

Sohlob romanization

Benct Philip Jonsson

(Updated 2016-06-28)

Contents

<u>Background</u>	2
<u>Overview of romanization schemes for Sohlob languages</u>	3
<u>Vowels</u>	3
<u>ASCII vowel digraphs</u>	4
<u>Consonants</u>	5
<u>Modifiers and diacritics</u>	6
<u>Prosody</u>	6
<u>Other diacritics</u>	7
<u>Overview of the native orthographies of the Sohlob languages</u>	8
<u>Vowels</u>	8
<u>Consonants</u>	9

Background

I'm in the process of switching from the old 'cool' but inscrutable Sohlob romanization scheme to a new uncool but more 'realistic' scheme, which comes in 'narrow', 'broad' and 'simplified' versions.

The old scheme harks back to Latin-1 days. The main problem with it beside odd mappings for *e* and *q* is that it looks too much like a roman orthography, which is something which these languages most certainly don't have. The dry and philological look of the new scheme is actually appropriate.

The 'simplified' scheme uses a raised dot or apostrophe to distinguish sequences like *kl-kh-l* and *kl-k'hl*. Strictly speaking the apostrophe marks a dropped vowel while the raised dot is purely a grapheme separator. Thus the instrumental of *ākāl* is *āk'lar/āk'h'lar* though I usually spell it *āklar/ākhlar*. Some of the alternatives represent differences between the languages and/or dialects or allophones while others are merely allographs. The main allophonic case is *š* vs. *ž* which are consistently distinguished in all the schemes except the old Latin-1 one in spite of being allophones of a single phoneme.¹

The 'new' scheme is actually not new at all. It is the longhand equivalent of the system of diacritics which I have used for phone*ic notation in shorthand for years.² Out of habit I just haven't used it for Sohlob, apart from diachronic stuff, or rather I was confused by the Latin-1 scheme and inconsistently used *e i y* with various diacritics for [i]. Earlier I have mostly used *ε đ k/h g t r* as longhandization of the macron-below notation for fricatives, and I might still revert to it.³

Actually there should also be a new version of the strict *transliteration* which faithfully mirrors the quirks of the native script, doing things like *ḁ ḃ ḅ ḇ ḉ ḋ ḥ ḥ' ḥ* for *p t k l ž s*. (That is e.g. *z* is *Ḍ* ⟨z⟩⁴ while *ž* is *Ḏ* ⟨ẓ⟩ and *h* is *Ḥ* ⟨h⟩ while *s* is *Ḧ* ⟨dh⟩ or *Ḩ* ⟨ḥ⟩. Needless to say it mostly makes sense historically!) I'm really at a loss what to do with what until now has been ⟨*e é x x̣*⟩, for *ī ä k š*, in the strict transliteration. I'm thinking that I may keep them, since the native letter *Ḕ* ⟨e⟩ actually is *e* [ɛ] in Chidilib, as opposed to *i* [i] in Classical Sohlob. There is no such excuse for *Ḷ* ⟨x̣⟩ though. The use of *Ḷ* ⟨y⟩ and *Ḹ* ⟨ỵ⟩ for *i y* in order to avoid ⟨*v*⟩ is also weird.

/bpj

¹I always thought that for example the spelling *jdreg* for the Chidilib word for 'bridge' (now *ždreg*, Classical Sohlob *äždräg*) looked somehow dyslectic.

²The diacritic shorthand phone*ics system is worth a tale in its own right. I might write that some day.

³I'm almost ashamed to admit that the main reason for using macron-below in longhand now is that I want to have room for more combining marks in my custom keyboard on my Android tablet, created with the [Multiling O Keyboard](#), which really rocks, never mind all those emoji which I don't use! :-)

⁴Which in origin is an upside-down *Ḷ* ⟨ḷ⟩!

Overview of romanization schemes for Sohlob languages

The following pages contain synoptic tables of the new and old romanization schemes with variants. Note that only those romanization symbols which use diacritics, digraphs or odd mappings in one scheme or other are listed; *b, d, f, g, h, i, k, l, m, n, p, r, s, t, u, v, z* are used with roughly their IPA values in all the schemes.

