+09CD+099A+09BF+09B2	TA+HA+VIRAMA+CA+SIGN I+LA

The following are the words in **BENGALI** which require ZWNJ so that they appear in the proper form :-

Bengali Word	Code Points	Names (abbreviated)
বন্ধ (correct form)	09AC+09A8+ 200C + 09CD +09A7	BA+NA+ ZWNJ + VIRAMA +DHA
বন্ধ (incorrect form)	09AC+09A8+09CD+09A7	BA+NA+VIRAMA+DHA
র্যাণ্ডম (correct form)	09B0+ 200C + 09CD +09AF+09BE+09A3+09CD+09A1+09AE	RA+ ZWNJ + VIRAMA +YA+SIGN AA+NNA+VIRAMA+DDA+MA
র্যাণ্ডম (incorrect form)	09B0+09CD+09AF+09BE+09A3+09CD+09A1+09AE	RA+VIRAMA+YA+SIGN AA+NNA+VIRAMA+DDA+MA
সফট (correct form)	09B8+09AB+200C +09CD+099F	SA+PHA+ZWNJ+ VIRAMA +TTA
সফট (incorrect form)	09B8+09AB+09CD+099F	SA+PHA+VIRAMA+TTA