


Writing Systems

Zdeněk Žabokrtský

partially based on slides of Magda Ševčíková and Anna Nedoluzhko

 January 4, 2024



Charles University
Faculty of Mathematics and Physics
Institute of Formal and Applied Linguistics



unless otherwise stated

A warm-up exercise

- (1) །ཀམ
- (2) λογική
- (3) ተጽጻጺዕዮ
- (4) 𐤀𐤁𐤂
- (5) nguòì
- (6) ひらがな
- (7) 𐌆𐌇𐌈
- (8) 인간은
- (9) ყველა
- (10) 𑆑𑆒
- (11) 𑆑𑆒𑆓
- (12) 漢
- (13) ДОЛЖНЫ
- (14) սարդիկ
- (15) الناس

- (a) Chinese
- (b) Arabic
- (c) Hebrew
- (d) Armenian
- (e) Neo-Tifinagh
- (f) Greek
- (g) Latin
- (h) Cyrillic
- (i) Ethiopic
- (j) Georgian
- (k) Thai
- (l) Hiragana
- (m) Devanagari
- (n) Mongolian
- (o) Hangul

Intro

Historical development of writing systems

Selected writing systems used today

A global view on world's writing systems

Final remarks

Intro

- writing – a system of recording language by means of visible or tactile marks
- the communication of meaning is the primary purpose of most writing
- today, more communication takes place in the written than in the oral form (Internet)
- writing is not a 'minor' form of communication

Writing vs. speech

- discrete
 - timeless
 - autonomous
 - permanent
 - visible
 - produced by hand
- continuous
 - bound to utterance time
 - contextual
 - temporal
 - audible
 - produced by voice

Writing systems of the world

- enormous diversity
- writing systems virtually located on a scale between semiography and phonography
 - semiography – meaning-based writing
 - phonography – sound-based writing
 - ideal, but not realizable writing systems
- no writing system is immediately comprehensible without instruction, initiation needed
- quite different learning curves/cognitive loads though

Five basic types of modern writing systems

1. **alphabets** – a symbol corresponds to a sound (a consonant or a vowel)
 - example: Latin(=Roman)/Cyrilic/Greek scripts
2. **abjads** – a symbol corresponds to a consonant (vowels to be inferred by readers)
 - example: Arabic and Hebrew scripts
3. **abugidas** – a symbol corresponds to a consonant-vowel pairing (vowels rendered as modifications of the base consonant shape)
 - example: Devanagari (used for Hindi)
4. **syllabaries** – a symbol corresponds to a syllable or a mora
 - example: Japanese Kana
5. **logo-syllabaries** – a symbol corresponds to a word or a morpheme
 - example: Chinese characters

Other basic notions

- a script = a set of symbol/character/sign/letter/glyph/marks
- an orthography = rules (conventions) for spelling, punctuation, capitalization...
- a writing system = a script + an orthography

Historical development of writing systems

Proto-writing – visible communication marks



Things



Wampums (shell beads)



Pictograms (cave paintings)



Quipu from Inca people (talking knots)

Invention of writing

- a relatively recent invention
- invented independently (at least) four times:
 - Mesopotamia – 3400 BC
 - Egypt – 3200 BC
 - China – 1300 BC
 - Mesoamerica – 900 BC
- very gradual
- writing – the first language technology!

Sumerian writing

- southern Mesopotamia
- clay tables, heavy but indestructible
- economic records (more than 80 % of the tablets)



Sumerian writing

- early tables – the primary referent of the signs were physical objects (animals and their body parts, tools, plants...)
- pictograms developed into cuneiform signs; around 1000 signs
- original meaning: concrete referents, abstract meaning added later (star → God)
- phonetic value added too: a picture of hand, 'hand' in Sumerian: 'šu', a picture of hand = phonetic syllable 'šu'

	3200 BCE	3000 BCE	2400 BCE	1000 BCE
sag 'head'				
gin 'to walk'				
šu 'hand'				
še 'barley'				
ninda 'bread'				
a 'water'				
ud 'day'				
mušen 'bird'				

