

Instructor:  
Email:  
Website (for Homework, Handouts, etc.):  
Location:  
Class Hours:  
Office Hours:

Jack Isaac Rabinovitch  
[jrabinovitch@g.harvard.edu](mailto:jrabinovitch@g.harvard.edu)  
[jackisaacrabinovitch.github.io](https://github.com/jackisaacrabinovitch)  
Barker Center 218 (Fong Room)  
TuTh 9:00–10:00 AM  
TBD

# Writing Systems and Language Conservation

## Week 3: Guiding Principles of Writing Systems

### 1 Goals for this Week

Here is what we will learn the week:

- (1)
  - a. More on the categorization and typology of writing systems (syllabaries, featural alphabets)
  - b. Script transmission, how do people fit languages into writing systems
  - c. Defective scripts and what we don't include in a script
  - d. What does it mean to be a 'good fit' for a writing system
  - e. Social reasons to keep non-phonemic scripts and diasystems

### 2 Syllabaries: Hiragana, Katakana, and Cherokee

**Syllabaries** are scripts for which each sign represents a syllable. While abugidas represent a syllable with a sign and diacritic pair, syllabaries have unique symbols for each syllable.

Japanese has three writing systems used in tandem. **Kanji** are Chinese characters used to represent the roots of words. Rather than Chinese, where characters correspond to a single syllable, Kanji can correspond to multisyllabic strings.

**Hiragana** are syllabic signs which represent single syllables, typically non-root morphology; they derive from simplified versions of quickly written Kanji which represent those sounds in their Sino-Japanese readings. **Katakana** are derived the same way, but represent ideophones, onomatopoeia, and loanwords.

Diacritical marks are used in Hiragana and Katakana to mark voicing and frication of the consonant of a syllable. A single character represents a final [n], the only non-geminate coda in Japanese. Why would these two attributes indicate that Hiragana and Katakana are not 'pure' syllabaries?

Two other syllabaries in common use today are the Cherokee syllabary and the Yi syllabary:

The Cherokee syllabary was invented by Sequoyah in the early 1800s to write the Cherokee language. While the signs used are reminiscent of the Latin alphabet, they have no sound correspondence: Sequoyah had access to English writing but was illiterate at the time of making the script.

Whereas Cherokee and Japanese have relatively simple syllable structure, Nuosu Yi is a tonal language with multiple kinds of codas. Thus the Yi syllabary contains 1,164 signs in order to represent every unique syllable.

What are the benefits of a syllabary, the costs? What is the dividing line between an abugida and a syllabary? Syllabaries commonly represent CV structure, but none represent only VC structure, why is that? Any guess at how many signs would English need to have a 'pure' syllabary?

无 ん	和 わ	良 ら	也 や	末 ま	波 は	奈 な	太 た	左 さ	加 か	安 あ
	為 ゐ	利 り		美 み	比 ひ	仁 に	知 ち	之 し	幾 き	以 い
		留 る	由 ゆ	武 む	不 ふ	奴 ぬ	川 つ	寸 す	久 く	宇 う
	恵 ゑ	礼 れ		女 め	部 へ	祢 ね	天 て	世 せ	計 け	衣 え
	遠 ゑん	呂 ろ	与 よ	毛 も	保 ほ	乃 の	止 と	曾 そ	己 こ	於 お

Figure 1: The Derivation of Hiragana from Kanji (Jeannebluemon-heo, CC BY-SA 4.0)

ア 阿	イ 伊	ウ 宇	エ 江	オ 於
カ 加	キ 機	ク 久	ケ 介	コ 己
サ 散	シ 之	ス 須	セ 世	ソ 曾
タ 多	チ 千	ツ 川	テ 天	ト 止
ナ 奈	ニ 仁	ヌ 奴	ネ 祢	ノ 乃
ハ 八	ヒ 比	フ 不	ヘ 部	ホ 保
マ 末	ミ 三	ム 牟	メ 女	モ 毛
ヤ 也		ユ 由		ヨ 與
ラ 良	リ 利	ル 流	レ 礼	ロ 呂
ワ 和	ヰ 井		エ 恵	ヲ 乎
ン 尔				

