

**TUTORIAL**  
**8.16**

# Developing an orthography

This tutorial will give an introduction and some principles and tools for the development of an orthography (a writing system) in a previously unwritten language.

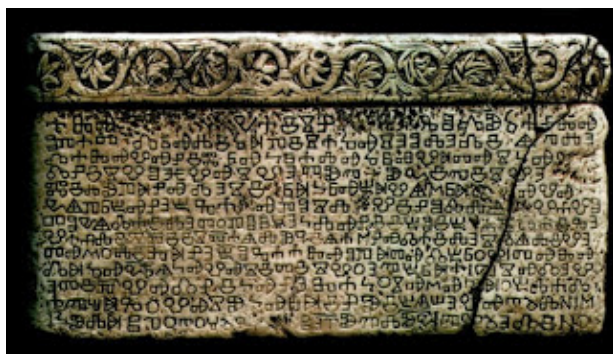
## Introduction

In this and the following two tutorials we will be looking at the process of developing a literacy project 'from the ground up', in a context where there is initially a very low level of literacy in the community, with no existing written language, and where all materials need to be developed so people can learn to read and write in their own language. This tutorial focuses on the development of an alphabet, and other aspects of developing a writing system in a previously unwritten language.

## What is an orthography?

Of the world's almost 7000 languages, less than half have an established writing system. Even though these languages have clear and complex rules of speech and established patterns of grammar, writing systems to record literature, or for people to communicate other than orally, have never been developed.

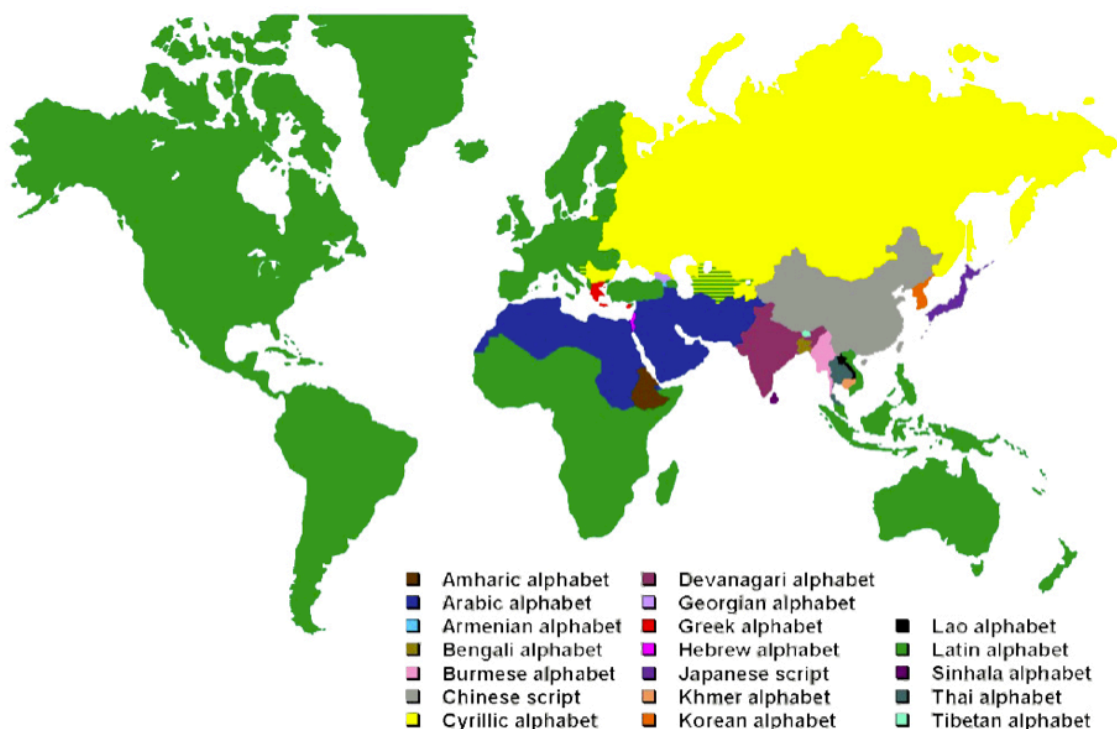
Writing is simply a method of representing language in a visual or tactile way, and consists of some kind of graphic marks on a durable or electronic surface. These marks must systematically relate to significant vocal sounds in the language so that readers can reconstruct a linguistic message and communication is achieved. The graphic marks used in writing are called 'graphemes', and they can be signs, characters, letters, diacritics, punctuation marks, etc.



