

Fiat Lingua

Title: The Romanization of Middle Pahrān

Author: George Corley

MS Date: 12-25-2015

FL Date: 01-01-2016

FL Number: FL-000034-00

Citation: Corley, George. 2015. "The Romanization of Middle Pahrān." FL-000034-00, *Fiat Lingua*, <<http://fiatlingua.org>>. Web. 01 January 2016.

Copyright: © 2015 George Corley. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License.



<http://creativecommons.org/licenses/by-nc-nd/3.0/>

The Romanization of Middle Pahran

A Case Study in Conlang Transliteration

George Alston Corley

Abstract: In this essay I expand on my “Design Parameters for Romanization” (Corley 2011), defining five parameters for designing and discussing conlang romanizations: elegance, accessibility, aesthetics, internal history, and technical factors. I apply this framework in a detailed discussion of my own process designing the romanization for my current conlang, Middle Pahran. I pay special attention to overspecifying the phonology for accessibility, and to the compromises I made due to the technical limitations of the software I use.

INTRODUCTION

In 2011, I wrote a blog post entitled “Design Parameters for Romanization” (Corley 2011), in which I collected some of my thoughts concerning conlangers’ choices as regards romanizing their conlangs. I broke down the motivations for these decisions into four parameters conlangers need to balance when determining how to represent their conlangs in written form: elegance, accessibility, aesthetics, and (internal) history. On the one hand, romanization is almost a trivial part of building a conlang, that is, not really a proper part of the language. On the other hand, the romanization is the first thing people will see when exposed to your conlang, and so it deserves quite a bit of consideration.

I have recently had an opportunity to apply these parameters to my current project, a conlang called Middle Pahran (a working name). Middle Pahran is intended as one stage in a language family set in a fantasy world. It is called Middle Pahran because it is an intermediate stage between the proto language (Old Pahran) and the modern Pahran languages, which I plan to develop in the future. Middle Pahran itself is intended to be a full language capable of producing historical documents.

The romanization of Middle Pahran was a major challenge, mostly due to the fact that it has nine vowels which form a complex vowel harmony system. In addition, vowels are distinguished on length and nasalization, which yielded further complications. In this paper, I want to guide people through my thought process and the various factors at play as I wrangled this language into the constraints of the Roman alphabet. There will be diacritics and headaches and gnashing of teeth. In the first section, I will review and expand on my original design parameters as outlined in the above mentioned blog post. In section two, I will introduce the phonemic inventory of Middle Pahran and a few of my initial and simpler choices for romanization. In section three, I will highlight potential alternative interpretations of Pahran phonology, and my choice to “overspecify” the inventory to a degree in my romanization. Section four will highlight my troubles with Unicode compliance and keyboards, which led to the final form of the romanization scheme.

1 DESIGN PARAMETERS FOR ROMANIZATION

In my original blog post, I identified four principles that need to be balanced in a romanization scheme. In principle, these parameters are also relevant to transliteration schemes using other scripts, such as Cyrillic, the Greek alphabet, or Chinese characters (though the considerations might look *very* different there). They can also be applied to writing systems that are not transcriptions or transliterations (for various reasons the weighting may

be very different in such situations). Here, I will briefly go over my old design parameters, adding a little to some of them.

The first parameter is **elegance**. For the purposes of discussing romanizations, I define elegance as the degree to which the scheme is phonemic, and how close it approaches a 1:1 ratio of graphemes to phonemes. One could also refer to this as *phonemicity*. I tend to like elegant romanizations, and in Middle Pahrn I did try to get that 1:1 correspondence. Elegant systems are easier to achieve in smaller phonologies than in larger ones, where copious diacritics and digraphs may be necessary to convey the full inventory.