Figure 2: The Derivation of Katakana from Kanji (Pmx, CC BY-SA 3.0)

Basic hiragana characters

	<i>a</i>	<i>i</i>	<i>u</i>	<i>e</i>	<i>o</i>
∅	あ	い	う	え	お
<i>k</i>	か	き	く	け	こ
<i>s</i>	さ	し	す	せ	そ
<i>t</i>	た	ち	つ	て	と
<i>n</i>	な	に	ぬ	ね	の
<i>h</i>	は	ひ	ふ	へ	ほ
<i>m</i>	ま	み	む	め	も
<i>y</i>	や	<sup>[5]</sup> ゆ	<sup>[5]</sup> ゆ		よ
<i>r</i>	ら	り	る	れ	ろ
<i>w</i>	わ	ゐ	<sup>[5]</sup> ゐ	ゑ	を

Figure 3: Hiragana syllables can be organized by onset and vowel

Consonant	<i>a</i>			<i>e</i>	<i>i</i>	<i>o</i>	<i>u</i>	<i>v</i> [ə]
∅	D a			R e	T i	ʃ o	ʊ u	i v
<i>g</i> / <i>k</i>	ʒ ga	ʉ ka		ʀ ge	ʎ gi	A go	J gu	E gv
<i>h</i>	ʰ ha			ʔ he	ʙ hi	ʈ ho	ʦ hu	ʡ hv
<i>l</i>	W la			ʤ le	ʀ li	G lo	M lu	ʁ lv
<i>m</i>	ʦ ma			ʈ me	H mi	ʃ mo	ʎ mu	G* mv
<i>n</i> / <i>hn</i>	θ na	ʈ hna	G nah	ʌ ne	h ni	Z no	ʁ nu	ʊ nv
<i>qu</i> [kʷ]	I qua			ʊ que	ʈ qui	ʎ quo	ʊ quu	ɛ quv
<i>s</i>	ʊ s	ʉ sa		4 se	b si	ʈ so	ʂ su	R sv
<i>d</i> / <i>t</i>	ʌ da	W ta		ʂ de	ʌ te	ʎ di	ʎ ti	V do
<i>dl</i> / <i>tl</i> [dʒ] / [tʃ]	ʂ dla	ʌ tla		L tle	C tli	ʈ tlo	ʈ tlu	P tlv
<i>ts</i> [ts]	C tsa			ʎ tse	ʈ tsi	K tso	d tsu	ʌ tsv
<i>w</i> [w]	G wa			ʂ we	ʈ wi	ʈ wo	ʎ wu	6 wv
<i>y</i> [j]	ʊ ya			ʂ ye	ʂ yi	ʈ yo	G yu	B yv

Figure 4: Cherokee syllables can be organized by onset and vowel

### 3 Featural Systems: Hangul

While Chinese characters are still used as Kanji in Japanese, modern Korean has made an switch to Hangul/Hangeul. Hangul was invented in the 1443 by King Sejong the Great of the Joseon Dynasty of Korea and is a featural alphabet.

**Featural Systems:** A writing system in which signs are not arbitrary but encode phonological features of the phonemes that they represent

Consonants and vowels were supposedly designed to represent the shape of the mouth in making said sounds; others have argued that the consonant graphs are actually derived from Tibetan.

Consonants and vowels are each represented by their own graphs, but the graphs are combined in a specific way to make syllable blocks, such that each syllable is represented by a composite sign.