Left: A stone tablet from around 1100 AD, written using the Glagolitic Alphabet, an early Croatian alphabet. It records a donation of land made to a Benedictine abbey.

An orthography is the combination of a set of graphemes (also called an alphabet) with the set of rules that governs their use in writing a particular language. This set of rules or conventions can include rules for spelling, word boundaries, punctuation, capitalization, hyphenation or emphasis. English is written using the Latin or Roman alphabet, which has 52 letters (including upper and lower case), plus punctuation marks. There is a set of rules for how the Latin alphabet should be used when writing English, for example, names must always begin with a capital letter, and questions must end with a question mark. Many other languages also use the Latin alphabet, and add various notations, diacritics or ligatures (two letters joined together) to represent necessary features of the particular language being written.

The major writing systems used in the world today are shown on the map below - the main ones are Latin, Cyrillic, Arabic and the Asian logographic scripts (like Chinese, where *units of meaning* are depicted rather than *sounds*). Each of these writing systems is used to represent many different languages. For example, the Cyrillic script that is used to write Russian is also used for around 100 other languages, the Chinese script is also used to write Japanese and other languages and dialects, and the Arabic script is used to write more than 40 separate languages.



Developing a writing system for an unwritten language is not simply a matter of analyzing the language and finding a linguistic solution. A writing system is always a significant social reality for a community of speakers, so sociolinguistic and political issues and the history and context of the

community need to be taken into account as well. There are also other practical issues such as font availability and how easy it will be for people to learn to read and write.

## Linguistic issues

A good writing system should seek to represent the language in an elegant and simple way. The first step is to do an analysis of the sounds of the language (this is called phonemic or phonological analysis - see Module 5, Tutorials 5.8 and 5.9 for an introduction). There are many good resources available to help you if you are involved in Phonological Analysis.

A phonological analysis will help you to decide:

- What are the distinctions of sound that must absolutely be represented in the language to avoid confusion? For example, [pit] and [bit] mean different things in English, and so it follows that /p/ and /b/ are different phonemes, and should be represented by different graphemes in the written language, otherwise meaning could not be communicated.
- How should the essential sounds be spelled? Should one sign represent one sound? Should each distinctive sound unit be represented by only one sign? For example, in English, /p/ in the initial position before a vowel is aspirated (i.e. is followed by a little puff of air, as in ‘pin’), but it is unaspirated after /s/ (as in ‘spin’). These two ‘p’ sounds are different, but they never make a difference in meaning in English and only change because of the contexts in which they occur. We say they are allophones of the same phoneme, so they can be represented by the same symbol or grapheme <p> in the written language.
- What is the grammatical structure of the language and how does this influence how the language is written? For example, decisions need to be made about where word breaks should be: in English, the word *walking* could possibly have been written as two separate words - *walk ing*.
- Does tone have to be represented, and are there any other units of meaning that must be written to convey a clear meaning? In many tonal languages there are a lot of words that would mean exactly the same thing if tone wasn’t represented in the written language. Chinese is a tonal language, but it uses characters that represent meaning in a logographic way rather than phonemically (based on sound). In English, intonation is represented by punctuation, for example:

*She’s bought a car.*      a statement of fact.  
*She’s bought a car?*      a question: is it true?  
*She’s bought a car!*      an exclamation: I don’t believe it!  
*She’s bought a car, ...*      an innuendo: ...and that’s not the end of it.  
*She’s bought a “car”.*      an implication: it may not be a “normal” car.

