

Typography and Education

<http://www.typoday.in>

Hello World

A Tool To Compare & Explore Multiple Writing-Scripts – A Work In Progress

Amrita Gada, Sophia Polytechnic, amy.gada@gmail.com

Abstract: This paper documents and discusses the development process of a model designed to explore writing-scripts from an educational point of view. Awareness about the variety of writing scripts is very less among people. While learning scripts via conventional methods, the learning curve is steep, time consuming causing an entry barrier. The study aims to find a solution to resolve that problem. The primary objective of the study was to showcase writing-scripts as a collective to give an overview of writing systems, and to simplify the process of learning scripts by focusing the study around familiarising the user with the writing-scripts.

Key words: *Writing-Scripts, Educational Tool, Color-coding, Transliteration.*

1. Introduction

There are more than 35 distinct scripts and 7000 languages are in use today, and the most popular being the Latin alphabet and English language. Of these, one is only aware of the scripts that are renowned like Latin, Hebrew or Kanji (commonly referred to as Japanese script). Not everyone is aware of the entire gamut of scripts that exist. As the prevalence of signages across the world is increasing, from signages inside malls to the ones across digital applications, multilingualism is increasingly being visible. Increasing number of fonts also are being designed for 'non-Latin' scripts. Information about writing systems, scripts, languages is available via books or videos or tutorials/lessons which are knowledge intensive. a lighter, visual based solution can fill in much gaps. Organising the data about scripts can sort them based on similarities and differences. Using tools like color coding

and interactive interfaces, accessing this data can be made a lot more systematic and aligned. A visual based solution can bridge the gap that is present at the moment. Information visualization offers a way to ease cognitive load to the human perceptual system through graphics and enables us to recognise meaningful patterns in data. We aim to breakdown this process/scripts and look at it with fresh eyes. To find out what scripts are made of and their connecting threads. An analogy that can be applied here is of the idea of world map for a view of the world.

2. Basic Terminology

Below is some basic terminology used in the paper:

The generic **term** text refers to an instance of written material. The act of composing and recording a text may be referred to as **writing**, and the act of viewing and interpreting the text as **reading**. A set of defined base elements or symbols, are individually termed as **characters** and collectively called a **script**. The method and rules of observed writing structure (literal meaning, “correct writing”), and particularly for alphabetic systems, includes the concept of spelling is called **orthography** which is understood and shared by a community, which assigns meaning to the base elements (**graphemes**). Their ordering and relations to one another is called a **language** (generally spoken) whose constructions are represented and can be recalled by the interpretation of these elements and rules.

Characters of a script represent speech sounds or represent a word, morpheme, or other semantic units or both in some cases. Following are the categories of scripts:

Logo Syllabograms (or morpho syllabograms) are units that function in written language to spell whole words or morphemes but which also correspond to discrete syllables in spoken language if, in the language in question, words are typically monosyllabic as is the case in Chinese. A **syllabogram** is a unit of writing that corresponds to a discrete syllable in speech and which is used for spelling any words whose spoken equivalents contain that syllable regardless of meaning. In an alphabet there are autonomous characters which can be put into correspondence with vowels as well as consonants. In a **featural** system ‘the shapes of the characters correlate with distinctive features of the segments of the language. The characters of an **abjad**, or consonantary, correspond only to consonants in spoken language while those of an **abugida** correspond to a consonant- plus- vowel sequence. Vowels in abugidas correspond to systematic additions to a base consonant character which on its own often represents a consonant plus /a/ as a kind of default vowel - an abugida is thus a vocally augmented abjad. Note that an abjad can, as in