In my original blog post, I defined **accessibility** in terms of a linguistically naïve audience, making choices that will lead them as close as possible to the desired pronunciation. I still believe this to be an important consideration, but as I worked on Middle Pahrn romanization, I found I was also taking into account linguistically sophisticated audiences as well. Conlangers who are most likely to read about my language may well be able to make sense of IPA transcriptions, but I would not be presenting every example that way, and it is helpful to assume a certain familiarity with different writing systems and different uses of the Roman alphabet to lead people in the right direction. We will see the assumption of that knowledge used in my choices of correspondences. In some cases, it may be helpful to have different transliterations for linguistically naïve and sophisticated audiences (or for audiences of different language backgrounds)¹.

I would like to note, as I alluded above, that the choice to romanize at all, or the choice to romanize rather than to transliterate into another system, is an accessibility consideration in itself. It assumes that your intended audience is familiar with the Roman alphabet at least to some degree. The choice may be different for a different audience. For instance, I can handle most romanizations once taught the correspondences (and some even if I am not taught, given the language is similar enough to ones I know), but I am completely lost trying to read examples in Cyrillic, the Greek alphabet, or the Arabic abjad, or really anything other than the Roman alphabet – even if there is a section teaching correspondences. I am just not familiar with those systems. I might be able to get through a “Hanification” into Chinese characters, but if the conlang is at all well done and not a relex of Mandarin, I will not know how to pronounce the characters, and should need to relearn the meanings of some of them.

Moving along from accessibility, there is the issue of **aesthetics**. Aesthetics here, as in all things, is subjective. Some conlangers will strongly prefer to have lots of digraphs, while others will like diacritics. You might have very strong opinions on what is the most beautiful character to use to represent /k/. Anything could be involved here, but subjective aesthetics do need to be balanced against the other factors. That is not to say you cannot consider it more important – by all means, create a special snowflake of a transliteration if you like – but it is still a choice to consider carefully.

In the original blog post, my discussion of **internal history** played fast and loose with the term *romanization*, as I mentioned languages that, per conworld history, use the Roman alphabet natively. More properly, true transliterations and romanizations are by definition *not* the native writing system of a language, and should not really be treated as such. There is a possibility that a language might have a romanization that is historical (for conlangs written in another script but intended for the real world, or for alllangs set in alternate histories where the Roman alphabet exists), but in general, there is not a strong reason for a romanization to accumulate the same kinds of irregularities that native orthographies tend to.

However, there is something to be said for considering the internal history of a language, as well as other internal factors, such as conworld geography and language ecology. If you have a conworld with multiple related languages (or even unrelated languages), you might want to maintain a certain degree of consistency in their

¹ In a recent interview, JS Bangs makes mention of using a separate “book orthography” for his conlangs (Conlangery Podcast 2015).

romanizations, whether to highlight similarities or to simply lighten the load on a reader by following the same rules. I will discuss some of how these internal factors apply to Middle Pahran in the following sections.

In addition to the original four parameters I discussed in the blog post, I think it is relevant to add a fifth: **technical factors**. We live in a world of computers, ruled by Unicode, and this is a major issue of practicality when we are doing romanizations. How well can your funny characters with their funny hats be printed in a word processor? Can you easily make a keyboard for it? I will discuss this in section 4, where I note a change I had to make in Middle Pahran’s romanization due to font problems, and the work I went through to build a Middle Pahran keyboard layout.

2 THE INVENTORY OF MIDDLE PAHRAN

Table 1 Middle Pahran consonants

	<i>Labial</i>	<i>Alveolar</i>	<i>Retroflex</i>	<i>Palatal</i>	<i>Velar</i>	<i>Uvular</i>	<i>Glottal</i>
<i>Voiceless Plosive</i>	p <p>	t <t>	ʈ <ṭ>		k <k>	q <q>	ʔ <ʔ>
<i>Voiced Plosive</i>	b 	d <d>	ɖ <ḍ>		g <g>	ɢ <ḡ>	
<i>Nasal</i>	m <m>	n <n>	ɳ <ṇ>				
<i>Voiceless Fricative</i>	f <f>	s <s>	ʂ <ṣ>				h <h>
<i>Voiced Fricative</i>		z <z>	ʐ <ṣ>				
<i>Rhotic</i>		r <r>	ɽ / ɻ <ṛ>				
<i>Approximant</i>	ʋ <w>			j <y>			
<i>Lateral</i>		l <l>					

Table 1 shows the consonant inventory of Middle Pahran given in IPA, with their current romanized values in angle brackets. The analysis taken here is somewhat “surfacy” in some places, and includes some phonemes that a linguist studying this language may or may not include if Middle Pahran was a natlang being studied. More on this surfacy analysis is discussed in the next section.