## Types of alphabets

Another major decision that has to be made is what *type* of alphabet will it be? For example an alphabet can be: a phonetic alphabet (where sounds are represented), a non-phonetic alphabet (like English), a syllable-based alphabet (meaning there is one character per syllable - Tibetan is of this type), or an abjad, (which is a consonant-only-alphabet, where all vowels are represented by diacritic marks - Hebrew and Arabic are examples of this type).

The decision about what type of writing system you will use will probably be influenced by socio-linguistic and practical issues, which we will look at now.

## Sociolinguistic issues

Sociolinguistic and other political factors have to work together with linguistic considerations when developing an orthography. The orthography really belongs to the community: to the speakers and the eventual readers and writers of the language. It is crucial that local people are enthusiastic and accept the orthography so they feel ownership of it and will be motivated to learn to use it. Development, testing and making adjustments to an orthography should all involve consultation and discussion with local people - we will look in a later tutorial at how this kind of collaboration can be a natural part of a community literacy project.

During the development stages of an orthography, local people will be your main resource in understanding socio-linguistic and political factors that need to be considered, such as:

- *Attitudes toward other languages or dialects in relation to their language.* Do they feel their language is inferior or superior to other languages or dialects and, if so, why? Is there pressure on their language from another major language? Is there another language that has a high status or is of commercial or educational benefit?
- *The influence of other orthographies.* What are the other orthographies that people have come into contact with in their region (those used for the national language or other major languages)? How many people are literate in other orthographies? How do people feel about using an existing familiar orthography or would they prefer a new and unique orthography for their own language?
- *Government policies.* What are the government policies relating to literacy, education and alphabets or scripts? Are there any existing government policies or projects supporting particular orthographies?

## Practical issues

Finally, there are a number of more practical issues to think about when developing an orthography, such as:

- *The utility, or practical simplicity of the orthography.* The orthography should naturally reflect the particular structure of the language in a way that makes sense to speakers of that language when they read it. A phonetic orthography, for example - where the written form represents the sound - will be easier for people to learn to read and write and will also allow younger members of the community to add new words - to spell them and write them phonetically, which helps the written language keep pace with natural language change.
- *Technical production issues.* What local fonts are most available for printing, word processing or other digital communication in the country? The representation of any characters other than those used in the major writing systems is going to be a problem because access to special fonts may not be available to the community. The safest option - to ensure usability of the orthography as well as safe long-term storage of digital files containing text written in that orthography - is to use only characters that can be found on a standard local keyboard or combinations of these characters (e.g. digraphs or combinations of letters with diacritics).



## ACTIVITIES

### *Developing an Orthography*

1. For detailed practical information and resources for developing an orthography, you can go to the *Resources for Developing Orthographies* link at the SIL (Summer Institute of Linguistics) website
2. You will find a helpful step-by-step guide in *Writing unwritten languages, Drafted by Clinton Robinson with Karl Gadelii, December 2003.*
3. For a wealth of information for further research, you can go to *Omniglot*, an online encyclopedia of writing systems and languages, or to the *Museum of the Alphabet*.
4. After you read the information below on *Colorbet*, (from omniglot.com) answer these questions:
  - What type of alphabet is *Colorbet*?
  - If *Colorbet* was chosen as the writing system for a previously unwritten language, are there any potential issues can you think of that might come up in linguistic,

socio-linguistic, political or practical areas? List any you can think of.