			<i>Bilabial</i>	<i>Alveolar</i>	<i>Alveolo-palatal</i>	<i>Velar</i>	<i>Glottal</i>
Obstruent	Stop (plosive)	Lax	p (ㅍ)	t (ㅌ)		k (ㄱ)	
		Tense	p* (ㅍ*)	t* (ㄷ*)		k* (ㄱ*)	
		Aspirated	p <sup>h</sup> (ㅍ <sup>h</sup> )	t <sup>h</sup> (ㄷ <sup>h</sup> )		k <sup>h</sup> (ㄱ <sup>h</sup> )	
	Fricative	Lax		s (ㅅ)			h (ㅎ)
		Tense		s* (ㅅ*)			
	Affricate	Lax			tʃ (ㅈ)		
		Tense			tʃ* (ㅈ*)		
		Aspirated			tʃ <sup>h</sup> (ㅈ <sup>h</sup> )		
Sonorant	Nasal		m (ㅁ)	n (ㄴ)		ŋ (ㅇ)	
	Liquid (lateral approximant)			l (ㄹ)			

Figure 5: The consonant correspondences of Hangul

What are the benefits of having a featural system? What are the costs? Does Korean do a good job of representing its sounds iconically? Why? What is the boundary between a featural system and an alphabet?

# 한글

한글은한국어의고유문자입니다.  
Hangeul is the native alphabet of the Korean language.

## Hangeul

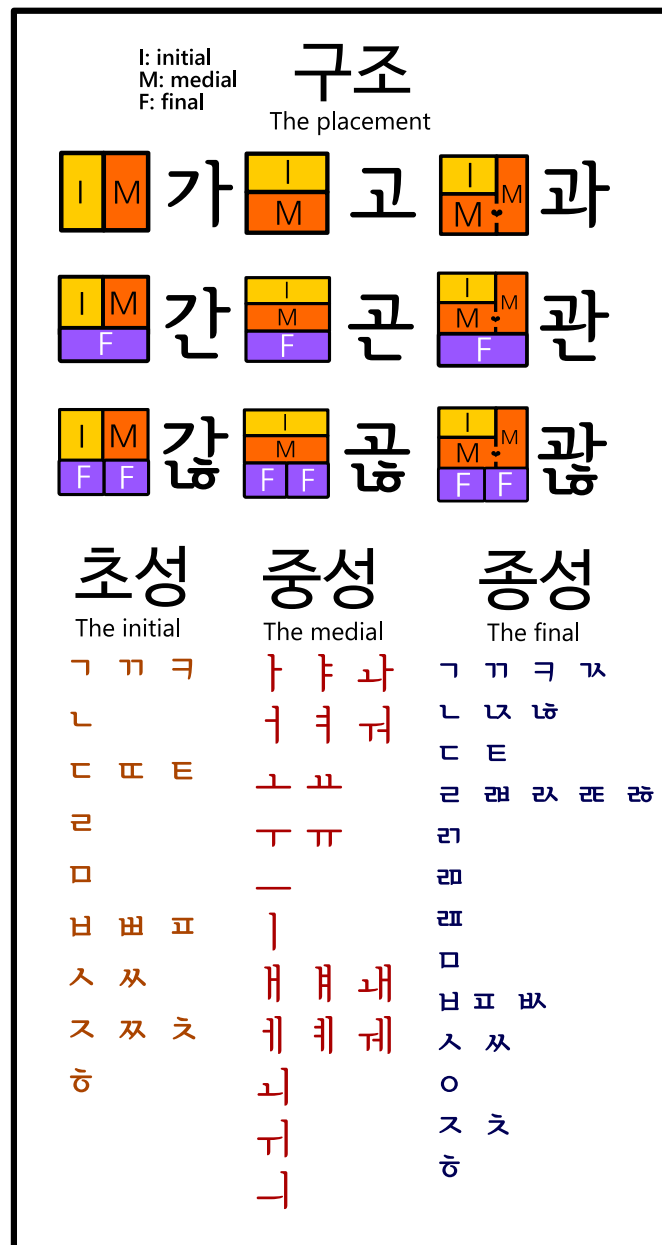


Figure 6: The syllabic structure of Hangeul character blocks (Metal-slick, CC BY-SA 4.0)

## 4 The Mechanisms of Script Transmission

### 4.1 Making Vowels: Uyghur adoption of Arabic script

As we have seen, Arabic script is an (imperfect) abjad, where vowels do not necessarily have their own symbol, and are typically inferred from context. For the Turkic languages, many of which adopted the Arabic script during the spread of Islam, such a system led to too much ambiguity; as a result many of these languages modified the script accordingly.