The schemes are, from left to right:

1. *IPA*: a key rather than a romanization scheme. Ties omitted in affricates.
2. *Narrow*: mainly used in diachronic-phone*ic contexts.
3. *Broad*: used in synchronic-grammatical contexts. Roughly phonemic.
4. *Simplified*: used for names and terms in metalanguage contexts (and in narrative fiction).
5. *ASCII*: useful for portable filenames and with really dumb software.
6. *ASCII-96*: the old ASCII scheme from the nineties.
7. *Latin-1*: the old Latin-1 scheme from the nineties which I'm finally abandoning.

Vowels

IPA	Narrow	Broad	Simplified	ASCII	ASCII-96	Latin-1
æ	ǣ	ǣ	ǣ/æ	ae	ae	æ/ǣ
ǣ	â (a)	a	a	a	a	a
ɒ	ǻ	ǻ	o/ǻ	o/ao	o	o
ɑ	ɑ	a (ɑ)	a/á	a (au)	a	a
ɑ	ɑ	a (ɑ)	a/à	a (ai)	a	a
ɛ	ĕ/ĕ	ĕ (ĕ)	e/ĕ	e/io	e	e
ɛ	ê/ĕ	ê (ĕ)	e/ê	e/ia	e	e
ɛ	ê/ə	ê	e/ê	e	e	e
e	ę/e	e (ę)	e/é	e/ei	e	e
ɛ	ę/e	e (ę)	e/è	e/ea	e	e
ɪ	î	î (î)	i/ie (î)	ie	e	e
ɪ	ï	ï	i/ie	ie/iu	e	e
ø	ö/ö	ö (ö)	ö/oe	oe/oei	oe	ö
œ	ÿ/ö	ö (ÿ)	ö/oe	oe/oea	oe	ö
o	ô	ô	o/ô	o/eo	o	o
o	o/o	o (o)	o/ó	o/ou	o	o
o	o/o	o (o)	o/ò	o/oa	o	o
u	û	û	u/û	u/eu	u	u
y	ÿ	ÿ	ÿ/ue	ue	ue	ÿ
a: etc.	ā	ā	aa	aa	aa	aa/á
ǣ etc.	ǣ	ǣ	a	ax	a	a

One could use *ï î ï û ü û* for [ɪ ɪ̃ ɪ̃ u y ʊ] and *î ï î û ü û* for [i ĩ ĩ u y u] but the distinction was not phonemic in any of the languages concerned, and was only ever mentioned in passing by native philologists.

ASCII vowel digraphs

The vowel digraphs in ASCII are rather *ad hoc*, and most have to be memorized. Some mnemonic hints have, however unsuccessfully, been attempted:

[back]e	front
i[back]	(oddly) back unrounded
[back unrounded]o	back rounded (ao)
e[back rounded]	central rounded
[front unrounded]e	central unrounded (ie)
[mid]a	low mid
[mid][high]	high mid
[low (central)][high]	(more) front/back (ai, au)
ox	short (somewhat odd!)
oo (doubled)	long

Examples of length marking:

aae	[æ:]
iuu	[u:]
aaö	[ɔ:]
euu	[u:]
iiie	[i:]
eeaa	[ɛ:]
eeie	[e:]
aaax	ǎ

Consonants

IPA	Narrow	Broad	Simplified	ASCII	ASCII-96	Latin-1
tʃ	č	č	ch	ch	tj	c
tɕ	ć	č	ch	ch (chy)	tj	c
tʂ	ç	č	ch	ch (chr)	tj	c
ts	c	ć	ts	ts	ts	ts
dʒ	ǰ/ǰ̣	ǰ	j	j	dj	j
dʒ̥	ǰ̣/ǰ̣̣	ǰ̣	j	j (jy)	dj	j
dʒ̥̣	ǰ̣̣/ǰ̣̣̣	ǰ̣̣	j	j (jr)	dj	j
dz	ǰ/ʒ	ǰ	dz	dz	dz	dz
ʃ	š	š	sh	sh	sj	ç
ɕ	ś	š	sh	sh (shy)	sj	ç
ʂ	ʂ	š	sh	sh (shr)	sj	ç
ʒ	ž	ž	zh	zh	zj	j
ʒ̥	ẓ̌	ẓ̌	zh	zh (zhy)	zj	j
ʒ̥̣	ẓ̣̌	ẓ̣̌	zh	zh (zhr)	zj	j
β	ḃ	ḃ	bh	bh	bb	v
ð	ḋ	ḋ/ḏ	dh	dh	dd	ð
ɸ	ḡ/ḡ̣	ḡ	gh	gh	q	q
ɸ̥	ḡ̣/ḡ̣̣	ḡ̣	ph	ph	ff	f
θ	ṱ	ṱ/Ṳ	th	th	tt	þ
χ	ḵ/h	ḵ	kh	kh	x	x
ɬ	ḷ	ḷ/Ḹ	ll	ll	ll	ll
ɮ	ḹ	ḹ	l	l (lly)	l	l
ɮ̥	ḹ̣	ḹ̣	l	l (ly)	l	l
ɮ̥̣	ḹ̣̣	ḹ̣̣	l	l (lh)	l	l
ɲ	ṅ	ṅ	ny	ny (ngy)	ny	ny
ɲ̥	ṇ̇	ṇ̇	ny	ny	ny	ny
ŋ	ŋ	ŋ	ng	ng	ng	ng
ŋɡ	ŋɡ	ŋɡ	ngg	ngg	ngg	ngg
j	y	y	y	y	y	y
ɟ	ɟ	ɟ'/ɟ̣	gy	gy	gj	gy
c	ḵ/ḵ'	ḵ'/ḵ̣	ky	ky	kj	ky
ɾ	ɾ/ɽ	ɾ	hr	hr	hr	hr
ɽ	ɽ/ɽ̣	ɽ̣	hl	hl	hl	hl
ɱ	ṁ/ṃ̇	ṁ	hm	hm	hm	hm
ɳ	ṅ/ṇ̇	ṅ	hn	hn	hn	hn
ɳ̥	ṇ̇/ṇ̣̇	ṇ̇	hny	hny	hny	hny
ɳ̥̣	ṇ̣̇/ṇ̣̣̇	ṇ̣̇	hng	hng	hng	hng
ɰ	ɰ/w	ɰ	hw	hw	hw	hw
ç	ɰ/ɰ̣	ɰ̣	hy	hy	hy	hy
ʔ	ʔ					

Modifiers and diacritics

IPA	Description	Narrow	Broad	Simplified	ASCII	ASCII-96	Latin-1
̃	Nasalization. ⁵	ⁿ /̃	ⁿ	<i>m/n</i>	\~/nn/n	\~	~
h	Aspiration.	ˈ/ˈh	<i>h</i>	· <i>h</i>	'h/`h	'h	<i>h</i>
j	Palatal(ized). ⁶	ɣ/ˈ/˘	ˈ/ɣ	<i>y/·y</i>	y	y	<i>y</i>
w	Labialized, for consonants.	w/°	w/ <i>w</i>	<i>w</i>	w	w	<i>w</i>
	Morpheme separator.	-	-	-	-	-	-
	Elision marker. ⁷	,	,	,	,	,	,
	Grapheme separator.			·	'/\`	'	(·)

Prosody

IPA	Description	Narrow	Broad	Simplified	ASCII	ASCII-96	Latin-1
:	Length. ⁸	̄/ː	̄	ː	ːː	ːː/˘˘˘	ˆ
◌̆	Short. ⁹	◌̆			◌x		
ˈ	Primary stress, on vowel. ¹⁰	◌ˈ/◌˘					
ˌ	Secondary stress, on vowel.	◌ˌ/◌˙					
◌̚	Syllabic.	◌̚	(◌̚)	(◌̚)	(◌̚)	(◌̚)	(◌̚)
◌̚̚	Non-syllabic, for vowels.	◌̚̚					

⁵In Yahab nasalization is a feature of *words* rather than of single phonemes. This is indicated with a prefixed ⁿ, e.g. ⁿ*riġ* [r̃iġ] ‘west’.

⁶Prime when distinguished from alveopalatal.

⁷In Classical Sohlub there isn’t really any synchronic difference between elision of unstressed vowels and non-insertion of svarabhakti vowels; both are normally unmarked.

⁸Alternatively and preferred macron for vowels.

⁹N.B. that [ɹ] would be ř!

¹⁰E.g. *äzdrág* with prothetic vowel.

