Home FAQ Languages ▸ **Fonts** ▸ **Keyboards** ▸ **Typography** ▸ **Maps** ▸ **About Me** ▸ **Bibliography Email**

APOSTROPHES IN NATIVE LANGUAGES

CHRIS HARVEY © 2006

CONTENTS

FUNCTION OF APOSTROPHE-LIKE SYMBOLS

The apostrophe comes in several shapes and sizes, and has many different uses in languages across the world. It is an unusual member of the Latin orthography because it can have very distinct functions.

- **Elision mark.** Elision refers to the omission of a sound which might otherwise have been pronounced. In English, the eliding apostrophe appears in words like *don't* and *they've*. This also appears commonly in French and Welsh. In Skicinuwatuwewakon (Maliseet) the eliding apostrophe indicates that, historically, there once was an initial consonant which is now not pronounced. Marking this elision is important because it shows that the following consonant is to be pronounced voiceless. For example: ʼposkomon "he wears it" is pronounced [pəskəmən], but posonut "basket" is [bəzənot].

- **Separation mark.** Sometimes an apostrophe is used to separate two letters which would otherwise be pronounced as one sound. In Uummarmiutun (a dialect of Iñupiaq), the letter «nng» is pronounced [ŋŋ], as in avinngaq "lemming". When the sequence should be two distinct sounds [nŋ], an apostrophe separates the two: tanʼngit "whites".

- **Punctuation mark.** In North American English, the apostrophe also functions as a closing quote-within-a-quote, as in the following. *My brother said, "The woman told the dog, 'Stay!' after it had barked at the neighbour."* In British English, the rôles of the single and double quotes are reversed. This type of usage may appear in Native language texts. In French, «double guillemets» and ‹single guillemets› are commonly used instead. As punctuation, the closing single quote has no effect on pronunciation.

- **Grammatical mark.** On occasion, orthographies can distinguish between two otherwise identical words, prefixes or suffixes with some sort of mark or accent. In French, a "have" and à "to" are pronounced the same way but are marked with an accent to help the reader distinguish between the two words. Similarly, the apostrophe in the singular English possessive marker *'s,* as in *the man's clothes*, reinforces a possessive meaning instead of the unpossessed plural suffix *-s.* The plural possessive after the unpossessed plural suffix is indicated solely by the apostrophe: *the lions' roars.* Use of the apostrophe in this way is not at all common in Native languages. One could say that in eastern dialects of Kanienʼkéha (Mohawk), a verb-final apostrophe indicates the punctual aspect: senón:niʼ "you all are making",

senón:ni "you all have made" – both words are pronounced [zɛnũːni]. However, in other dialects of Mohawk, the final apostrophe in the punctual *is* pronounced.

- **Alphabetic letter.** Many Native languages have orthographies where the apostrophe is used as a letter of the alphabet. Many languages represent the glottal stop (IPA [ʔ]) with an apostrophe, as in the Hul'q'umin'um' (Cowichan) word 'i' "and". In Listuguj Mi'gmawi'simg (Restigouche Micmac), the apostrophe can represent a schwa, as in ms't "all", pronounced [msət].

- **Digraphs.** The apostrophe can also be used in a digraph or trigraph (a single sound written with more than one letter). English has several digraphs, such as: *sh*, *th*, and *oo*. In some languages, one of the letters of the digraph or trigraph is an apostrophe. Breton has a letter ‹c'h› which is pronounced [x]~[ħ]; the apostrophe is part of the trigraph and cannot be removed, nor does it indicate that some other letter has been elided. Native languages use apostrophes in similar ways. Dënesųłiné (Chipewyan) digraphs with an apostrophe indicate that the consonants are ejectives, as in erihtł'ís "book". This apostrophe is not a glottal stop, which in this language is written with the symbol ‹ʔ›, as in ʔihchogh "parka". In Yup'ik, the apostrophe represents a doubling of the preceding consonant, thus Yup'ik is pronounced [jupːik]. Many languages use the apostrophe both as a letter and as part of a digraph. In Gwich'in, it serves both as a glottal stop and ejective: t'aii'ee "hooked onto". Here the first apostrophe is part of the ejective digraph /t'/, and the second is a glottal stop.