Ultimately, assigning graphemes to the consonants was not difficult. In many cases, the IPA values themselves are quite suitable, and in other cases there were very commonly used pairings that fit in easily, such as <ʔ> for /ʔ/ and <y> for /j/. For the retroflexes, I initially used an overdot except for <ṭ> and <ḍ> (which don’t look good with overdots), but on looking into it, I decided it would be better to borrow the tradition from Indic languages of using underdot for all retroflexes – again, leveraging the fact that some in my audience will be familiar with these languages. Looking into this even inspired me to change my phonology somewhat – the retroflex rhotic was originally the retroflex approximant /ɻ/, but on finding that <r> often represents a retroflex tap, I decided that there would be dialectal variation between [ɽ] and [ɻ], which fits nicely with a planned divergence in daughter languages, where this phoneme will merge with /z/ in some daughters, and with /r/ in others.

Similarly to the underdot convention, I borrowed <ḡ> from a convention used in Arabic for a similar sound to Middle Pahran /ɢ/ — and I have already planned to look into what happens to /ɢ/ in Arabic dialects as inspiration for changes in Middle Pahran’s daughter languages. It also helped that this fit with the visual style of using dots elsewhere, which was absent with other choices I had gone with early on (namely <gh> and <ḡ>).

The vowels were much more challenging. Large vowel systems are always hard to romanize because the Latin alphabet was only designed to have five. Middle Pahran has nine vowels, shown with IPA values in Table 2, which also needed to be distinguished on length and nasalization. There was never a question for me that I would represent long vowels by doubling the vowel character. It was always the most intuitive way for me to represent long vowels, and in general I will always choose that option unless there is some reason it comes into conflict with another convention. A note on the representation of vowels – [y ø ʉ] are actually not phonemic in Middle Pahran, being allophones of /u o i e/ respectively. As such, I will always place these phones in square brackets []. I will discuss my reasons for representing them separately in the romanization in the next section.

Of course, because I need to use vowel doubling, that makes it unworkable to use digraphs for vowel quality, so I needed one unique character for each vowel quality. This led to one of my first drafts of the vowel romanization scheme, presented in Table 3. There are a few things to note in this romanization scheme. Notice that I have used <y> for [y]. At this point in my work, /j/ had been represented with <j> rather than <y> to make room for this choice. I also chose <v> to represent [ʉ], not too far from the Cherokee use of <v> for /ǝ/.

Table 2 Middle Pahran Vowels

	front		back		front		back		front		back	
	unround	round	unround	round	unround	round	unround	round	unround	round	unround	round
high	i	y	ʉ	u	i:	y:	ʉ:	u:	--			
mid	e	ø	ʉ	o	e:	ø:	ʉ:	o:	ẽ:	ø̃:	ʉ̃:	õ:
low	a				a:				ã:			

However, as my plans formed for the future of Middle Pahran, I decided that one future divergence of dialects would include changing [ʉ] to [v] and [j] to [ʒ], and possibly the introduction of [dʒ]. This means that, if I want to keep romanizations consistent in daughter languages, it would be very useful to plan ahead now and keep <v> and <j> available. Another change in later revisions of the phonology was in the method of denoting nasal vowels. Through the process of learning Hoocąk², I have gotten used to the idea of using the ogonek for nasalization, which opened up some options for the other vowels. I dislike stacking up multiple diacritics on top of a character, so I had avoided using the umlaut on the grounds that it would lead to using <ö> for [ø] and thus leave me with <öö> or <õõ> for [ø̃:].³ Your mileage may vary, but that just looks ugly to me – that <o> just has too many hats. However, if I switched to ogoneks, I'd get <õõ>, which looks quite nice and balanced.

