

Proposal to Modify the Encoding of Deseret Alphabet in Unicode

Kenneth R. Beesley
Xerox Research Center Europe
Ken.Beesley@xrce.xerox.com

25 April 2002, marked revisions 17 May 2002

1 Summary

It is proposed that the encoding of Deseret Alphabet in Unicode be augmented at the earliest opportunity to include four additional characters, used in some versions of the Alphabet, being

1. DESERET CAPITAL LETTER OI (IPA /ɔʰ/)
2. DESERET SMALL LETTER OI (IPA /ɔʰ/)
3. DESERET CAPITAL LETTER EW (IPA /j^hu/)
4. DESERET SMALL LETTER EW (IPA /j^hu/)

[Revision 17 May 2002: Citation glyphs for these characters are shown in N2473.]

2 A Short History of Deseret Alphabet Versions

The Deseret Alphabet went through a number of versions during its history. Several of the versions were used to write significant, and still extant, manuscripts that are of interest to historians for their content and to linguists for the phonological clues they provide to the speech of the writers.

Some of the versions of the Deseret Alphabet had only 38 letters, each in uppercase and lowercase; other versions had 40 letters, the two extra letters being the ones proposed herein for addition to the Unicode encoding.

Summary of known Deseret Alphabet versions:

1. Stout Version. Printed on a four-page broadside (copies are extant) on or shortly before 21 March 1854, when it was seen by Hosea Stout and copied into his journal. 38 letters. No known manuscript texts exist.
2. Remy Version. Seen in Salt Lake City in 1855 by travel writers Jules Remy and Julius Brenchley and printed as a plate in their published account (*A Journey to Great-Salt-Lake City*, London:W. Jeffs, 1861). 40 letters. No known manuscript texts exist, but Remy wrote, “The new characters, intended for the printing-presses of the Salt Lake, were cast at St. Louis; but up to this day nothing has been published, as far as we know, with these singular types. We have known them used in private correspondence, and seen them on some shop signs.”
 [Revision 17 May 2002: The Remy Version was used in a letter from George D. Watt to Brigham Young, 21 Aug 1854, Brigham Young Collection, Archives, Church of Jesus Christ of Latter-Day Saints. Watt’s letter contains parallel example texts in the Remy Version and in a radical proposed revision of the Alphabet.]
3. Haskell Version. Used for several months in 1859 by Thales H. Haskell to write his journal, which is now held in the library of Brigham Young University. 40 letters. This version was also used briefly, again in an extant journal, by M.J. Shelton.
4. Speller Version. Used in an extant undated manuscript, “The Deseret Phonetic Speller”, in the Archives of the Church of Jesus Christ of Latter-Day Saints, Salt Lake City, Utah. 40 letters. The glyphs of this version suggest that it was written circa 1859.
 [Revision 17 May 2002: On reexamination, the Speller looks like it could have been written anytime between 1855 and 1859.]
5. Deseret News Version. Used in 1859, 1860 and 1864 to print articles in the Deseret News newspaper. 38 letters.
6. Book Version. Used in 1868 and 1869 to print four books: The Deseret First Book (a primer), The Deseret Second Book (a primer), The Book of Mormon, Part I (intended as an advanced reader), and The Book of Mormon (full text). 38 letters.

At least four versions of the Alphabet, the Haskell Version (40 letters), the Speller Version (40 letters), the Deseret News Version (38 letters) and the Book Version (38 letters), are used in significant extant texts that historians and linguists might want to encode in Unicode. Other texts may come to light.

There is one other known version, the Watt Version, proposed and briefly used in a single extant letter from George D. Watt to Brigham Young, dated 21 August 1854. However, this was a radical proposal for changing the Alphabet, and there is no evidence (so far) that it was ever used outside the original letter of proposal.

[Revision 17 May 2002: As noted above, this letter does contain an example text written in the 40-character Remy Version of the Alphabet.]