Other diacritics

IPA	Description	Narrow	Broad	Simplified	ASCII	ASCII-96	Latin-1
	Frontness—backness reversal. ¹¹	ö	ö	ö	oe		
ö/ø	Central(ized).	ô	(ô)	(ô)	(eö)		
◌̥	Closer/raised, for vowels.	◌̥	(◌̥)	(◌̥)	◌̥i/◌̥u		
◌̦	Open/lowered, for vowels.	◌̦	(◌̦)	(◌̦)	◌̦a		
◌̜	Rounded, for vowels.	◌̜	◌̜	◌̜	◌̜o	◌̜	
◌̚	Voiceless. ¹²	◌̚	◌̚	h◌̚	h◌̚	h◌̚	h◌̚
◌̥̚	Voiced. ¹³	◌̥̚					
(ʃ)	Palatoalveolar. ¹⁴	◌̥̚	◌̥̚	◌̥̚h	◌̥̚h		
(ç)	Alveopalatal. ¹⁵	◌̥̚	◌̥̚	◌̥̚h	◌̥̚hy		
◌̥̚	Apical, for consonants.	◌̥̚					
◌̥̚	Laminal, for consonants.	◌̥̚					
v	Velar(ized). ¹⁶	◌̥̚	(◌̥̚)		(◌̥̚)	(◌̥̚)	
ʁ	Uvular(ized). ¹⁷	◌̥̚					
◌̥̚	Fricative/fricativization. ¹⁸	◌̥̚/◌̥̥̚	◌̥̥̚	◌̥̚h	◌̥̚h	(◌̥̚)	
(kx)	Affrication. ¹⁹	◌̥̥̚/◌̥̥̥̚	◌̥̥̥̚	◌̥̥̚h ²⁰	◌̥̥̚`◌̥̚h		
◌̥̥̚	Approximant. ²¹	◌̥̥̥̚					
◌̥̥̥̚	Retroflex/retracted cons.	◌̥̥̥̥̚			(◌̥̥̥̥̚r)		
◌̥̥̥̥̚	Advanced (or fronted) cons.	◌̥̥̥̥̥̥̚					

¹¹Thus *ü* is front rounded, *i* is back unrounded (or central).

¹²Note *ɸ* for [ɸ]!

¹³Note *ɸ* for [ɸ]!

¹⁴In Broad Cl.S. *ʃ* etc. and Chid. *ʃ* etc. are written alike as *ʃ* etc.

¹⁵Also *ç* *ç* *ç* for mediopalatals. N.B. *ç* *ç* are fricatives, *y* *y* are approximants.

¹⁶Note that *k* *g* are normally velar. Hilib distinguishes velar *ɸ*, also written *ɸ* (and *ɸ* in ASCII), from front and weakly palatalized *ɸ*.

¹⁷Note that *ç*/*ç* *ç*/*ç* are normally uvular in Classical Sohlob and Hilib.

¹⁸Alternatively above for letters with descenders.

¹⁹Thus *ɸ* [pɸ], *ç* [kç] etc. (Sibilant affricates are preferredly based on *c* *ç*/*ç* but can be based on *ç* *ç*.)

²⁰Contrast *kçh* [kç] and *kçh* [x:].

²¹Hilib *r* is normally a retroflex approximant.

Overview of the native orthographies of the Sohlob languages

The native script is an alphabet written in lines running right to left. Like the syllabaries which it replaced the alphabet was originally heavily underspecifying. Most letters could stand for two phonemes, with values left to be inferred by the reader from context. In particular fortis vs. lenis stops, velars vs. palatals and vowels vs. semivowels were not distinguished. Soon after the adoption of the alphabet scribes began to double consonant letters to distinguish fortis stops from lenis stops. This doubling was later extended to the spelling of the second sound in each of the pairs *r—l*, *z—j*, *h̃—š*, *h—s*, for semivowels vs. vowels and word initially for *v* vs. *f*. When the alphabet was adopted in the west scribes began to use an overdot to distinguish the front rounded vowels of the western languages from back rounded vowels, and soon also to indicate low mid vowels. In time the overdot was also adopted in the writing of Classical Sohlob to distinguish *ä* from *ī*. The overdot soon also became used to indicate an omitted letter doubling, and eventually the dot could be used interchangeably with doubling for consonant spellings. After the invention of printing doubling more and more fell out of use in favor of dotting, except in the case of ⟨s⟩ which was still usually written as a somewhat stylized ⟨hh⟩. In the writing of Hilīb, which began rather late, ⟨rr⟩ was used for the velar *ɫ*, in contrast with both ⟨r⟩ for dental *l* and ⟨r⟩ for *r*. In the centuries after the invention of printing Chidilib spelling went through a reform. Fortis stops came to be indicated with an ⟨h⟩ after or even on top of the base letter, similar to how *l* and *r* had traditionally been written as ⟨l⟩ and ⟨r⟩ with a preceding ⟨h⟩. The convention of indicating voicelessness with a preceding ⟨h⟩ had extended to the voiceless sonorants of Chidilib much earlier. At the same time *b* and *d* were written as ⟨ḅ⟩ and ⟨ḍ⟩ while the old unadorned ⟨b⟩ and ⟨d⟩ were used for *v* and *ɖ*. Since Chidilib had lost the distinction between *h* and *h̃* the dot on *š* was omitted.