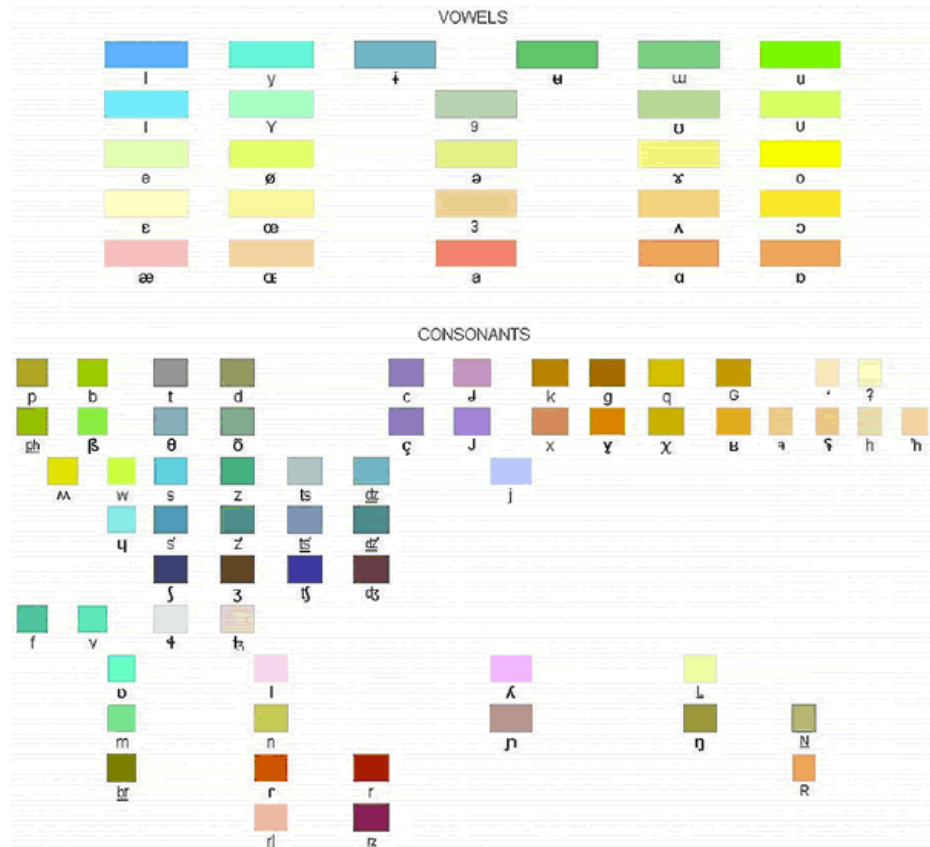
**Colorbet:**

Colorbet is a colour alphabet based on the International Phonetic Alphabet, and was developed by Vitaly Vetash, a Russian painter and linguist. The first variant of the colour alphabet was done in 1983, on the basis of psycholinguistic investigations. Theoretically, any language can be written using Colorbet.

The model is based on the kinship of psychological meaning of sounds and colours. According to the system of Goete (which now is used in T.V. transmission), the whole variety of colors is based on a fusion of three main rays (red, yellow and blue), light and shadow. A similar law can be found in phonetics, where the variety of vowel comes from the combination of the triangle of the main sounds (A, I, U). Therefore, the colours of vowels are: A - red, I - blue, U - green (not yellow, but green, as used on TV).

According to psycholinguistic investigations, one can give to non-sonorous consonants darker colours than to sonorous ones, voiced ones have brighter colours than voiceless ones, and fricatives are more coloured than plosives. That is, brightness depends on sonority: from rich colours (of sonants) to dry tints (of voiceless plosive sounds). Sonorous consonants, having more clear colours, are more close to vowels (R is ruby-coloured, and burr R is orange; velar L is yellow-white and palatal L is white-rosy. Nasal vowels: mat-green M, mat-beige N, light-violet Y, dark-yellow W).

On the next page the colour scheme according to the IPA (International Phonetic Alphabet) is represented. It shows that tints of kindred sounds are close. However, plain (flat) reproduction of sounds doesn't reflect the character of sounds exactly, because in the strict way one has to reproduce the texture also. Thus, colours of vowels must have additional luminescence, and consonants produce material sensations. For example, sibilants (S, Z, TS) could be reflected by metallic surface, and gutturals coincide with shaggy (wooden or board) surface. A dim, lacquered surface fits for resonant sounds (L,M,N etc.). Vowel sounds are written as longer coloured sections, to distinguish them more easily from consonants, which are shorter.



The example below of *Colorbet* writing says, “colour alphabet”:

