

The World's Writing Systems: A XeLaTeX document showing typesetting in a variety of writing systems

Jake Warde

August 20, 2025

Sections 1-5 show examples of in different writing systems. The sections are organized according to recognized categories of writing systems. The XeTeX Companion (Section 3.1.3) contains an overview of types of writing systems.¹ Another recommended resource on writing systems is *Omniglot: The online encyclopedia of writing systems and languages*² Sections 6-7 are a summary of the technologies used to typeset text for all the world's languages. The document may also be a resource to someone learning to use XeLaTeX. XeLaTeX source is here: myprojectnotebook.com. No claims to being a XeLaTeX expert are made. In fact the document was created while learning to use XeLaTeX.

1 Abjads-Consonant Alphabets

Abjads, or consonant alphabets, represent consonants only, or consonants plus some vowels. Full vowel indication (vocalisation) can be added, usually by means of diacritics, but this is not usually done. Most of abjads, with the exception of Ugaritic, are written from right to left. There are three abjads currently in use: Arabic, Hebrew and Syriac. Abjads, or consonant alphabets, represent consonants only, or consonants plus some vowels. For more about abjads visit: <http://www.omniglot.com/writing/alphabetic.htm>

1.1 Arabic

عَلَى يَعْطَلُ وَأَنْ الْوَطَنَ شَانَ إِعْلَاءِ فِيهِ جُضِدٌ كُلُّ يَبْدُلُ وَأَنْ وَجِيرَانِهِ أَهْلِهِ وَمَعَ نَفْسِهِ مَعَ وَصَادِقًا أَمِينًا يَكُونُ أَنْ الْإِنْسَانَ عَلَى يَجِبُ
لِلتَّصْفِيَةِ مَثَالٌ وَهَذَا الْخَاصَّةِ الْمُنَوَّعَةِ عَلَى الْعَامَّةِ الْمُنَوَّعَةِ يَقْدَمُ بِأَنْ إِلَّا ذَلِكَ لَهُ يَتَمُّ وَلَنْ . لِلنَّاسِ السَّعَادَةَ يَجْلِبُ مَا

¹<http://xml.web.cern.ch/XML/lgc2/xetexmain.pdf>

²<http://www.omniglot.com/index.htm>

2 Alphabets

Alphabets, or phonemic alphabets, are sets of letters, usually arranged in a fixed order, each of which represents one or more phonemes, both consonants and vowels, in the language they are used to write. In some cases combinations of letters are used to represent single phonemes, as in the English sh, ch and th. For a history of alphabets, lists of alphabets currently in use, alphabets used to a limited extent, and alphabets that are no longer used visit: <http://www.omniglot.com/writing/alphabets.htm>.

Most specialists believe that the Phoenician alphabet, an abjad, was adopted for Greek during the early 8th century BC, perhaps in Euboea. The earliest known fragmentary Greek inscriptions date from this time, 770–750 BC⁴, and they match Phoenician letter forms of c. 800–750 BC.⁵ The innovation of the Greek writing system was that symbols were used to represent vowels. The introduction of vowels into the script (vocalization) ”and the abstraction of consonants as a process in speech – meant that the number of signs (or characters) actually used to represent spoken language could be reduced to a relative handful (around 30) because even in complex languages, such as some Inuit dialects, the number of actual consonantal and vocal sounds is finite. <http://people.ucalgary.ca/~mwolfe/Syllabary,alphabet.html>

2.1 Greek

μῆνιν ἄειδε θεὰ Πηληϊάδεω Ἀχιλῆος
οὐλομένην, ἣ μυρὶ Ἄχαιοῖς ἄλγε’ ἔθηκε,
πολλὰς δ’ ἰφθίμους ψυχὰς Ἄϊδι προΐαψεν
ἠρώων, αὐτοὺς δὲ ἐλώρια τεῦχε κύνεσσιν
οἰωνοῖσί τε πᾶσι, Διὸς δ’ ἐτελείετο βουλή,
ἐξ οὗ δὴ τὰ πρῶτα διαστήτην ἐρίσαντε
Ἀτρεΐδης τε ἄναξ ἀνδρῶν καὶ δῖος Ἀχιλλεύς.
τίς τ’ ἄρ σφωε θεῶν ἔριδι ζυνέηκε μάχεσθαι;

Note: The above text is the first 8 lines of the Illiad by Homer. See: <http://www.sacred-texts.com/cla/homer/greek/ili01.htm>

⁴The date of the earliest inscribed objects; A.W. Johnston, ”The alphabet”, in N. Stampolidis and V. Karageorghis, eds, Sea Routes from Sidon to Huelva: Interconnections in the Mediterranean 2003:263-76, summarizes the present scholarship on the dating