One example of this which is still used today is the Arabic script as used for Uyghur: just like how Greek modelled its vowel letters after Phoenician consonantal letters, Uyghur uses Arabic consonantal letters to represent vowels.

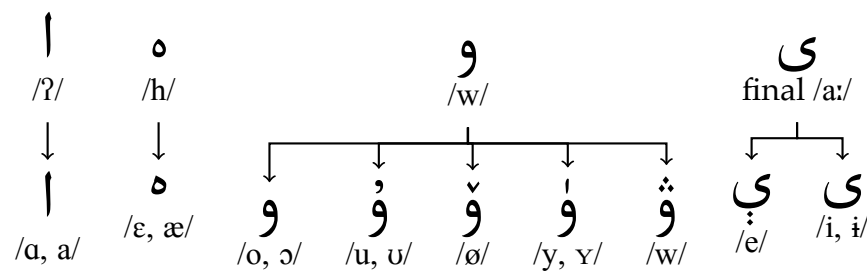


Figure 7: Arabic signs and original values (above) and their corresponding Uyghur signs and values (below).

Note that the Arabic letter waw ۋ represents all rounded vowels (and semi-vowels) in Uyghur, differentiated only by diacritics to determine frontness and height. How is this similar or different to diacritics used in abugidas? What kind of (phonological) meaning does ۋ have in Uyghur?

There is a line of reasoning that abjads always beget alphabets, due to alphabet's nature as "more complete". Does the case of Uyghur offer evidence for such a hypothesis? Why or why not? Could you imagine a language adopting an alphabet and turning it into an abjad? What would be required for such a change?

## 4.2 Showing Mutations: Irish adoption of the Latin script

Irish has a series of consonant mutations, including lenition (*séimhiú*), which occurs in special morphosyntactic environments (such as after certain prepositions and possessive pronouns).

Normal	<i>Séimhiú</i>	Gloss
<i>peann</i> /pʲaːn̪ˠ/	<i>pheann</i> /fʲaːn̪ˠ/	"pen"
<i>teach</i> /tʲax/	<i>theach</i> /hax/	"house"
<i>ceann</i> /caːn̪ˠ/	<i>cheann</i> /çaːn̪ˠ/	"head"
<i>bean</i> /bʲaːn̪ˠ/	<i>bhean</i> /vʲaːn̪ˠ/	"woman"
<i>droim</i> /d̪r̪ʲiːm̪ˠ/	<i>dhroim</i> /ɣr̪ʲiːm̪ˠ/	"back"
<i>glúin</i> /g̪l̪uːn̪ˠ/	<i>ghlúin</i> /ɣl̪uːn̪ˠ/	"knee"
<i>máthair</i> /mʲaːhəɾʲ/	<i>mháthair</i> /waːhəɾʲ/, /vaːhəɾʲ/	"mother"
<i>súil</i> /sʲuːl̪ˠ/	<i>shúil</i> /huːl̪ˠ/	"eye"
<i>freagra</i> /fʲr̪ʲaɡr̪ʲə/	<i>fhreagra</i> /r̪ʲaɡr̪ʲə/	"answer"

Figure 8: Examples of Irish lenition with orthography and IPA.

Such a mutation meant that a single word would have different pronunciations given the context that it appears in. (How) would such differences be represented in the script?

Gaelic scripts were typefaces of the Latin alphabet that were typically used to write Classical Gaelic (the precursor to Irish and Scottish Gaelic) until around the 1700s in Scotland and 1900s in Ireland.

**Typeface:** the design of lettering of a script (including variations in size, weight, slope, width, etc.)



Δ b c d e f g h i l m n o p r s t u  
 Δ b c d e f g h i l m n o p r r t u  
 a b c d e f g h i l m n o p r s t u

### *Lenited consonants*

ḃ ḃ ċ ċ ḋ ḋ ḟ ḟ ġ ġ ṁ ṁ ṗ ṗ ṡ ṡ ṫ ṫ  
 bh ch dh fh gh mh ph sh th

Figure 9: Gaelic script and its non-Gaelic equivalents. By Moillead-óir - Own work, CC BY-SA 4.0

In Gaelic scripts, lenition was denoted with a dot placed above the consonant. In non-Gaelic (modern) script, it is represented by the addition of an <h> after the lenited consonant.

Note that for some sounds, such as /x/, the written form <ch> is the same in both places where it derived from lenition (*cheann*) and not (*teach*). For others, such as f, the written form when derived from lenition <ph> (*pheann*) is distinct from when it is written in a non-lenited position <f> (*freagra*). Because lenition is simply noted with some modification of the letter, the same value derived through lenition from two different underlying sounds were represented differently (such as <mh> and <bh> which both represent /v/)

What are the benefits of marking mutations of sound in such a way? Would it be better to represent sounds with a one-to-one correspondence rather than by showing their derivation? Why? Does the Gaelic or non-Gaelic form seem more intuitive or natural?

Would you consider Gaelic 'scripts' to be a separate script from other Latin scripts, or just a different typeface? Where is the boundary between one and the other? What does the dot in Gaelic script and <h> in the non-Gaelic script for Irish represent?

### 4.3 Dealing with large Inventories: Xhosa adoption of Latin script

Xhosa (isiXhosa) is a Nguni Bantu language spoken in South Africa. It has an incredibly large consonant inventory, including ejectives, implosives, various kinds of voicing, and click consonants.

		Labial	Dental/Alveolar		Post-alveolar	Velar		Glottal
			central	lateral		central	lateral	
Click	tenuis/ejective <sup>[17]</sup>		kl' <c>	klɭ' <x>	kl' <q>			
	aspirated		klʰ <ch>	klɭʰ <xh>	klʰ <qh>			
	slack voice		ǵlʰ <gc>	ǵɭʰ <gx>	ǵlʰ <gq>			
	nasal		ŋl <nc>	ŋɭ <nx>	ŋ! <nq>			
	slack-voice nasal <sup>[note 1]</sup>		ŋlʰ <ngc>	ŋɭʰ <ngx>	ŋlʰ <ngq>			
	prenasalized ejected <sup>[18][note 2]</sup>		ŋl' <nkc>	ŋɭ' <nkx>	ŋl' <nkq>			
Plosive	tenuis/ejective	p' <p>	t' <t>		t͡ʰ' <ty>	k' <k>		ʔ <sup>6</sup>
	aspirated	pʰ <ph>	tʰ <th>		t͡ʰʰ <tyh>	kʰ <kh>		
	slack voice	b̥ʰ <bh>	ɖ̥ʰ <d>		ɖ̥ʰʰ <dy>	ɡ̥ʰ <g>		
	implosive	ɓ <b>						
Affricate	ejective		ts' <ts>		t͡ʃ' <tsh>	kx' <kr>	kl̥' <kl> <sup>5</sup>	
	aspirated		tsʰ <ths>		t͡ʃʰ <thsh>	kxʰ <krh>		
	slack voice		ɖ͡ʒʰ <dz> <sup>3</sup>		ɖ͡ʒʰʰ <j>			
Fricative	voiceless	f <f>	s <s>	ɬ <hl>	ʃ <sh>	x <rh>		h <h>
	slack voice	v̥ <v <sup>7<td>z̥ &lt;z&gt;</td><td>ɬ̥ &lt;dl&gt;</td><td>ʃ̥ &lt;zh&gt;<sup>2</sup></td><td>χ̥ &lt;gr&gt;</td><td></td><td>ɦ &lt;h&gt;</td></sup>	z̥ <z>	ɬ̥ <dl>	ʃ̥ <zh> <sup>2</sup>	χ̥ <gr>		ɦ <h>
Nasal	fully voiced	m <m>	n <n>		ɳ̹' <ny>	ŋ <ngʷ>		
	slack voice	m̥ <mh>	ɳ̥ <nh>		ɳ̥̹' <nyh>	ŋ̥ <ngʰ> <sup>4</sup>		
Liquid	fully voiced		r <r> <sup>1</sup>	ɭ <l>				
	breathy voiced		ɾ̥ <r> <sup>1</sup>	ɭ̥ <lh>				
Semivowel	fully voiced				j <y <sup>8<td>w &lt;w&gt;</td><td></td><td></td></sup>	w <w>		
	slack voice				j̥ <yh>	w̥ <wh>		

Figure 10: The consonant inventory of Xhosa, with orthography

Xhosa has adopted the Latin alphabet, but has not added any letters beyond the 26 we see in Modern English and the use of the apostrophe <'>.

<h> represents aspiration after stops and clicks (orthographic <p, t, k, c, x, q>), breathy voice after nasals (orthographic <m, n, ng>), and a glottal fricative when on its own.

**Featural notation:** writing system that include symbols which represent individual features of phonemes rather than phonemes.

<d> and <g> represent the coronal and velar slack voiced plosives respectively; <bh> represents the labial one, while <b> represents the implosive. This doesn't follow the same kind of cross-place patterning of featural notation, why might this breaking of featural notation be useful?

Xhosa uses <ng'> to represent /ŋ/, but <ng> to represent the breathy /ɿ/. Why is the apostrophe there in the first place? Why is it absent in the second?

Orthography	Phonology	Resulting Correspondence Assumption
<finger>	/fɪŋgɹ̥/	∴ <ng> ≡ /ŋg/
<singer>	/sɪŋɹ̥/	∴ <ng> ≡ /ŋ/

Figure 11: Deficiency in English orthography: ambiguity arising from multiple readings of <ng> in English

#### 4.4 Two cases of tone representation: Vietnamese and Bai adoptions of Latin script

Bai and Vietnamese are two languages which have complex tone systems and employ the Latin script for writing.

For Bai, tones are represented by letters written after the syllable; some of these take on two roles, such as <l> which also represents /l/ in initial position in Bai. There is one 'default' tone which has no corresponding letter.

For Vietnamese, tones are represented by diacritics written over the nucleic vowel of a given syllable; there are other diacritics which represent changes in vowel quality; these diacritics stack to give complex multi-diacritic marked vowels in Vietnamese. There is one 'default' tone which has no corresponding diacritic.

Why choose to use letters to represent tones? Why use diacritics? What parallels does this have with marking vowels in alphabets and abugidas? What does the placement of supersegmental features such as tone say about how people perceive what the tone is tied to?

Why choose a tone to not be represented overtly? What might the change of this 'underlying tone' in the 1982 and 1993 spellings of Bai tell us about Bai's tonal system?

Is there any reason why Bai chose those specific letters to represent the tones they use? How about the diacritics that Vietnamese employs to represent tone? What is the benefit of Vietnamese using the grave accent for falling and acute accent for rising tone? What about the 'dot below'?

Pitch contour and phonation	1982 spelling	1993 spelling
high level (55), <i>modal</i>	-l	-l
mid level (33), modal	-x	-x
mid falling (31), <i>breathy</i>	-t	-t
mid rising (35), modal	-f	-f
mid-low falling (21), <i>harsh</i>	(unmarked)	-d
high level (55), <i>tense</i>	-rl	-b
mid-high level (44), tense	-rx	(unmarked)
mid-high falling (42), tense	-rt	-p
mid falling (32), modal		-p/-z

Figure 12: Bai Tones

Diacritic	Symbol	Name	Contour
unmarked	N/A	Ngang	mid level, ˥
<i>grave accent</i>	à	Huyền	low falling, ˨˩˦
<i>hook above</i>	â	Hỏi	mid falling, ˨˩˦ (Northern); dipping, ˨˩˦˥ (Southern)
<i>tilde</i>	ã	Ngã	glottalized rising, ˥˩˦˥˩ (Northern); slightly lengthened Dấu Hỏi tone (Southern)
<i>acute accent</i>	á	Sắc	high rising, ˥˩˦˥
<i>dot below</i>	ạ	Nặng	glottalized falling, ˥˩˦˥˩ (Northern); low rising, ˨˩˦˥ (Southern)

Figure 13: Vietnamese Tones