Table 3 Early draft vowel romanization

	front		back		front		back		front		back	
	unround	round	unround	round	unround	round	unround	round	unround	round	unround	round
high	i	y	û	u	ii	yy	ûû	uu	--			
mid	e	ø	v	o	ee	øø	vv	oo	ěě	ø̃ø̃	ṽṽ	õõ
low	a				aa				ãã			

Thus, internal history considerations edged out <v> and <y> as vowel letters. /y/ was now represented with <ü>, in symmetry with having <ö> for [ø], freeing <y> to once again be used for /j/. In addition, when I

² Hoocąk is a Siouan language spoken by a Native American nation of the same name (also called Ho-chunk or Winnebago, though the last can be taken as pejorative) in a number of places in Wisconsin and on the Winnebago Reservation in Nebraska.

³ More on why I still maintain the double-letter system in nasalized vowels in the next section.

presented this along with several other romanization options to the Constructed Languages Facebook Group and to the /r/conlangs subreddit, quite a few people disliked <û> for [u]. Looking at it, I understood the sentiment. It is rather off the wall, especially in my earlier systems, where it was the only non-systematic diacritic. To counter this, I did at one point try <ô> for [ɤ], as well as <î ê> for [u ɤ], but there’s also just no precedent for using circumflex for either unround or back that I know of. Ultimately, the issue with back unrounded vowels is that there are few options. Coming from natlangs, your options for [u] are <ı> used in some Turkic languages, and <ʉ>, which is used in Vietnamese and brings with it an option for [ɤ] in the form of <σ>. Both of these options were repeatedly suggested to me, but I strongly resisted them on aesthetic grounds – I didn’t like the look of them. However, I discovered in my experimentation that <ǫǫ> seemed a pretty cute representation for [ɤ:]. I was aware that <õ> existed of course, but in my opinion, that would actually be somewhat detrimental to accessibility. Anyone familiar with Vietnamese would see <õ> and think of the Vietnamese “breaking” tone, and I did not want to confuse people. So, I went with horns, ogoneks and umlauts, producing the romanization seen in Table 4.

Table 4 Horn-ogonek-umlaut romanization

	front		back		front		back		front		back	
	unround	round	unround	round	unround	round	unround	round	unround	round	unround	round
High	i	ü	ʉʉ	u	ii	üü	ʉʉʉʉ	uu	--			
Mid	e	ö	σ	o	ee	öö	σσ	oo	ęę	ö̇ö̇	σ̇σ̇	ȯȯ
Low	a				aa				ąą			

However, this was not the final romanization for the vowel system of Middle Pahran, and I ended up using instead <ı> for [u] and <ë> for [ɤ]. More on that decision is discussed in section 4. I also have not exhaustively listed all the options that I considered, nor all the suggestions I received as I presented different romanization schemes to the community. In the following section, I will go over a few notes about how the romanization I chose perhaps overspecifies the phonology of Middle Pahran.

3 PHONOLOGICAL ANALYSIS AND OVERSPECIFICATION

By *overspecification*, I mean drawing distinctions that are not present in the phonology. It is not entirely uncommon for romanizations to overspecify the phonology of a language. For instance, in Mandarin Chinese, it is possible to analyze the palatals [tɕ tɕʰ ɕ] as allophones of /k kʰ x/ (Duanmu 2007)⁴, but most romanizations of the language have represented them with separate characters.