⁵Pierre Swiggers, Transmission of the Phoenician Script to the West, in Daniels and Bright, The World’s Writing Systems, 1996

2.2 Cyrillic Script Languages

В этой статье рассматриваются две конструкции современного русского языка: конструкция идентификации с существительным в именительном падеже (NomC) и сравнительная конструкция с существительным в творительном падеже (InsC). Эти две конструкции рассматриваются в условиях, когда они максимально сближаются: в контексте связки быть. Материал Национального корпуса русского языка показывает, что, вопреки встречающимся в литературе утверждениям, NomC все еще частотна в современном русском языке. Хотя процент InsC и возрастает, тем не менее распределение NomC и InsC до и после 1950 г. изменилось не очень сильно. В работе показывается, что распределение двух конструкций зависит от грамматической формы глагола. При употреблении связки настоящего времени используется исключительно NomC, в то время как при употреблении связки прошедшего и будущего времени превалирует конструкция InsC. Для разных именных групп распределение употреблений NomC и InsC различается. Процент NomC больше для тех именных групп, которые по своей семантике удовлетворяют требованиям NomC, а процент InsC больше у тех именных групп, которые по своей семантике удовлетворяют требованиям InsC.⁶

3 Syllabic Alphabets

Syllabic alphabets, alphasyllabaries or abugidas consist of symbols for consonants and vowels. The consonants each have an inherent vowel which can be changed to another vowel or muted by means of diacritics. Vowels can also be written with separate letters when they occur at the beginning of a word or on their own.

When two or more consonants occur together, special conjunct symbols are often used which add the essential parts of first letter or letters in the sequence to the final letter. Sample syllabic alphabets are shown at: <http://www.omniglot.com/writing/syllabic.htm>. Two syllabic alphabets are shown.

3.1 Devanāgarī

Devanagari, also called Nagari (Nāgarī, the name of its parent writing system, is an abugida alphabet of India and Nepal. Because of its use to write the Hindi language, devanagari is one of the most used and adopted writing systems in the world. The use of the name devanāgarī is relatively recent, and the older term nāgarī is still common. The rapid spread of the term devanāgarī may be related to the almost exclusive use of this script to publish sacred Sanskrit texts. This has led to such a close connection between

⁶Sample text from: Kuznetzova,J;Diachronic distribution of predicate nouns in Russian: Russian Linguistics 2013 <http://link.springer.com/article/10.1007/s11185-012-9107-1>

devanagari and Sanskrit that devanagari is now widely thought to be the Sanskrit script; however, before the colonial period there was no standard script for Sanskrit, which was written in whatever script was familiar to the local populace.⁷

संस्कृतम् जालपत्रे सुशवागतम्

संस्कृतभाषा भारतीय परम्परा तथा वृधेः आधारः । अस्माकं वृधेः, परम्परा, जीवनमुल्यवोध नमितिं प्रतदिनि जीवने संस्कृतभाषायाः आवश्यकम् अस्ति । वेदः, रामायणं, महाभारतं, श्रीमद्भगवद्गीता आदि धर्मग्रन्थाः संस्कृतभाषायां लिखितम् । इमे ग्रन्थाः नसिन्देह अन्यग्रन्थाणाम् अपेक्षा महतरं भवन्ति ।

संस्कृतभाषा दशसहस्र वर्षादपि पूरातनम् । सहस्रवर्षव्यापि इयं भाषा अस्माकं कथितभाषा आसीत् । इयं संस्कृतभाषा सर्वभाषाणां जननी अपि भवति । कथं इयं मुख्यभाषा तथा महत्वपूर्ण भवति सा स्वयम् कथयति । यदा अस्माकं भाषाज्ञानम् भवति तदा वयम् अस्माकं भाषायाः परम्परायाः ग्रन्थाणां च प्रशंसा तथा ज्ञान्तुं समर्थाः भविष्यामः ।

अतः आगच्छन्तु प्रतदिनि-जीवने वयं शुद्धं संस्कृतं कथयिष्यामः ।

Note: The web page from which the above text was copied supports the use of Sanskrit in daily life in the cultural life of India. The text is a welcome to Sanscritam.com <http://www.sanscritam.com>

3.2 Bengali

The Bengali alphabet is the writing system for the Bengali language. The script with variations is shared by Assamese and is basis for Meitei, Bishnupriya Manipuri, Kokborok, Garo and Mundari alphabets. All these languages are spoken in the eastern region of South Asia. Historically, the script has also been used to write the Sanskrit language in the same region. It is written from left to right and lacks distinct letter cases. It is recognizable by a distinctive horizontal line running along the tops of the letters that links them together, a property it shares with two other popular Indian scripts: Devanagari (used for Hindi, Marathi and Nepali) and Gurmukhi (used for Punjabi). The Bengali script is, however, less blocky and presents a more sinuous shape. Because of the large population of literate Bengali speakers, Bengali script is also one of the more widely used writing systems in the world.