Egyptian hieroglyphs

- probably developed from preliterate visual art (proto-hieroglyphic paintings 3500 BC...)
- 900 signs in the Middle Kingdom period (around 2000 BC)
- combined logographic, syllabic and alphabetic elements (e.g. phonetic glyphs representing consonants)

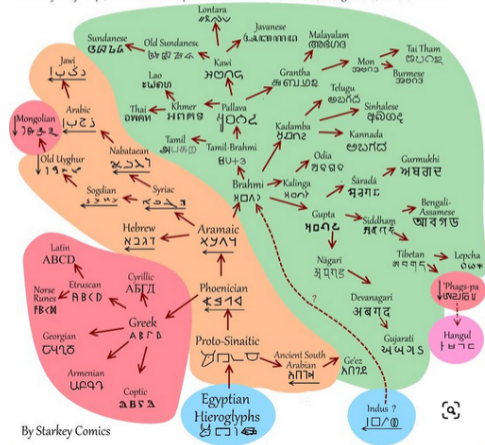


From hieroglyphs to today's writing systems

The ABCD Family Tree

- True Alphabets
- Abjads
- Featural Alphabets
- Abugidas
- Logograms

This infographic shows how 57 different scripts are descended from ancient Egyptian hieroglyphs. The colours show the type of script, and the red arrows show how the scripts are related. For scripts which are not read from left to right, the black arrows show the direction of writing. Where possible, letters corresponding to the A, B, C, and D symbols in the Latin alphabet are shown. In nearly every script, these letters correspond to sounds similar to the /a/, /b/, /g/ or /k/, and /d/ sounds.

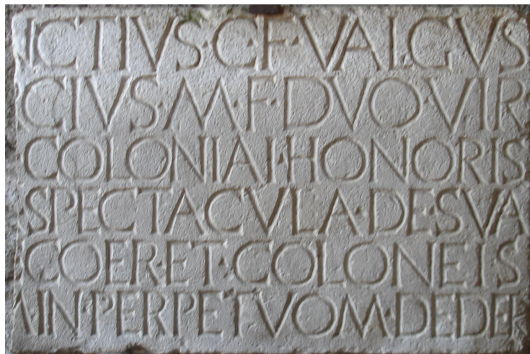


By Starkey Comics

Selected writing systems used today

Latin (Roman) alphabet

a Pompeii amphitheatre inscription



various modifications:

- sometimes more consonants need to be distinguished:
 - Wykład o systemach pisma
 - **The** lecture about writing systems
- or more vowels:
 - Előadás az írásrendszerekről
- or both:
 - Ein Vortrag **über** **S**chriftsysteme
 - Přednáška o písmových systémech
- or tones:
 - Bài giảng về hệ thống chữ viết
- or clicks:
 - Kung (a language spoken in Namibia)

not a perfect (one-to-one) alphabet

- a single symbol pronounced differently in English: cat, city, cello, ocean, muscle
- many non-1-to-one exceptions even in relatively phonetic orthography of Czech (i/y, ú/ů, mně/mě, s/z, d/t ...)
- a crosslingual perspective: (almost) the same sound written differently across languages
– č cz tsch cs ch

a sample from UDHR (the 1st article)

सभी मनुष्यों को गौरव और अधिकारों के मामले में जन्मजात स्वतन्त्रता और समानता प्राप्त है। उन्हें बुद्धि और अन्तरात्मा की देन प्राप्त है और परस्पर उन्हें भाईचारे के भाव से बर्ताव करना चाहिए।

Devanagari, cont.

- a left-to-right abugida
- based on the ancient Brahmi script,
- used in a number of languages the northern Indian subcontinent (Hindi, Marathi, Kashmiri, Nepali...)
- round shapes, a horizontal line connecting letters
- 48 primary characters, including 14 vowels and 34 consonants

Devanagari, cont. – an example of vowels in Hindi

अ	आ	इ	ई	उ	ऊ	ए
a	ā	i	ī	u	ū	e
[ʌ]	[a]	[i]	[i:]	[u]	[u:]	[e]
प	पा	पि	पी	पु	पू	पे
pa	pā	pi	pī	pu	pū	pe
ऐ	ओ	औ	अं	अः	अँ	ऋ
ai	o	au	aṅ	aḥ	āṅ	r̥
[æ:]	[o]	[ɔ:]	[aŋ]	[əh]	[ā:]	[r̥]
पै	पो	पौ	पं	पः	पाँ	पृ
pai	po	pau		paḥ	pāṅ	pr̥

a sample from UHDR in Korean

모든 인간은 태어날 때부터 자유로우며 그 존엄과 권리에 있어 동등하다. 인간은 천부적으로 이성과 양심을 부여받았으며 서로 형제애의 정신으로 행동하여야 한다.