There are several places where the Middle Pahran romanization could be seen as overspecifying the phonology. The most obvious place is, ironically, in the vowel system. As I noted in the previous section, the secondary cardinals in Middle Pahran [y ʉ ø ɤ] are not actually phonemic. This is a result of the evolution of Middle Pahran’s vowel system. Old Pahran had a three vowel system /i a u/. There were two sources for the mid vowels, nasalization and lowering due to a preceding /q/ or /h/. The primary cardinal mid-vowels /e o/ became phonemic when /h/ merged with /h/ (and, in some environments, was subsequently deleted).

The only source for [y ʉ ø ɤ] is vowel harmony. I do have a simple sound change planned for daughter languages that will make [y ø] phonemic, but in Middle Pahran they are still allophonic variants of /u/ and /o/. Ironically, if I kept to a more elegant system, I would have had a much easier time with the vowels – I would only need to use five. The choice to represent these sounds was instead a nod to accessibility – I wanted to highlight

⁴ This is only one possible analysis. In addition, Duanmu (2007) has an interesting analysis for Mandarin vowels which collapses all the mid vowels to underlying /ə/.

the existence of the vowel harmony system at this stage of the language, and aid pronunciation. In addition, the potential for these vowels to become phonemic in the future makes figuring out how to represent them now quite useful.

Another potential overspecification is seen in the retroflexes. The only historical source I have for retroflexes in Middle Pahran (other than [r̥]) is the result of assimilation of a preceding retroflex rhotic, which was subsequently deleted with compensatory lengthening. Since /l/ and /r/ also underwent deletion with compensatory lengthening, it is possible to find minimal pairs – at the time of this writing, none existed in the lexicon, however I was able to create a term *haas* ‘monkey’ (< *hals), in opposition to *haaş* ‘foot’ (< *haʎs)⁵.

However, this change is still morphologically visible, and the deleted codas can still be recovered in some cases, as is illustrated in (1-2). In the examples below the perfective is formed with the suffix –su, and the irrealis with –un (the glottal stop is deleted when following a consonant). The examples in (1) show both the application of retroflex assimilation and the recovery of the underlying /ɹ/ in the verb stem when followed by a vowel. The examples in (2) show that both /l/ and /r/ can also be recovered in the same way.

- | | | |
|------------------------------|--|------------------------------------|
| (1) <i>makeep</i> ‘to push’ | <i>mūmakeeşö</i> ‘she pushed (perfective)’ | <i>mūmakerön</i> ‘she might push’ |
| <i>guuṇip</i> ‘to turn’ | <i>muguuṇiṣu</i> ‘she turned (perf.)’ | <i>muguuṇirun</i> ‘she might turn’ |
| (2) <i>sannuup</i> ‘to cook’ | <i>musannuusu</i> ‘she cooked (perf.)’ | <i>musannurun</i> ‘she might cook’ |
| <i>fyakaap</i> ‘to yell’ | <i>mufyakaaso</i> ‘she yelled (perf.)’ | <i>mufyakalon</i> ‘she might yell’ |

Thus, a particularly reductive phonologist could conclude that the retroflex series is entirely the result of assimilation to a preceding /ɹ/, and that the deletion with compensatory lengthening is still active in the languages. There is one argument against this analysis: where there is a long vowel that is not followed by a retroflex, the identity of the deleted liquid is often not recoverable, as not all environments will be subject to the morphophonology present above. For instance, in *briiza* ‘donkey’, it’s not possible to determine whether the deleted liquid was /l/ or /r/. In *taaba* ‘body’, it’s impossible to eliminate any of the liquids, as the long vowel is not followed by a coronal. In any case, the point is that I chose to represent the retroflexes separately not to support some particular phonological analysis for my own conlang, but to help linguistically-savvy readers accurately pronounce the language.

I will not belabor the point by going into great detail, but there are some similar arguments to be made for nasalized vowels, which also have only one source (nasal deletion), with the etymology recoverable through morphology (arguably much more consistently than for retroflexes and long vowels derived from liquid deletion). In that case, I considered simply leaving the etymological coda nasals to mark nasalization, but the nasalization rule is fiddly enough that I felt it would be good to mark it. However, I will point out that I did get at least one suggestion not to bother with doubled vowels for nasalized vowels, since they are always long. Again, I could have made them single vowel letters, but I felt I wanted to remind the reader of the fact that they are long. Also, future daughter languages may also have short nasal vowels, so it is helpful to be consistent.