ধারা ১: সমস্ত মানুষ স্বাধীনভাবে সমান মর্যাদা এবং অধিকার নিয়ে জন্মগ্রহণ করে। তাঁদের বনিক এবং বুদ্ধি আছে; সুতরাং সকলেরই একে অপরের প্রতি ভ্রাতৃত্বসুলভ মনোভাব নিয়ে আচরণ করা উচিত।

Note: The sample text in Bengali is from *WikiTravel: The Free Travel Guide* http://wikitravel.org/en/Bengali_prasebook. The text is Article 1 of the Universal Declaration of Human Rights (by the United Nations): Article 1: All human beings

⁷<http://en.wikipedia.org/wiki/Devanagari>

are born free and equal in dignity and rights. They are endowed with reason and conscience. Therefore, they should act towards one another in a spirit of brotherhood.

4 Syllabaries

A syllabary is a phonetic writing system consisting of symbols representing syllables. A syllable is often made up of a consonant plus a vowel or a single vowel. In a true syllabary there is no systematic graphic similarity between characters that share a common consonant or vowel sound. That is, the characters for 'ke', 'ka', and 'ko' have no similarity to indicate their common "k" sound (e.g. hiragana ケカコ).⁸

4.1 The Cherokee Syllabary

The Cherokee syllabary is a syllabary invented by Sequoyah, also known as George Gist, to write the Cherokee language in the late 1810s and early 1820s. His creation of the syllabary is particularly noteworthy in that he could not previously read any script. He first experimented with logograms, but his system later developed into a syllabary. In his system, each symbol represents a syllable rather than a single phoneme; the 85 (originally 86)[1] characters in the Cherokee syllabary provide a suitable method to write Cherokee. Some symbols do resemble the Latin, Greek and even the Cyrillic scripts' letters, but the sounds are completely different (for example, the sound /a/ is written with a letter that resembles Latin D). Each of the characters represents one syllable.

ᏍᏊᏃᏅᏗ ᏃᏘᏂᏚ ᏚᏃᏗᏂᏚᏚᏚ ᏚᏃᏍᏅᏗ ᏃᏅ ᏚᏍᏂᏚᏚᏚ ᏃᏂᏗ ᏃᏘᏂᏚ,
ᏚᏂᏚᏚᏚᏚᏚᏚ ᏘᏚ, ᏃᏅ ᏚᏚᏚᏚ ᏃᏚᏂᏚᏚᏚ ᏘᏚ, ᏚᏃ ᏘᏚᏃᏅᏗ, ᏂᏚᏚᏚᏚᏚ ᏚᏚᏚᏚᏚ,
ᏃᏃᏚ ᏚᏂᏚᏚ, ᏃᏂ, ᏂᏚᏚᏚ ᏚᏚᏚᏚ ᏃᏚᏚᏚᏚᏚ ᏚᏚᏚᏚ, ᏚᏚᏚᏚ ᏚᏚᏚᏚ ᏚᏚᏚᏚᏚ,
ᏚᏚᏚᏚᏚ ᏚᏚᏚᏚ, ᏃᏅ ᏚᏚᏚᏚ, ᏚᏚᏚᏚ ᏚᏚᏚᏚᏚ ᏚᏚᏚᏚ, ᏚᏚᏚᏚ ᏚᏚᏚᏚᏚ ᏚᏚᏚᏚᏚ,
ᏃᏅ ᏚᏚᏚᏚᏚᏚ ᏚᏚᏚᏚᏚᏚ ᏚᏚᏚᏚᏚ, ᏚᏚᏚᏚᏚ ᏚᏚᏚᏚᏚ ᏚᏚᏚᏚᏚ, ᏚᏚᏚᏚᏚ ᏚᏚᏚᏚᏚ,
ᏚᏚᏚᏚᏚᏚᏚᏚ, ᏃᏅ ᏚᏚᏚᏚᏚᏚᏚ ᏚᏚᏚᏚᏚᏚ ᏃᏅ ᏚᏚᏚᏚᏚᏚ ᏚᏚᏚᏚᏚᏚ, ᏚᏚᏚᏚᏚ
ᏚᏚᏚᏚ ᏚᏚᏚᏚ ᏚᏚᏚᏚᏚ ᏃᏅ ᏚᏚᏚᏚᏚᏚ ᏚᏚᏚᏚᏚᏚ ᏚᏚᏚᏚᏚ ᏚᏚᏚᏚᏚᏚ ᏚᏚᏚᏚᏚᏚ

5 Semanto-Phonetic Writing Systems

Semanto-phonetic writing systems often represent both sound and meaning. As a result, these scripts generally include a large number of symbols: anything from several hundred to tens of thousands. In fact there is no theoretical upper limit to the number of

⁸<http://en.wikipedia.org/wiki/Syllabary>

symbols in some scripts, such as Chinese. These scripts could also be called logophonic, morphophonemic, logographic or logosyllabic. Two semanto-phonetic scripts are currently in use, namely Chinese and Japanese (Nihongo). For more on semanto-phonetic writing systems visit: <http://www.omniglot.com/writing/semanto-ponetic.php>

5.1 Chinese

审讯揭秘伊以欧洲秘战

以向欧盟施压定性真主党恐怖组织

NICHOLAS KULISH 报道

本周在塞浦路斯一家小法庭里举行的一场少人问津的审讯中，一名24岁的真主党地下成员罕见地向人们揭示了以色列与伊朗在全球范围内进行的一场秘密战争真主党成员涉嫌炸以色列客车

Note: The text excerpt is from the New York Times Chinese Edition, February 23, 2013. Nicholas Kulish covers the upcoming elections in Cyprus.⁹

5.2 Japanese (Nihongo)

6 Technologies used to set the type in the worlds writing systems

6.1 Overview of Technical Issues in Supporting Scripts

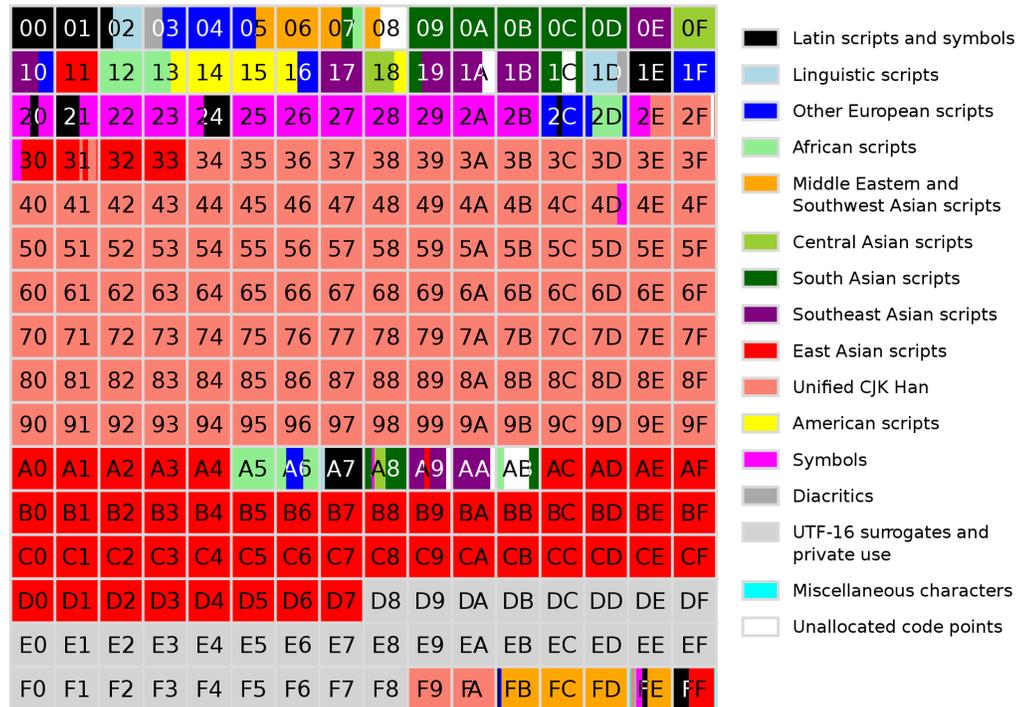
This section is based on a discussion of technical challenges in typesetting the worlds scripts presented by SIL International <http://scripts.sil.org>

”There are a number of challenges in the typesetting of non-Roman scripts. These include problems of interaction between the font and typesetting system, problems of very large character sets, and considerations of typesetting the particular script. Happily, the advent of new computer technology has considerably lessened some of these problems.”

⁹<http://cn.nytimes.com/>

6.2 Font Basics

6.3 Overview of Unicode Character Set



6.4 Codepages

More to be developed.