- **Diacritics (accents).** Apostrophes can also occur as an accent above letter. Typically the diacritic modifies the usual sound of the letter in some way. In Ancient Greek, this diacritic marks the absence of an [h] sound at the beginning of words, as in ἀρετή "excellence". Typically, the apostrophe accent indicates an ejective or glottalised sound, like the Diitiidʔaatx (Nitinaht) word for "three", qaqac̓.

APOSTROPHE SHAPES

The apostrophe itself can take on different forms, depending on the font. Generally speaking, it is either shaped like a filled in number '9' or a slanted line. The shape of the apostrophe no more changes its function than a two-storied a or single-storied ɑ in English; it's simply a difference in typeface design.

'Palatino'   'Times New Roman'

The standard computer keyboard does not come with an apostrophe key as such. It has what is called a *dumb quote*, or *straight quote* key, which types a raised vertical line as in *don't*. Typographically, this character has no real use, and is a hang-over from the typewriter days when the dumb quote was used as a short cut for proper opening and closing single quotation marks. James Felici, in his book *The Complete Manual of Typography* writes, "The use of typewriter-style quotation marks instead of typographic ones should always be seen as a mistake."

Because of the limitations of the standard computer keyboard, many people have been forced to use the dumb quote as there was no way to input the proper curly apostrophe without inserting it directly from a character table. This has led to lots of poor quality typography in virtually all languages which use the Latin orthography. To remedy this problem, word processor and desktop publisher software designers have included *auto-correcting*, which automatically changes dumb quotes to curly quotes: *don't* becomes *don't*. This works well when the only apostrophe which occurs at the beginning of words is the opening quote or 6-shaped quote. Auto-correction has disastrous results when an eliding apostrophe or apostrophe letter can appear word-initially. For example, the language Tohono 'O'odham uses the apostrophe as a glottal stop. But when auto-

correcting is turned on, and the keyboard can only produce dumb quotes, the language name Tohono 'O'odham (which is typographically ugly because of the dumb quotes) becomes *Tohono 'O'odham (which is just plain wrong). For this reason, it is imperative that all **Native language typists turn off auto-correction in their word processors and use a keyboard layout which replaces the dumb quote with a curly apostrophe**.

Native language keyboard layouts should automatically type in the curly quote as per the language's orthography; Dakota requires the 9-shaped apostrophe while Hawai'ian uses the 6-shaped one. Some orthographies, like Deloria's Lakota, need both shapes: śk'e' "they say". Certainly using the dumb quote here would be impossible. Most of the keyboards on Languagegeek reflect this convention.

An asterisk * before a word means that *it's form is incorrect.

APOSTROPHE ACCENTS

For those languages which use an apostrophe accent, a lack of properly designed keyboards and fonts has led to the practice of typing the accent as an apostrophe before or after the main character: writing the Secwepemctsín (Shuswap) place name for Alkali Lake Esk'et instead of the more correct Esk̓et, or in Heiltsuk, writing 'Wúyalitx̌v instead of Ẃúyalitx̌v (for Uyalit). Now that proper keyboards and Unicode fonts are available, there should be no more need to cut corners like this.

APOSTROPHES IN UNICODE

Unicode, the universal encoding scheme for computers world wide, has encoded the apostrophe in several different ways. It is important to understand the various characters to ensure that one is using the proper symbol. I'll outline the major players here.

- **U+0027. The dumb quote '.** Should not be used under any circumstances. This character is unstable – word processing software may change this to another character. Reasons against its use have been discussed above. This is the character on the standard English keyboard.

- **U+02B9. Modifier Letter Prime ′.** This character has exactly the same shape as U+2032. If punctuation code points are to be preferred over modifier letters, then U+2032 is the better choice. See below under *Modifier Letter Apostrophe* for arguments against using spacing modifier letters.

- **U+02BB. The Modifier Turned Comma ʻ.** This is a spacing modifier letter, i.e. a spacing version of the accent U+0312. See below under *Modifier Letter Apostrophe* for arguments against spacing modifier letters.