I did not always bow down to accessibility and overspecify my romanization. At a certain point, always overspecifying will just lead to insanity – I might as well just give people audio files or, worse, spectrograms. That is not going to be helpful at all. And in any case, there are other considerations. For instance, there is a rule that causes non-coronal stops to become labialized before a rounded vowel – with a related rule that turns stop+/v/ into a labialized stop. I decided not to represent surface forms there, partly as a nod to elegance, but mostly because an attempt to represent all the labialization occurring due to rounded vowels would lead to an unsightly proliferation of <w> -- in other words, an aesthetic consideration.

⁵ It is interesting that identifying phonemes in a conlang is different from doing so in a natlang. In a historically-derived conlang, where all sound changes are known, I imagine you could mathematically prove that two phones will always appear in complementary distribution.

4 KEYBOARD AND UNICODE TROUBLES

In section 1, I noted that in addition to the original four parameters, a fifth parameter for evaluating romanizations might be *technical factors*. The fact is, computers rule our world, and if you have any intention of sharing your conlang to the world, it will be quite useful for people to be able to type it. There are materials out there about converting original scripts into fonts (Watkins 2013; Peterson 2015:chap. 4), but here my focus is on romanizations, and if your conlang contains any diacritics or letters not used in English, you are most likely going to have to wrangle with Unicode.

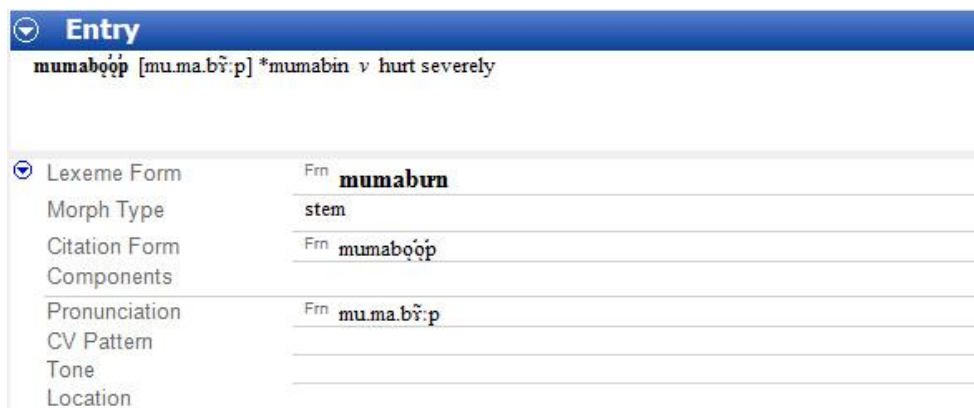


Figure 1 Problems rendering <ǫ>

As I mentioned in section 2, I had to get rid of my horns in the final version of my romanization because of technical issues. To illustrate this point, Figure 1 shows what happens when I attempt to put <ǫ> into SIL FieldWorks Language Explorer (FLeX), which I have been using to manage my lexicon. The problem arises due to the use of combining characters. In Unicode, there are some diacritic-bearing characters, such as <ü>, <ö>, <u>, <á>, <ñ>, and many, many others, which have their own code-points. These tend to be characters that are frequently used in natural languages, so that a separate code point would be useful. Even some characters with multiple diacritics, such as the aforementioned <õ>, have their own code points. However, when you are putting multiple diacritics on a character, or even just applying a diacritic to characters it's not usually used with, you will probably have to resort to adding combining diacritics. In theory, a combining diacritic is supposed to work like a ligature – they are designed to neatly attach to the previous character to form a new visual character. Some programs just convert all characters with diacritics to a sequence of main character plus combining diacritic. If you ever hit backspace and delete just a diacritic, rather than the whole character, you are dealing with combining diacritics.

Unfortunately, fonts do not always fully support all combinations, and that is what happened with <ǫ>. I created my <ǫ> by taking <q> (U+01EB: LATIN SMALL LETTER O WITH OGONEK) and adding the combining form of the horn (U+031B: COMBINING HORN), and FLeX's font choked on that, improperly attaching the horn. I also tried to go the other direction, starting with <σ> and adding a combining ogonek, but I ran into a similar problem with the ogonek not attaching properly. Ultimately, I ended up ditching the horns entirely, since it would not make

much sense having only one. For [ɣ] I chose <ë>, since the umlaut plays pretty nicely with the ogonek⁶, and for [ʉ] I went with <ı>⁷. The final form of the vowel romanization scheme is presented in Table 5.

Table 5 Final Middle Pahran vowel romanization

	front		back		front		back		front		back	
	unround	round	unround	round	unround	round	unround	round	unround	round	unround	round
high	i	ü	ı	u	ii	üü	ıı	uu	--			
mid	e	ö	ë	o	ee	öö	ëë	oo	ēē	ōō	ëë	oö
low	a				aa				aą			

With my romanization squared away, I wanted to create a keyboard in order to be able to type Middle Pahran. For the task I used Microsoft Keyboard Layout Creator (MSKLC), because it is free and I use Windows. I did not have to change anything about the romanization due to my keyboard, but I thought I would mention a couple of points in the design of the keyboard. Figure 2 shows the default values of each key. Figure 3 (page 9) shows the keyboard values when Shift is held.

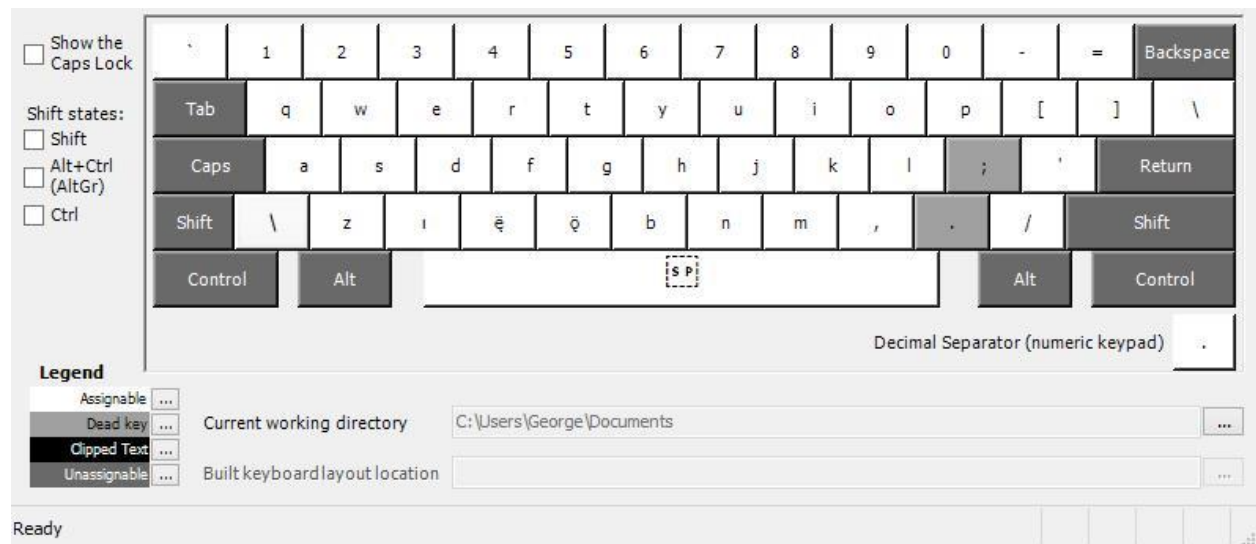


Figure 2 Middle Pahran Keyboard (default state)

The three grey keys are dead keys. Dead keys return different values depending on the next key pressed. They are useful for typing diacritics and ligatures. I chose to use the semicolon key as a dead key for ogonek because I was already familiar with that from the Hooçak keyboard I use. The quotation marks, similarly, are the dead key for diaeresis on the US-International keyboard. I chose the period for all my dotted consonant characters <ř ş ı đ z ğ> simply because it seemed the most intuitive choice.