3 History of Deseret Alphabet in Unicode

Currently only the Book Version (38 letters, 76 characters) of the Deseret Alphabet is accommodated in Unicode. John Jenkins of Apple, who championed the addition of the Deseret Alphabet, was honestly unaware that earlier 40-letter versions of the Alphabet had really been used.

4 Justification for the Augmentation

4.1 Practical Need for 40-Letter Encoding

It has been proposed that the two extra letters in the 40-character versions could be handled as ligature glyphs, thus existing only in the realm of fonts rather than in the underlying encoding. I believe that this would be a mistake. Phonological arguments for the phonemic status of /ɔʰ/ and /jʉ/ are presented below.

In the 40-letter versions of the Alphabet, the two letters in question were presented and used in a manner parallel to the other 38 letters. In practical use today, if the two extra letters were treated as ligatures, then the underlying encoding would presumably be something like the sequences corresponding to ɔɪ (or oɪ) and ju; when using 40-letter fonts, special rendering routines would presumably need to detect these sequences and render them with single glyphs. This would preclude the possibility of encoding a distinction between the /ɔʰ/ and the /jʉ/ diphthongs, on one hand, and the sequences /ɔɪ/ (or /oɪ/) and /ju/ on the other. Such sequences of vowels might appear in 40-letter texts as spelling mistakes¹ or as useful and accurate encodings of words such as *sawing* /sɔɪŋ/, *drawing* /drɔɪŋ/, *cawing* /kɔɪŋ/ or *blowing* /bloɪŋ/, *throwing* /θroɪŋ/ etc. where there really is a sequence of vowels, separated by a morpheme boundary, rather than a diphthong.

It would be highly annoying to have a rendering engine collapse an intentionally written sequence of vowels as in /sɔɪŋ/ or /bloɪŋ/ into an unintended diphthong glyph. I reiterate my belief that treating OI and EW as ligature glyphs would be a mistake.

¹I believe it should be possible to encode spelling mistakes accurately.

4.2 Historical Practice and Arguments

In texts written in the 38-letter Book Version, the phonemes represented by OI and EW were necessarily written with two letters. However, the variations in the alphabets and statements from the time indicate real doubt and debate about the status of EW and OI. In the Deseret News Version, which is almost identical to the Book Version, the editors (at least in the article printed 23 February 1859) offered the following apology: “Since the arrival of the matrices, &c, for casting the Deseret Alphabet, it has been determined to adopt another character to represent the sound of EW, but until we are prepared to cast that character, the characters [corresponding to] *iu* will be used to represent the sound of EW in NEW.”² The Pitman “phonotypy” alphabets of the day, which were the inspiration for the Deseret Alphabet, also vacillated on the issue of 38 vs. 40 letters.

4.3 Modern Phonological Justification

In the 19th century, “phoonetics” was a hot science, but the concepts of the phoneme and phonology were not fully understood. Isaac Pitman was constantly modifying his phonotypy alphabets to capture new “sounds” that he thought he was hearing.

From a modern phonological point of view, the Deseret Alphabet was clearly intended to be a phonemic alphabet, and the question properly reduces to this: Are OI and EW really phonemes in (standard dialects of) English? For OI, the answer is uncontroversially yes. The diphthong in *boy* /ɔ¹/ is parallel phonetically to the other English diphthongs in *high* /a¹/ and *how* /a^ʊ/ . The vowels in *hey* and *hoe* and also diphthongized in most cases: /e¹/ and /o^ʊ/ . It is of course possible to represent all such diphthongs orthographically using sequences of two separate characters, but this does not reflect the reality of their single-phoneme status in English. A diphthong, as described by phonologist Peter Ladefoged³ is a single vowel phoneme that “involves a change in quality within the one vowel” (op. cit. p. 76). In a phonemic alphabet like Deseret Alphabet, where /a¹/, /e¹/, /a^ʊ/ and /o^ʊ/ are (properly) encoded as single characters and rendered with single glyphs, there is no good justification for encoding or rendering the diphthong /ɔ¹/ differently.