Hangul, cont.

- used for Korean (both in South and North Korea)
- created in 1443 CE by King Sejong the Great, letters for consonants reflect the shape of the speech organs
- modern Hangul: 14 consonant letters and 10 vowel letters
- technically an alphabet (one symbol – one sound)...
- ...however, with letters written in morphosyllabic blocks (at least one consonant and one vowel), not separately
- historically written top-to-bottom, nowadays left-to-right

Korean Alphabet Chart

Consonants	Vowels									
	ㅏ (a)	ㅑ (ya)	ㅓ (o)	ㅕ (yo)	ㅗ (oh)	ㅛ (yo)	ㅜ (ow)	ㅠ (you)	ㅡ (er)	ㅣ (ee)
ㄱ(G)	가	갸	거	겨	고	교	구	규	그	기
ㄴ(N)	나	냐	너	녀	노	뇨	누	뉴	느	니
ㄷ(D)	다	댜	더	더	도	뎐	두	듀	드	디
ㄹ(R/L)	라	랴	러	려	로	료	루	류	르	리
ㅁ(M)	마	먜	머	며	모	묘	무	뮤	므	미
ㅂ(B)	바	뵤	버	벼	보	뵐	부	뷰	브	비
ㅅ(S)	사	샤	서	셔	소	쇼	수	슈	스	시
ㅇ Silent	아	야	어	여	오	요	우	유	으	이
ㅈ(J)	자	쟸	저	져	조	쥌	주	쥬	즈	지
ㅊ(CH)	차	챤	쳐	쳐	초	쵸	추	쥬	츠	치
ㅋ(K)	카	카	커	켜	코	코	쿠	큐	크	키
ㅌ(T)	타	타	터	터	토	토	투	튜	트	티
ㅍ(P)	파	파	퍼	퍼	포	포	푸	퓨	프	피
ㅎ(H)	하	하	허	허	호	호	후	휴	흐	히

©Modern Seoul Magazine (2012)

a sample from UHDR in Arabic

المادة 1

يولد جميع الناس أحرارًا متساوين في الكرامة والحقوق. وقد وهبوا عقلاً وضميرًا
وعليهم أن يعامل بعضهم بعضًا بروح الإخاء.

Arabic script, cont.

- used as the primary script for various language families because of the Quran
- an abjad for the Arabic language (letters only for consonants, vowels can be indicated by vowel marks)
 - fits well the inflection nature of Arabic (k-t-b)
- some modifications for other languages used as true alphabets with mandatory vowels, e.g. for Kashmiri or Persian (the Perso-Arabic script)
- no capital letters

Arabic script, cont.

ا	ب	ت	ث	ج	ح	خ
'a	b	t	ṯ	ǧ	ḥ	ḫ
د	ذ	ر	ز	س	ش	ص
d	ḏ	r	z	s	š	ṣ
ض	ط	ظ	ع	غ	ف	ق
ḏ	ṭ	ẓ	'	ǧ	f	q
ك	ل	م	ن	ه	و	ي
k	l	m	n	h	w	y

人人生而自由，在尊严和权利上一律平等。他们赋有理性和良心，并应以兄弟关系的精神相对待。

Chinese characters (Hanzi)

- traditional and simplified forms
- going back to 1400 BC (scripts on “oracle bones”)
- oldest writing system used continuously
- usually 1 character = 1 word or morpheme
- square-shaped, made of strokes
- over 100,000 characters