⁶ However, while preparing this for submission, I attempted to use LyX to make a LaTeX version, and found out that LyX cannot handle combining diaeresis properly. The nightmare only continues.

⁷ I did also test capital letters for everything, though I have presented only the lower case versions. One thing that made me reticent about using <ı> was what to do about the capital letters, until I looked it up and found out that in Turkish, capital <ı> is in fact <İ>, an <I> with an overdot. It looks funny, but it satisfied me well enough.



Figure 3 Middle Pahran Keyboard (Shift active)

Since <x>, <c>, and <v> are not used for Middle Pahran, those were free to use for other characters. I mapped <i> to <x> (and correspondingly changed <l> to <i>) simply because I felt it was logical for it to have its own space. The letter <i> is far more common than <l>, so there was no point in loading another dot character on the period dead key. However, adding <ë> and <ö> to their own keys was purely a technical issue. MSKLC keyboards do not seem to like stringing multiple dead keys together, and also do not like to have characters with more than one Unicode code point mapped to said dead keys (necessary for <ë> and <ö> as they both use U+0308: COMBINING DIAERESIS.), so I simply had to give them their own keys for the keyboard to work. If someone has a less hacky solution to that, I would love to hear it, but I do not think there is another way to do it in MSKLC.

In short, do not forget to consider the technical requirements for your romanized script. If you use all digraphs, or simply do not need many sounds, you can probably get away without any of this technical wrangling, but if you plan to use diacritics, be prepared to do a little playing around with Unicode and keyboards.

CONCLUSION

My goal here is not to advocate for a particular way of creating a romanization, but simply to give an example of how different factors need to be considered in the construction of one, particularly for languages with large numbers of sounds that the Roman alphabet is ill-equipped to handle. Different conlangers will find a different balance than I have, and I might even find a different balance myself, as I am considering having alternative romanizations for less savvy audiences, or places where my diacritics might fail (such as if I include Middle Pahran in printed fiction in the future). These principles might also apply to different degrees to invented scripts as well. Certainly accessibility is a lower priority there (I plan to make the Pahran writing system annoyingly opaque and historical), and in *The Art of Language Invention*, David Peterson mentioned a technical consideration to invented scripts in the fact that left-to-right alphabetic scripts are easier to work with on a computer (Peterson 2015:227).

In all of these cases, I think considering the success of an orthography as a balance of the parameters of elegance, accessibility, aesthetics, internal history, and technical factors will be helpful in deciding which course to take with your orthography, and what benefits and drawbacks each choice may have. In any case, I hope that this

discussion has been helpful to others who are working on orthographies for their conlangs. It may be only a small part of the process of conlanging, but it is one of the most visible. Thank you, and happy romanizing!

REFERENCES

- Conlangery Podcast. 2015. Conlangery #113: Interview with JS Bangs. *Conlangery Podcast*.
<http://conlangery.com/2015/08/03/conlangery-111-interview-with-js-bangs/> (20 October, 2015).
- Corley, George A. 2011. Design Parameters for Romanizaion. *George's Blog*.
<http://www.gacorley.com/blog/2011/11/14/design-parameters-for-romanization.html> (20 October, 2015).
- Duanmu, San. 2007. *The Phonology of Standard Chinese*. Second Edi. Oxford University Press.
doi:10.1353/lan.2003.0139.
- Peterson, David J. 2015. *The Art of Language Invention*. New York: Penguin Books.
- Watkins, Britton. 2013. Conscript Creativity with Computer-Based Fonts. *The Fifth Language Creation Conference*.
https://youtu.be/vmMN_wDmftI.