The status of EW, i.e. /j^u/, is rather more interesting, and still somewhat controversial. Peter Ladefoged (op. cit., pp. 77-78) argues that it should be treated as a diphthong (i.e. a single phoneme) in English.

The last diphthong, [ju] as in “cue”, differs from all the other diphthongs in that the more prominent part occurs at the end.

²The editorial introduction is otherwise rather confused, suggesting also that a new letter would also be added for representing the vowel in *hair*.

³A *Course in Phonetics*, 2nd ed. New York: Harcourt Brace Jovanovich, 1982.

Because it is the only vowel of this kind, many books on English phonetics do not even consider it as a diphthong; they treat it as a sequence of a consonant followed by a vowel and symbolize it by [ju] (or [yu], in the case of books not using the IPA system of transcription). I have considered it to be a diphthong because of the way it patterns in English. Historically, it is a vowel, just like the other vowels we have been considering. Furthermore, if it is not a vowel, then we have to say that there is a whole series of consonant clusters in English that can occur before only one vowel. The sounds at the beginning of “pew, beauty, cue, spew, skew” and (for most speakers of British English) “tune, dune, sue, Zeus, new, lieu, stew” occur only before /u/. There are no English words beginning with /pje/ or /kjæ/, for example. In stating the distributional properties of English sounds, it seems much simpler to recognize /ju/ as a diphthong and thus reduce the complexity of the statements one has to make about the English consonant clusters.

4.4 Comparison with the Shavian Alphabet

The Shavian Alphabet was a 20th-century attempt at promoting a phonemic alphabet for English. It includes single symbols for the diphthongs /^ju/ and /ɔ¹/, parallel to all the other diphthongs.

4.5 Fonts and the Deseret Alphabet

The versions of the Deseret Alphabet differ not only in their phonemic inventory (38 vs. 40 letters) but also in the glyphs. Even the Deseret News Version and the Book Version, both 38 letters and both of which were cast into printing type, differ in the shape of the letter used for /e¹/; in the Deseret News version, it opens to the left much like the digit 3—in the Book Version, it opens to the right. The glyphs used (in 40-letter versions of the Alphabet) for EW and OI varied, and there were other variations in short vowels and even consonants. The handling of variant glyph shapes is almost certainly best handled by using different fonts to render phonemically standardized underlying encodings.

5 Summary

Based on the arguments above, I urge that the encoding of Deseret Alphabet in Unicode be augmented to provide single-character encodings for the

following diphthongs:

1. DESERET CAPITAL LETTER OI /ɔʰ/
2. DESERET SMALL LETTER OI /ɔʰ/
3. DESERET CAPITAL LETTER EW /j^hu/
4. DESERET SMALL LETTER EW /j^hu/

I understand that sufficient code space was left after the current encoding for just such an eventuality.

This expanded encoding will allow faithful (and phonological justified) encoding of Deseret Alphabet documents encoded in 40-letter versions of the Alphabet without prejudicing the encoding of 38-letter versions. It will allow, with 40-letter fonts, a justifiable distinction between the encodings of diphthongs in words like *boy* /bɔʰ/ and non-diphthong vowel sequences in words like *sawing* /sɔɪŋ/ or *blowing* /bloɪŋ/.

[Revision 17 May 2002: There are two reasonable candidates for the Unicode citation glyphs for OI and EW:

1. The glyphs used in the letter from George D. Watt to Brigham Young, 21 Aug 1854, and which were seen by Remy & Brenchley (op. cit.) in 1855. The same OI and EW glyphs are used in the (undated) Deseret Phonetic Speller Version.
2. Or the glyphs used in the Haskell and Shelton journals of 1859.

See N2473 for the recommended citation glyphs.]