Arabic, have optional diacritics corresponding to vowels whereas the vocalic augmentation in abugidas is obligatory.¹ An **abugida** is similar to an alphabet: all vowels are indicated, but normally, vowels are written as diacritics, and one vowel is not written. The Indian tradition considers /a/ to be inherent in the consonant symbol. Vowels, other than short /a/, are shown by diacritics. Consonant clusters are written as **ligatures**. In an abugida, the vowels are written as diacritics on the consonants and one vowel is not explicitly written. An abugida, also called an alphasyllabary, is a segmental writing system in which consonant-vowel sequences are written as a unit: each unit is based on a consonant letter, and vowel notation is secondary. Abugida scripts: Devanagari, Gujarati, Gurmukhi, Bengali, Kannada, Malayalam, Oriya, Sinhala, Tamil, Telugu. In an Abugida, vowel graphemes have two allographs - one free, and the other bound. At the beginning of an orthographic unit, the free allograph is used; otherwise the bound allograph is used, i.e., the vowel is written as a diacritic on the preceding consonant. A consonant cluster is written as a ligature, with the individual graphemes of the cluster combined into a single symbol. A ligature is treated as a single unit with respect to vowels. The vowel diacritics are attached to this conjunct symbol exactly as to a single consonant symbol.²

3. Script learning programs

There are many routes a person can take to learn languages. One can use books, online tutorials, professional tutors, online videos, take language classes etc. Video tutorial - Youtube, Omniglot³, UCLA Center for World Languages⁴ and its Language Materials Project⁵

4. Learning Strategy

4.1 Translation

Translation is a learning strategy to study languages that has been employed traditionally. The **conversion of scripts** or writing is a procedure of replacing text written in one script or writing system with the characters of another script or system in order to make the text (e.g., proper names) legible for users of another language or script.

Transliteration: Graphemic conversion, i.e., from one script to another.

¹ Daniels 1996: 4

² Writing Systems, Henry Rogers

³ <http://www.omniglot.com>

⁴ <http://www.international.ucla.edu>

⁵ <http://lmp.ucla.edu/>

Transcription: Phonemic conversion, i.e., from one writing system into another, possibly including script conversion.

4.2 Transliteration to learn script

Most systems of transliteration map the letters of the source script to letters pronounced similarly in the target script, for some specific pair of source and target language.

Transliteration is opposed to transcription, which maps the sounds of one language into a writing system. If the relations between letters and sounds are similar in both languages, a transliteration may be very close to a transcription. In practice, there are some mixed transliteration/transcription systems that transliterate a part of the original script and transcribe the rest. For instance, a Latin transliteration of the Greek phrase "Ελληνική Δημοκρατία", usually translated as 'Hellenic Republic', is "Ellēniké Dēmokratía".

Google Translate is a great tool that translates, transliterates across 90 languages at various levels. Texts written in the Greek, Devanagari, Cyrillic and Arabic scripts can be transliterated automatically from phonetic equivalents written in the Latin alphabet. Google Translate does not translate from one language to another ($L1 \rightarrow L2$). Instead, it often translates first to English and then to the target language ($L1 \rightarrow EN \rightarrow L2$). However, because English, like all human languages, is ambiguous and depends on context, this can cause translation errors. Google Translate, like other automatic translation tools, has its limitations.

5. Visual design – a Probable Solution

Visual representation gained popularity with greater acceptance of empirical approaches. A visual solution was sought for the autonomy that it provides the user with. The user can view it at his own pace and form concepts and comparisons on his own. Writing and reading are both visual interactions.

People are aware of the sounds. But not of how it is visually represented by specific characters in various scripts. The key problem lies here, that this information is not visually represented anywhere in the script. We attempt to solve this by color coding the characters and using a base script as the reference key to the color code. Also by breaking down the orthography and labelling it with the corresponding matching characters of the base script.

The concepts we wish to highlight are pluralism of the languages, and the virtue of use of design solutions & colors.

5.1 Color Coding

In case of confusingly similar things, Color coding them has two aspects: assigning the same color to things, groups them together. Assigning multiple colors help tell things apart. Color can add dimensionality and richness to communications. Color is typically used to differentiate information into classes. The challenge is picking colors that are discriminable. A systematic approach to choosing colors can help us find a lineup effective for color coding. Occasionally, authors use a sequence of colors, such as the 'rainbow' color scheme, to represent a range of values.

6. A Tool To Compare & Explore Multiple Writing-Scripts

A visual solution, as stated earlier, can be a great solution to bridge the gap of a missing platform for scripts. The respective scripts can be displayed collectively, the orthography be defined in a simplistic manner by color coding and labelling it to a common base - the phonetics. An interactive tool was developed where the user can type in the letters and the tool transliterates and presents the results in other scripts. The output of the process is colour coded for enhancing understanding of scripts for the user.

6.1 Scope and Limitations

This project was focused to develop simplistic tools to showcase the range of scripts used today and briefly explain their orthography, using design. The tools developed are prototypes and work to a limited point.

Phonetics/phonemes was identified as the common base among all scripts [except logographic scripts like chinese which are not based on sound, but on meaning, i.e. a character stands for a word]. in other scripts, each character stand for a sound.

6.1.2 These guidelines were established to streamline the development process

1. The writing-systems should be simplified preserving the qualities (add)
2. Preserve the uniqueness of all scripts
3. The Orthography Chart should be easy to follow and this should be evaluated on the basis of following aspects:s
 - 3.1 Consistency and accuracy of color coding
 - 3.2 Accuracy of orthography breakdown

4. For the purpose of study, use Noto family of fonts
5. For the prototype, use the phonetic base of Latin/English
6. Design the model for web first

6.1.3 Choices of Script that is shown in Flowchart

Major scripts used across the world, by maximum number of people were selected. from Latin with more than over 4900 million users to Cherokee to about 0.02 million users. Of these, ones which could be broken down into parts and transliterated, as per our limited knowledge, have been included in the chart.

Latin - more than over 4900 million users across the world, has been used as the base language for transliteration as it the most widely used script in the world.

6.2 Research methodology

Major scripts being used today were determined. The alphabet sets of these were collected and displayed.

A test word, hello world, was selected and transliterated across 15 scripts. The model was developed following a combination of following three methods.

6.2.1 Defining the orthography

6.2.2 Phonetic matching

6.2.3 Color coding

6.2.1. Defining the orthography

Base consonant was identified and marked. The diacritics - superscript, subscript, prescript, postscript were similarly colored. This was verified with list of diacritics. They were placed at a little distance from each other to be able to gauge where they attach, upon writing. '+' sign represents the attachment/detachment of diacritic

A test word, hello world in this case is used to represent how scripts are written with color code

[Representing abugidas is difficult because there is non linear placement of diacritics. diacritic conversion requires expert knowledge about matching characters and accurate diacritic conversion requires a suitably sophisticated code in the background

Abjad representation in interactive tool is difficult because middle ending beginning forms need to be coded to present in the output.]

6.2.2. Phonetic matching

Matching characters - one is to one / one is to many: a test word was transliterated in all scripts where the characters used have the same phonetic connotation as the Latin set, only difference being they look different because the other script looks different.

Interactive tool - background coding: all the characters were matched one is to one, and remaining ones, which did not have a corresponding sound in Latin, were left out.

6.2.3. Color coding

The universal concept of sounds are represented by universal colors which transcend all boundaries of countries and dialects and thus can be applied to all scripts and understood by all people.

Each phoneme is assigned a color in the test word and applied across the corresponding character in other scripts. By typing any words the tool activates and change the value in color in the respective script and can help in reading the pattern,

The position, indentation, and line length of the row tracks the text of various script.

Color Code gives each row the same visual prominence, making it easier track the color in each script that can help in avoiding the ambiguity that occur with plain black text. It is easy to compare different parts of the script. The view gives different insight into the data.

6.3 A model consisting of three parts was developed

6.3.1. Alphabet Set

6.3.2. Orthography Chart

6.3.3. Interactive Tool

6.3.1. Alphabet Set

Scripts in Alphabet Set - 35 nos.: Arabic, Hebrew, Aramaic, Tifinagh, Burmese, Ge'ez, Inuktitut, Javanese, Khmer, Lao, Sudanese, Thaana, Thai, Tibetan, Bengali/Assamese, Devanagari, Gujarati, Gurmukhi, Kannada, Malayalam, Oriya, Sinhala, Tamil, Telugu, Georgian, Mongolian, Hangul, Armenian, Cyrillic, Greek, Latin, Chinese, Cherokee, Kana, Modern Yi

6.3.2. Orthography Chart

Scripts in Chart - 16 nos.: Arabic, Hebrew, Bengali/Assamese, Devanagari, Gujarati, Gurmukhi, Kannada, Malayalam, Oriya, Sinhala, Tamil, Telugu, Armenian, Cyrillic, Greek, Latin

hello world हेल्लो वर्ल्ड

Figure.1



Figure.2



Figure.3



Figure.4



Figure.5



Figure.6

6.3.3. Web Tool

Scripts in Interactive Tool - 5 nos.: Latin, Cyrillic, Greek, Armenian, Georgian

Web is used the medium because it has broader reach, better interaction, easier to update, economical. The site was built using Bootstrap. Javascript is used to write the logic. For loops are used to iterate through each letter and create the corresponding word in different scripts. Letter dictionaries are maintained for each script. A color dictionary is maintained. Each letter has a corresponding color.

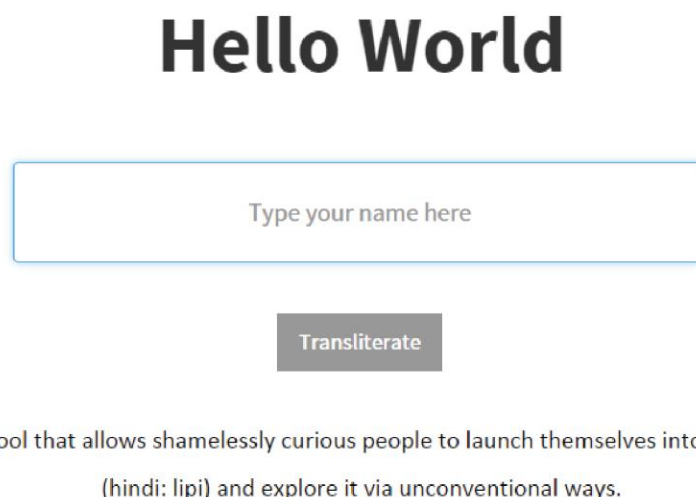


Figure.8

English : hello world
Cyrillic : хелло world
Greek : -ελλο world
Armenian : հելլո world
Georgian : ჰელო world

Figure.9

The tool makes it easier for the people to relate to the idea of multiple scripts and comparison based learning.

Encoding and Data Entry Mechanism: Logic: The user enters input in latin script. Letter by letter transliteration of characters takes place for all the specified scripts. Similarly letterwise color coding takes place. And the result is output to the page.

6.4 Accuracy and Functionality:

The model has been executed in a non expert environment. So there are various flaws and limitations of many kind, resulting in a prototype that should only be seen as a concept. It works to a certain level, under limited set of phonemes.

7. Use Cases

The tool helps in teaching scripts to curious people, by identifying the similarities and differences between scripts. It is a great learning platform for natives of a certain language, to learn a different script using the tool, for example a Latin user can learn Armenian. Designers as well can benefit from this medium, like graphic designers or illustrators can get a variety of scripts to use, with correct information and understanding. On the other hand, font designers can go through various scripts and understand the basic technicality involved in the placements, sounds, strokes and so on.

8. Observation [20%, 600 Words]

Every script is a bunch of characters written in a certain way, based on the set of universal sounds that the user already knows. This type of simplification brings out the central uniqueness of characters and enables the learner to immediately grasp it.

Translation is not regarded as a direct contributor in the process of acquisition of a new language or script.

9. Limitations

Not all scripts can be perfectly transliterated across due to defective orthography. Some difficulties were faced with IPA being a large and complex set, and not easy to transliterate. Lack of expertise about linguistics, lack of time and resources were faced as well.

Limitations with Design:

Limited distinguishable color palette - limited contrasting colors. As the number increases, contrast decreases, which gives us lesser options to work with. Another problem is the unavailability of free fonts for certain scripts and how browsers do not render certain non popular scripts.

Solution: Sometimes lesser options gives us the ability to come up with simpler designs. Grouping of sounds to reduce variables and effectively use limited colours can help us solve the issue at hand.

10. Possibilities of Future Work

The model uses a limited phonetic base currently. This base can be expanded to include all the sounds included in the IPA - The International Phonetic Alphabet⁶. Also, more research will be needed to define the phonetic and linguistic rules of each script in detail, to be able to add more scripts to the current model. Non-Latin scripts have less documentation about the linguistic and phonetic rules that specify when and how a character should be used to represent a certain sound. Thus ambiguity arises and thorough study is required to decode the information. Also, addition of scripts to the interactive tool and orthographic chart to complete the set. Also, striving for better matching of characters among scripts for transliteration is required.

11. Summary

Learning scripts can be made simpler by applying design principles, such as color coding, labelling the orthography, and by analogies and contrast. The model essentially introduces the user to all the major existing scripts, in an extremely simplified and brief manner. The flowchart in the model shows the structure of the scripts in brief. Simplification of the scripts contributes a lot towards achieving and retaining the interest of the user. By placing all script characters on a common phonetic base, confusion is removed and a

⁶ The IPA is an alphabetic system of phonetic notation based primarily on the Latin alphabet. It was devised by the International Phonetic Association as a standard representation of the sounds of oral language.

degree of familiarity for the learner gets introduced. This can contribute to greater chances of the user actually completing the learning process and not leaving midway through the lessons, because of having run out of enthusiasm and vigour. It makes the whole learning process interesting and visually-rewarding. Transliteration tool lets the user enter a word his own choice to explore scripts. This gives greater control to the user over the learning process, and he can customise it to suit his pace and needs. Color coding the script characters, based on phonetics, again removes confusion and the user can solely concentrate on the structure of the letters. The concept of learning by comparison has two implications. One, it takes away the necessity of remembering and recalling the shape and sound of the letters, of the script being studied. This lessens the burden the user may feel. Two, it encourages exploration and formation of concepts. The user is encouraged to make connection between corresponding letters of the neighbouring script and play around with them. The tool can be used to test the theories they come up with.

12. Conclusions

Script learning can be an interesting journey, often leading to tangential concepts of culture and origins. Geographically closer scripts have a lot of similarities and often common origins as well. The idea lands us to a question we must ask ourselves : can scripts be taught like this? This might look like an unconventional idea to some, but considering the advanced stages that the human race is in at present times, anything can happen. Design has evolved with the changing times, this era lives in the web world, where information is available at the tap of the thumb. Learning through online courses, apps or tools is not so unconventional nowadays. Looking at scripts like an old fashioned forgotten world might be a limited imagery that people are carrying. History only brings us closer to the future, which is why we need to value something that had emerged out of the great past and built it well in the present. This tool has great advantages for the learners, as it skips on non relevant rules which otherwise increases the initial learning material, the efforts required from the learner and retaining the initial unreadability of the script(s). By visually representing the sound of each character we attempt to add an additional layer of information which may help to decode the script. We attempt to reduce the entry barriers and increases productivity with a feeling of fulfilment from the users, which is necessary in the initial stages of learning. We intend to present this tool as a unique creation for independent learning and understanding the complexities involved in understanding script.

Acknowledgements

Prof Satyamangal Rege

Prof Meenal Joshi

Dr. Girish Dalvi

Anurag Chowdhary

Piyush Verma

Fonts Used

Noto Family, <https://www.google.com/get/noto>

Ek Mukta Devanagari <https://github.com/girish-dalvi/Ek-Mukta>

References

Rogers, Henry. Writing Systems: A Linguistic Approach. Malden, MA: Blackwell Pub., 2005. Print.

Daniels, Peter T., and William Bright. The World's Writing Systems. New York: Oxford UP, 1996. Print.

"Point of View, Color Coding." <http://www.readcube.com/articles/10.1038%2Fnmeth0810-573>.

Nature Methods, Vol.7 No.8, 2010 , Aug. 2010.

"Omniglot - the Online Encyclopedia of Writing Systems and Languages." Omniglot - the Online Encyclopedia of Writing Systems and Languages. Web. 08 Feb. 2016. <<http://omniglot.com/>>.

Harper, Douglas. "Orthography." Online Etymology Dictionary. ©2001-2016 Douglas Harper, Web. 8 Feb. <<http://www.etymonline.com/index.php?term=orthography>>.

"Google Translate." Google Translate. Web. 08 Feb. 2016. <<https://translate.google.com/>>.