ADAPTING UNIFIED ENGLISH BRAILLE (UEB) FOR LANGUAGES

OTHER THAN ENGLISH

by Joseph E. Sullivan

President, Duxbury Systems, Inc.

(1992-2004) Chair of the "Objective 2 Committee" of the UEB development project

Member, UEB Code Maintenance Committee of the ICEB

August 2015

It is a widely respected principle of braille that it should accurately parallel the customary written form of a language — the letters, numbers, punctuation marks and other symbols that occur in print — so that blind readers are in effect on the same page, orthographically speaking, as print readers. That characteristic derives from Louis Braille's original work, and is surely one of the key reasons that his code survives to this day as a practical as well as elegant writing system. However, because print symbols are not only much more numerous than the 64 basic braille cells but also are used very differently in the World's various languages, it is normally necessary as a practical matter for a braille code to be designed with a particular natural language as its primary focus, even if it is capable of being applied to a wider class of material.

Unified English Braille, or UEB, is no exception to that general rule. It is designed for English as the underlying natural language, with provision also for technical notation such as for mathematics, computer programming, chemistry and other sciences. But while it can express the occasional accented loanword or proper name from French and other languages, and also can represent Greek letters such as may occur in math, academic society letters and the like, its way of handling accented letters and non-Latin letters makes it impractical for passages actually in those languages. Consequently, it is not the position of this paper that existing braille codes should necessarily be replaced by UEB or even a variation thereof; every case will be different and in most situations, even if changes are under consideration, a code that is in current use is probably the best place to start. As always, those who speak a language and who will be reading and writing the corresponding braille code should have the final say as to what that code is like. Certainly the development of UEB followed that principle — and it took 12 years (1992-2004) to develop an acceptable system. You can't rush these things!

On the other hand, where the establishment or expansion of a braille code is deemed to be useful, UEB's design principles and even, to the extent feasible, its symbol choices, may be worthy of consideration as one way to promote such overall consistency as may be possible. In a sense, UEB would only be returning the favor if it were to play a role in the design process for other languages, as the committee that designed UEB looked closely not only at the prior English codes but also at current French, Spanish and other codes, as well as Louis Braille's original conception, in carrying out its own work [Ref. 1].

It is especially advantageous to consider adapting UEB for another language in those places where English is also widely spoken or at least studied along with that language, so that there is consistent representation of material such as math that is pretty much the same regardless of the host language. Not surprisingly — indeed it was anticipated during the development of UEB — South Africa is a prime example of such a place as it has 11 official languages, English being just one, and so Christo de Klerk, Susan van Wyk and their colleagues in South Africa set to work adapting the other 10 braille codes for general UEB consistency immediately upon UEB's adoption in 2004. Likewise, certain other languages that share a region with English have been adapted along UEB lines — Maori in New Zealand, Welsh in the UK, Irish in Ireland, and (implicitly) Hawaiian in the US. In addition, a less expected development has been that users of braille for some languages not usually associated with English have also adapted, or begun the process of adapting, those codes towards UEB. At this writing, such projects exist for Romanian, Maltese, Mongolian and Filipino.

UEB is a suitable general model because it was designed for both clarity and expressiveness. As is especially important for technical materials, symbols are to be clearly readable regardless of position or other aspects of context — that is, there is to be no ambiguity, or reliance on prior knowledge of the semantics, involved when it comes to understanding what symbols are represented in the braille. New symbols can be defined in a systematic way so that UEB can be extended, if necessary, without disturbing the existing elements of the code. More specifically, the main characteristics of UEB that are relevant to possible adaptation for languages other than English can be enumerated as follows:

1. Multi-cell symbols are formed according to a well-defined pattern so that a reader can always tell where a given symbol begins and ends even if the meaning of the symbol is not yet familiar. Most such symbols comprise one or more "prefixes" (any of the patterns with only right-hand dots, or dots 3456) terminated by a "root" (any other pattern with at least one dot). This general way of using braille cells is not new with UEB; it is common in many braille codes because cells with only right-hand dots are most easily read when "up against" other cells. The difference in UEB is that the principle is rigorously rather than informally applied, such that it can be proven mathematically that symbol extent is always clear [Ref. 1, Appendices B and C]. The result is to provide for an indefinitely extensible symbol set, with over 400 such symbols currently assigned.

2. Any symbol has only one basic, that is "grade 1", meaning.

3. In making basic symbol assignments, alphabetic roots a-z are used only for symbols that are inherently alphabetic, such as the Greek letters, or that contain a letter in stylized form, such as most currency symbols and the "at-sign". In particular, the English names of symbols have not been used as the basis for any basic symbol assignments.

4. The symbol series consisting of dots 456 followed by one of the 26 Latin alphabetic roots has been reserved for an additional alphabet that has not yet been decided upon. The original idea was to provide a parallel to the way that Greek (with characteristic prefix dots 46) is now handled in UEB, probably for either Cyrillic or Hebrew, which are found in some (mostly advanced) math. But while Greek letters are fairly common in English texts, especially in math, letters from any other alphabets are quite rare in practice, so much so that a specific assignment of this reserved 456-letter series has not yet been considered a priority. (UEB provides a "transcriber-assigned symbol" series for the odd occasion when a symbol that does not have a specific assignment must be represented. Consequently there is no urgent need to assign highly uncommon symbols.) Instead, this 456-letter series has been found useful in practice for languages, such as Maori and Hawaiian, where vowels may commonly have a "long" modifier (macron) and the 456 prefix is not only less cumbersome than the 2-cell UEB macron sign but also an already established custom.

5. Like other braille codes, UEB employs "modes" for special purposes, wherein symbols may have other than their basic meaning. The most important of these are numeric mode (wherein the symbols for the letters a through j represent digits instead) [Ref. 6], contracted mode (also called grade 2) and uncontracted mode (or grade 1). Other modes are provided for text emphasis, extended capitals, arbitrary shapes, arrows and lines. The exact extent of any mode is always well defined.

6. Contracted braille is most common in English, and by default is the normal mode. Hence "grade 1" indicators are provided to distinguish symbols, words and passages wherein symbols that could be read as contractions have their basic meaning. However, when an entire work is uncontracted, such grade 1 indicators are not required.

7. If contractions are used, they can be used in math and other technical notation as well, without ambiguity. This can be helpful in technical expressions involving ordinary words, such as

Interest rate × current principal balance ÷ number of days in the year

= daily interest

from a Web page containing a few “business math” examples [Ref. 4] or

{*t* | over half the members of ***I*** believe *t*}

from a recent book on category theory [Ref. 5, pg. 148]. However, grade 1 passages can also be used for any extensive algebraic math that uses primarily "letter" variables, thereby cutting down on "indicator clutter".

8. Contractions are of two types: those that always represent a certain letter-group inherently (e.g. dots 5-134 = "mother") and those where some contextual requirement must be met in order for the contraction meaning to apply, the basic meaning being understood otherwise. An example of the latter would be dots 256, which means "dis" at the beginning of a word and otherwise represents a full stop. Similarly, a braille "f" means "from" when, and only when, it is standing alone. In UEB, such contextual requirements as "beginning of a word" and "standing alone" are always defined precisely in terms of orthography, that is the exact pattern of punctuation marks or other symbols that must be present in order for the context condition to apply.

9. The rules regarding the use of "shortform" contractions within larger compound words are a notable instance of this preciseness principle. Shortforms are small letter groups that, when standing alone in contracted braille, represent a longer word, e.g. "ab" = "about". Shortforms may also be used in larger compound words standing alone but, in UEB, such usage is restricted to a certain specified list. That list is in turn maintained by a committee based upon human criteria. Ultimately, a compound will go on the list only if it is judged common enough and otherwise easily recognized. So the "about" shortform is used in "roundabout" but not in "squareabout" (a word I just made up). [Ref. 2, Section 10.9 "Shortforms"]

So, if one is thinking of adapting UEB to some other language, what are the considerations? The primary questions are how similar the writing system is to that of English, and whether or not contractions are to be used.

If the language in question is customarily written using only unmodified Latin letters, and if any existing braille code is uncontracted and has followed the normal assignments for those letters (which is likely), then the uncontracted aspects of UEB COULD be adopted as-is. Doing that would probably mean, however, some changes to common punctuation signs such as parentheses and also to indicators such as for italics and capitals. English readers had to adjust to some such changes from the earlier English literary code, so it is not necessarily a huge hurdle to learn a few signs that do not occur very frequently in typical literary text. However, if customs with respect to these basic signs have become deeply ingrained, it may be possible to consider some rearrangements of the basic symbol assignments that would better accommodate existing customs. Such rearrangements would need very careful consideration, however, given that internal consistency as well as compatibility with English could be affected. For instance, if the existing braille code uses dots 46 as a capital indicator, it would be tempting to interchange the role of dot 6 and dots 46 throughout the code. But while such a modification could be made to work structurally, e.g. in terms of symbol formation rules, the effect on the symbol set would likely outweigh the benefit obtained, as not only capital indicators but also Greek letters and a variety of nonletter symbols would be impacted.

If only unmodified Latin letters are in use but there are contractions to consider, the general mechanisms provided by UEB can be applied in most cases to preserve non-ambiguity — i.e. grade indicators to distinguish contractions from any symbols that have another meaning in uncontracted braille, adoption of precise definitions for context specifications such as "beginning of word" and "standing alone", and carefully defined limitations, based upon orthography, as to when contractions of the "shortform" type may be used in larger compound words. Of course the symbols that will require a grade 1 indicator in contracted braille will typically not be the same as in English because the corresponding contractions will naturally be different, but that in itself is not a problem as it's the basic principle that matters, not the specific signs that are affected. For instance, when a superscript occurs in contracted English, the superscript indicator (dots 35) must be preceded by the grade 1 indicator (dots 56) because otherwise dots 35 would be read as the contraction for "in". But if dots 35 has no contraction meaning in a given language, then there would be no need for the grade 1 indicator in that instance.

There may be cases, however, where some further measures may be necessary to cope with certain signs or other aspects of a given contraction system. UEB reflects this fact, as 9 of the 189 contractions defined in the prior English codes, along with the former practice of "sequencing" (omitting the space between certain word combinations), were dropped because of various conflicts with UEB principles. For example, the former "ble" contraction, dots 3456, is defined as a prefix in UEB and so the usage as a contraction would violate UEB symbol-formation rules. Likewise the former "ation" contraction, dots 6-1345 after the beginning of a word, was dropped because that combination naturally means a capital N, and interior capitals are now a fairly common occurrence in English words (especially commercial names and the like). While it would have been possible in that instance to retain the contraction and require a grade 1 indicator before any interior capital N, that solution was judged just too messy given the rather low net value of the contraction (which is only one cell shorter than "a" plus the "tion" contraction). Sequencing, of course, introduces a formal ambiguity between phrases such as "For the" and words such as "Forthe" that could conceivably occur in special contexts (such as a computer program) even though they are not normal English words. That, plus the complications involved in teaching blind children that the words are actually separate in print orthography, led to the decision not to retain sequencing in UEB. Such considerations may or may not enter into decisions regarding the contraction system of another language, but in entering into the design process, it is worth remembering that preciseness is important when technical notation is, or may be, present.

If the language is written with Latin letters but modifiers such as accents or ligatures are common, it is unlikely that the UEB provisions for these features will be satisfactory because they add two cells to every affected letter. Instead, the provisions of the existing code are likely to be the best place to start, at least, to provide guidance as to what is practical and likely to be acceptable. In cases where a prefix other than dot 6 or dots 56, such as dot 4 or dots 456, is currently used to indicate the presence of the modifier before the letter in its ordinary form, it is possible in most instances to retain that current symbol and, if by some chance UEB has assigned that combination to a symbol, simply to use the grade 1 indicator (dots 56) before the symbol when it has its basic UEB meaning. In Welsh, for instance, the accepted sign for the letter a with circumflex accent is dots 4-1, which is also the UEB "at-sign". So in Welsh braille, any instance of an at-sign is represented as dots 56-4-1. In other words, the at-sign is treated much as if the accented a were a contraction, although the 56 prefix is needed on the at-sign in uncontracted as well as contracted Welsh. Even though this means that at-signs require three cells each, the overall frequency of that symbol was judged low enough that it would not be worthwhile to complicate the matter further by assigning another, non-conflicting symbol. As another example, the letter d with circumflex below is customarily represented in Tshivenda braille as dots 46-145, which is the UEB sign for Greek delta. Consequently, in the adapted Tshivenda code, a Greek delta is dots 56-46-145, even in grade 1. Still another example is from the Northern Sotho language, in which the letter s with caron is represented by dots 4-234, which is the UEB dollar sign ($). So when the dollar sign appears in Northern Sotho braille as adapted, it is represented as dots 56-4-234.

This approach is an especially obvious choice when the currently used prefix is dots 456 because, as already mentioned, that prefix with an alphabetic root is already in the "reserved" series with no conflicting UEB basic assignment. By extension, one might consider changing to a 456 combination if it is necessary to make a change for some other reason.

A complication arises, however, if the current representation for a modified letter can also have a contraction meaning in some circumstance — e.g. when standing alone. As a hypothetical example, let us imagine that the abovementioned sign for s with caron in Northern Sotho, dots 4-234, had been used in the prior Northern Sotho braille code as an “alphabetic” contraction for the word “šala” (“stay”) when that word was standing alone. Then, in order to retain that contraction when adapting for UEB, 56-4-234 would naturally be needed for the s with caron when standing alone. Consequently that combination couldn’t also stand for the dollar sign; we would have had to either drop or redefine the contraction, or adopt a different dollar sign in Northern Sotho. (Fortunately there was no such contraction, so this actual situation didn’t arise; indeed such situations involving two-cell signs would be rare.) If a case such as this should be encountered, defining the basic symbol is likely to be easier and more acceptable to readers short-term, but it lowers the consistency with UEB and runs the long-term risk of symbol conflict as UEB adds symbols to its repertoire. An alternative that could theoretically be considered would be to devise a special "escape to UEB" indicator symbol that would generally apply to the following symbol when it is to have its UEB meaning rather than its normal meaning within the host language. Such a mechanism could be generalized to apply to multiple symbols, in the end amounting to "language switching" akin to UEB's existing facilities for handling inclusion of non-English text. But while possible in principle, it is hard to imagine how such an approach could be made practical for individual symbols here and there — most likely, a single-cell root symbol would have to be reserved for the purpose, but such symbols are usually already assigned in most braille codes, including UEB. All things considered, alteration or elimination of the contraction is the best solution technically, and at least does not affect basic symbols, but of course the matter of acceptance must also be taken into account. It must be said that the elimination of 9 of 189 contractions in English was controversial — once even being likened it to removal of letters from the print alphabet!

Though not essentially different from the two-cell case considered above, problems with representing modified letters become even more pervasive when, in languages following the French model, modified letters already have customary single-cell representations. (For instance, in such languages, an accented "a" may be dots 12356, which in UEB is the opening general fraction indicator.) This situation requires examination of each such symbol and, where there is a corresponding basic symbol assignment in UEB (as is very likely), any instance of the UEB symbol would need to be prefixed by a grade 1 indicator, dots 56. That is all that is required unless the symbol also has a contraction meaning in the host language — in which case, if ambiguity is to be avoided, it would be necessary to proceed as discussed in the preceding paragraph for the similar case involving two-cell signs. That is, the first option to consider would be to eliminate or redefine the contraction. If that option is unacceptable, then it will be necessary either (1) to define the basic symbol differently than in UEB, thus creating some current and potentially greater future incompatibility with UEB, or (2) to set up some “mode switching” convention, which can get heavy-handed if needed too frequently.

Finally we can consider the situation where the host language is written in a script other than the Latin alphabet. Thanks to hard work in the early 1950s [Ref. 3, especially "International Meeting on Braille Uniformity" on pg. 141], there is a generally respected international convention whereby the representation of non-Latin scripts, even those that are not in themselves alphabets, effectively involves at least an implicit "Romanization" so that the resulting braille utilizes the alphabetic and other signs in much the same way that they are used in French or English. For instance, the Arabic "mim" is represented by dots 134, just like the letter "m" in braille for Latin-based languages. (But that is a particularly simple example; correspondences between alphabets are not 1-to-1 in many instances.) To the extent such conventions have been followed for the host language's braille, it may be possible (though not necessarily easy) to consider the script as thus Romanized and thence to apply the same kinds of adaptations that are discussed above for Latin alphabets with modified letters. Beyond that it is difficult to generalize, much less to specifically consider all the issues that may arise as writing systems for natural languages vary so widely. At this writing, an adaptation of UEB for Mongolian, which is normally written in Cyrillic script, is in progress. As things stand, UEB math may be intermixed with Mongolian text, but the braille does not provide a specific indicator as to which script was used in the original — technically an ambiguity, though perhaps not an important one in practice. In any case, undoubtedly there are lessons yet to be learned from this project, and chapters yet to be written on the subject of adapting UEB for other languages.

REFERENCES

1. International Council on English Braille, Unified English Braille Code Research Project. The Reader Rules (The January 2004 Report of the Objective II Committee with corrections and amendments through February 15, 2004.) [This report served as the defacto primary definition of UEB from 2004 until the first edition of "The Rules of Unified English Braille" in 2010. It describes the code mainly as to what the symbols mean to the braille reader, which was the perspective taken during the code development. It is available at http://www.iceb.org/c2r0401.html]

2. International Council on English Braille. The Rules of Unified English Braille (Second Edition 2013). Edited by Christine Simpson. [This is the current official manual, and describes the code mainly from a transcriber perspective. It is available at http://www.iceb.org/Rules%20of%20Unified%20English%20Braille%202013.pdf]

3. UNESCO. World Braille Usage (A survey of efforts towards uniformity of Braille notation). by Sir Clutha MacKenzie (Chairman, World Braille Council). [This is available as a PDF file at <http://unesdoc.unesco.org/images/0007/000711/071103eb.pdf>]

4. Viewed August 2015 at

<http://www.myfedloan.org/billing-payment/about-interest/>

5. Cheng, Eugenia. How to Bake π (An Edible Exploration of the Mathematics of Mathematics). Basic Books, New York, 2015.

6. International Council on English Braille, Unified English Braille Code Research Project. From the Committee 2 Archives: The Second Debate on Numbers. Available at

<http://www.iceb.org/ueb_num_debate2.html>

(This is the full text of the principal debate leading up to the decision to retain the traditional “upper” numbers in UEB. A summary of the reasoning is also given in Ref. 1, Appendix D.)