- **U+02BC. The Modifier Letter Apostrophe ʼ.** This is a spacing modifier letter, i.e. a spacing version of the accent U+0313. The Unicode Standard 5.0 says under the entry for this character:

  - = apostrophe

  - glottal stop, glottalization, ejective

  - spacing clone of Greek smooth breathing mark

  - many languages use this as a letter of their alphabets

  With regards to its identity as a glottal stop, glottalisation, ejective, it is not up to Unicode to decide the phonetic value of a given character. Why should a language, like Goyogǫhó꞉nǫʼ (Cayuga), which employs the apostrophe for a glottal stop, use U+02BC while Breton uses U+2019? After a discussion on the Unicode mail-list, there are reasons for using U+2019 and reasons for using U+02BC. In my opinion, the best solution is to use U+2019 in almost *all* instances of apostrophe (except for

IPA and spacing clones). There are several important reasons why:

- ○ Here's a situation. A Cayuga writer is typing in the Native language, using U+02BC for apostrophe like the Unicode Standard suggests. At some point, this writer may want to add some English word with an apostrophe, or perhaps a surname like O'Brian. It would be unfair and unrealistic to expect this writer to realise that when typing Cayuga or English, different apostrophe characters are required. Imagine French and English having a different code point for the letter *M*, so that an English person typing *Montréal* would need to use the French M instead of the English M. Except in a very few cases where a language encoding standard has already been established using U+02BB and/or U+02BC, neither of the modifier letter apostrophes should be used.

- ○ As a spacing clone of the Greek smooth breathing mark, U+02BC must *always* be a 9-shaped apostrophe. It is not permissable that this character be the slanted shape because this would then look virtually identical to the Greek acute accent. A similar situation exists for some languages which use the Latin script, like Heiltsuk. In Heiltsuk, a letter, like ‹u› can take a Combining Comma Above (U+0313) as in the word u̓w̓íƛ̓itx̣ʷ, and it can take a Combining Acute (U+0301) as in q̓ʷúqʷay̓aítx̣ʷ. If we were to discuss the functionality of thesese two different diacritics in a grammar or textbook, we would use U+02BC and U+02CA. For example, "In Heiltsuk, the apostrophe accent ʼ indicates a glottalised sound, whereas the acute accent ´ marks high tone." This way the two characters would be distinct regardless of the style of the typeface. It is clear that if U+02BC could take the slanted shape, it would be indistinguishable from U+02CA. Therefore, wherever it is possible in a language for an apostrophe to appear in the slanted-line form (as in the example from the font Palatino above), we can be sure that it cannot be U+02BC. In every Native language I have seen in North America, the apostrophe can appear in various forms, for example, in Maliseet: dumb quote, 9-shape, slanted line. Consequently, the U+02BC character is not appropriate.

- ○ Why use U+02BC or U+02BB? There are two reasons which support using these characters for the apostrophes. 1) When you triple click to select a word, the quotation apostrophes might break the word into two parts. 2) Web site addresses cannot contain puntuation. These are good points, but are not strong enough reasons to outweigh the arguments above. Furthermore, for languages like Lingít (Tlingit), SENĆOŦEN, and Kanyen'kéha, punctuation marks are used for letters anyway (like the period, comma, and colon). So using a modifier apostrophe doesn't gain one anything.

- **U+2018. The Left Single Quotation Mark '.** Languages such as Hawaiʻian, and Deloria's Lakota should use this character. In the case of some Polynesian languages, they may have already officially decided to use U+02BB. In these cases, there will be minimal confusion between their apostrophe letter ʻ and the punctuation apostrophe ʼ as they are graphically different. Even for Hawaiʻian (specifically mentioned in Unicode 5.0), slanted-line shaped apostrophes are common, as in the book He Iʻa Wau: Pehea Koʻu ʻAno, so U+02BB is not universal by any means.

- **U+2019. The Right Single Quotation Mark '.** This is the correct code point for the any apostrophe: punctuation, phonological, or otherwise. Wherever languages are using U+02BC instead, problems will be encountered similar to those outlined above. It is an important point that different functions of the same letter in various orthographies does not merit multiple characters. English uses the apostrophe in three different ways: elision, grammatical, and punctuation. Breton also has two uses: elision and in trigraphs. Yet both always use the same character U+2019.

- **U+2032. The Prime ′.** As seen above, it can be preferable to use punctuation marks

instead of spacing modifier letters in natural language. For languages which use a spacing stress mark this would then be the correct character. It is very important not to use the dumb quote here, because it may be changed into a curly apostrophe by auto-correction.

There are a few Unicode characters for apostrophe accents. The two most relevant to this discussion are below.

- **U+0313. The Combining Comma Above.** This is the preferred character for Native North American languages. With a good font, this diacritic will appear over its base character, as in Upper Chehalis ċéx̣mesč "was tired of". This same accent can go atop another accent, as in čʼósos "children". Using a regular spacing apostrophe (*čʼósos) in these cases is unacceptable for many languages.

- **U+0315. The Combining Comma Above Right.** I am not aware of any Native languages which require this character. I have seen it used occasionally in digraphs in place of a regular apostrophe: *Gwichˊin (U+0315) instead of Gwich'in (U+2019). Here U+2019 is correct. I have also seen instances where capitals and letters with ascenders (tall letters like *l k d b*) which have an apostrophe accent use U+0315 instead of U+0313: as in Hən̓q̓əmin̓əm̓ typing *sʔəl̓eləxʷ (U+0315) instead of sʔəl̓eləxʷ (U+0313). Here U+0313 is correct. Different codepoints should never be used for glyphic variants of the same character. Use of U+0315 in this manner is strongly discouraged.

POTENTIAL PROBLEMS

A well designed orthography is one with as little ambiguity as possible. For any language which uses apostrophes for more than one purpose, difficulty can arise. In the case of English, with its three types of apostrophes: elision, punctuation, grammatical, there are some situations where the reader may be unsure. The sentences

1. Those girls are my sisters

2. Those girls are my sisters'

have very different meanings in English writing although they are pronounced the same way in speech. When we put these into single quotes (as used in Britain) potential for confusion arises.

1. 'Those girls are my sisters'

2. 'Those girls are my sisters''

In long sentences or complex quotes-within-quotes, this is going to become difficult to manage. In languages like Onǫda'géga' (Onondaga) where most words end in an apostrophe letter anyway, things get very messy very quickly.

To avoid this sort of ambiguity, there are several strategies which could be employed for orthographies with apostrophes.

- Find a method of representing quotation other than apostrophe-like symbols.

  a. Guillemets. These are small, arrow like marks which work just like quotation marks, with opening and closing forms. There are «double guillemets» for primary quotations and ‹single guillemets› for quotes-within-quotes. This is my preferred choice for indicating quotations as it does not interfere with apostrophe letters and it already has precedence in North America (French). For example: *My brother said, «The woman told the dog, ‹Stay!› after it had barked at the neighbour.»*

b. Long dash. Some languages, like Italian, begin a quotation with a long dash —. There is no closing form. This system might fall apart somewhat for quotes-within-quotes.

c. Nothing. Many Native languages have grammatical devices to show that something is a quotation, through affixation or particles. In these cases, perhaps there is no need for punctuation marks.

- Don't use single quotes as primary quotation punctuation. Double quotes are preferable although even these can get confusing. In the rare instances of quotes-within-quotes are required, the reader should be able to work out the correct interpretation through context. If at all possible, the methods above are better; an excess of apostrophe-like characters can truly result in a confusing jumble.

- Although there are several language standardisation projects for Native North American languages (Mushkego Cree, NWT Dene, Ojibway, Mohawk), they typically pay little attention to punctuation and capitalisation. And where specific guidelines are given, they are often ignored in practice. As a general rule, where Native speakers are bilingual in English, they use "English quotes", and where they are bilingual in French, «French quotes» are more common (note that there are not necessarily spaces surrounding the guillemets as in continental French). There is a lot of individual preference in this matter.

Home   Previous Page

JavaScript Menu Courtesy of Milonic.com