A global view on world's writing systems

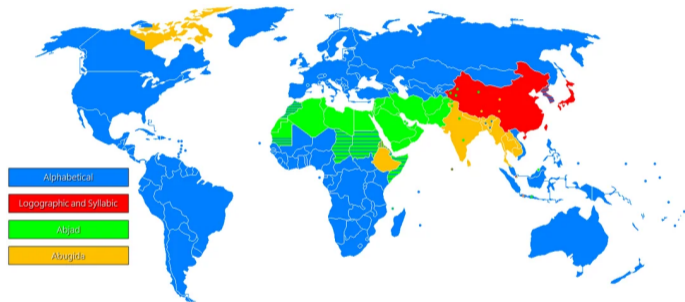
The world's most popular scripts (worldatlas.com)

Rank	Name of script	Type	Population actively using (in millions)
1	Latin Latin	Alphabet	over 4900
2	Chinese 汉字漢字	Logographic	1340
3	Arabic العربية	Abjad	660+
4	Devanagari देवनागरी	Abugida	608+
5	Bengali-Assamese বাংলা-অসমীয়া	Abugida	300
6	Cyrillic Кириллица	Alphabet	250
7	Kana かなカナ	Syllabary	120
8	Javanese	Abugida	80
9	Hangul 한글 조선글	Alphabet, featural	78.7
10	Telugu తెలుగు	Abugida	74
11	Tamil தமிழ்	Abugida	70
12	Gujarati ગુજરાતી	Abugida	48
13	Kannada ಕನ್ನಡ	Abugida	45
14	Burmese မြန်မာ	Abugida	39
15	Malayalam മലയാളം	Abugida	38

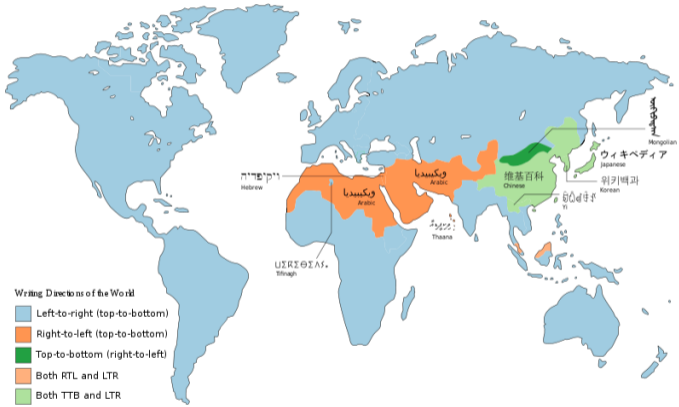
The world's most popular scripts (worldatlas.com), cont.

16	Thai ไทย	Abugida	38
17	Sundanese	Abugida	38
18	Gurmukhi ਗੁਰਮੁਕੀ	Abugida	22
19	Lao ລາວ	Abugida	22
20	Odia ଓଡ଼ିଆ	Abugida	21
21	Ge'ez ግዕዝ	Abugida	18
22	Sinhala සිංහල	Abugida	14.4
23	Hebrew עברית	Abjad	14
24	Armenian Հայոց	Alphabet	12
25	Khmer ខ្មែរ	Abugida	11.4
26	Greek Ελληνικό	Alphabet	11
27	Lontara	Abugida	7.6
28	Tibetan ལྷོ་	Abugida	5
29	Georgian ქართული	Alphabet	4.5
30	Modern Yi ꯀ	Syllabary	4
31	Mongolian ᠮᠣᠩᠭᠣᠯᠢ	Alphabet	2
32	Tifinagh	Abjad	1
33	Syriac	Abjad	0.4
34	Thaana	Abugida	0.35
35	Inuktitut ᐃᓄᐱᓂᓂᓂ	Abugida	0.035
36	Cherokee ႡᎦᎵᎾ	Syllabary	0.02

Type of the dominant writing system



Writing direction



Final remarks

Take-home message

- diversity, no writing system immediately intelligible
- writing systems invented independently a couple of times
- a scale from semiography to phonography (in reality always something in between)
- five basic types (vague boundaries): alphabets, abjads, abugidas, syllabaries, logo-syllabaries

- Relationships between a writing system and a language?
- Pros and cons of individual types of writing systems?
- Writing is not an object of linguistics (A paradox: linguistics tends to base the analysis on writing-induced concepts (phoneme, word) but subscribes to the principle of the primacy of speech)