Vowels

Roman	Cl. Sohlob	Chidilib	Hilīb	Linjīb
<i>a</i>	ʌ	ʌ	ɑ ʌ	ɑ ʌ
<i>ä</i>	ä̇		ä̇	
<i>e</i>		ɛ ɛ		ɛ ɛ
ɛ				ä̇
<i>i</i>	ɨ	ɨ	ɨ	ɨ
<i>ī</i>	ī̇ ɨ		ɨ	
<i>o</i>	ɑ̇ ɔ	ɔ ɔ	ɑ̇ ɔ	ɔ ɔ
ɔ				ʌ
<i>ö</i>			ɔ̇	ɔ̇
<i>u</i>	ʊ	ʊ	ʊ	ʊ
<i>ü</i>			ʊ̇	ʊ̇

Consonants

Roman	Cl. Sohlob	Chidilib	Hilīb	Linjīb
<i>b</i>	ⲃ	Ⲅ	ⲃ	ⲃ
<i>d</i>	Ⲅ	ⲅ	Ⲅ	Ⲅ
<i>g</i>	ⲅ	Ⲇ	ⲅ	ⲅ
<i>p</i>	Ⲅ	Ⲇ	Ⲅ	Ⲅ
<i>t</i>	ⲅ	ⲇ	ⲅ	ⲅ
<i>k</i>	Ⲇ	ⲇ	Ⲇ	Ⲇ
<i>ǰ/ǰ̃</i>	Ⲇ	Ⲇ	Ⲇ	
<i>č</i>	ⲇⲄ	ⲇⲄ	ⲇⲄ	
<i>f</i>	ⲇ	ⲇ	ⲇ	ⲇ
<i>t̥</i>				Ⲇ
<i>s</i>	Ⲉ	Ⲉ	Ⲉ	Ⲉ
<i>š</i>	ⲇ	ⲇ	ⲇ	
<i>h</i>	Ⲉ	Ⲉ	Ⲉ	ⲇ
<i>v</i>	ⲇ	ⲃ	ⲇ	ⲇ
<i>d̥</i>		Ⲅ		Ⲇ
<i>z</i>	Ⲇ	Ⲇ	Ⲇ	
<i>ǰ̃</i>	Ⲇ			ⲇ
<i>r</i>	Ⲇ	Ⲇ	Ⲇ	Ⲇ
<i>l</i>	Ⲇ	Ⲇ	Ⲇ	Ⲇ
<i>l̥</i>			ⲆⲆ	
<i>r̥</i>		Ⲇ		
<i>l̥</i>	Ⲇ	Ⲇ	Ⲇ	
<i>w</i>	Ⲇ	Ⲇ		Ⲇ
<i>y</i>	Ⲇ	Ⲇ	Ⲇ	Ⲇ
<i>m</i>	Ⲇ	Ⲇ	Ⲇ	Ⲇ
<i>n</i>	Ⲇ	Ⲇ	Ⲇ	Ⲇ
<i>ñ</i>	ⲆⲆ	ⲆⲆ	ⲆⲆ	
<i>ŋ</i>	ⲆⲆ	ⲆⲆ	ⲆⲆ	ⲆⲆ
<i>ŋg</i>	ⲆⲆ	ⲆⲆ	ⲆⲆ	ⲆⲆ
<i>w̥</i>		Ⲇ		
<i>ɣ</i>		Ⲇ		
<i>m̥</i>		Ⲇ		
<i>n̥</i>		Ⲇ		
<i>ñ̥</i>		ⲆⲆ		
<i>ŋ̥</i>		ⲆⲆ		