Guidelines on usage of the terms

December 4, 2024

Vyshantha Simha (Institute of Digital Humanities, University of Cologne, Germany)

This document is intended to give guidance on how the terms *script*, *font* and *writing system* are interwoven yet clearly distinct based on contemporary knowledge.

The sets under consideration are intentionally human created graphs or shapes (conditionally graphemes or glyphs) on a medium and could be one or more of the following types: glotto-graph, or semasio-graph, or pasi-graph or kineto-graph or an unclassified.

Glotto-graph : graphs used to encode linguistic units
Semasio-graph : graphs used to encode non-linguistic units, like symbols
Pasi-graph : graphs is a type of semasio-graph used to encode abstract concepts
Kineto-graph : graphs is a type of semasio-graph used to encode movement

Writing System

A **writing system**[1] is a set of visible or tactile signs used to represent units of language in a systematic way, with the purpose of recording or communicating ideas.

A system that allows capturing of linguistic (glottography) and non-linguistic information (like numerals, non-phonetic suprasegmental like punctuations, semasio-graph or kineto-graph or pasi-graph, etc.) for a specific language (either spoken or signed) is a writing system.

A writing system could include more than one type of script (multi-scriptality) not limited to heterograms to encode part of the lexicon of a language. The writing system for languages Japanese and Old Korean required multiple scripts for encoding information.

A language might employ multiple writing systems (polygraphy) for encoding information. The languages Bosnian, Inuktitut, Kashmiri, Konkani, Mongolian, Serbian, Uzbek are a few examples which employ multiple writing systems in different contexts and / or niches.

Bhaisuki, or Garay, or Arabic, or mediæval Latin documents by Cisterians as a writing system has text required to be read or written in a particular direction, while the numerals modify this direction.

Pinyin had been designed to be a writing system and is a font-variant of the Latin script (inclusive of diacritics).

Hebrew, Devanagari, Thai, Tibetan, Grantha scripts with cantillation marks are also writing systems. Notation systems (like mathematical, musical, pasigraphy) are writing systems by themselves without being scripts.

All scripts (including font variety) are part of writing systems, however not all writing systems can be scripts. Maps are not writing systems and only visual information systems or graphical communication tools.

An idealistic system that allows encoding all probable and plausible thoughts in a particular speech (inclusive of verbal & non-verbal parts of language) is a complete writing system.

Script

A **script**[1] is a set of graphic signs used to represent units of language in a systematic way. It serves as the visual representation of a writing system, encoding linguistic units such as sounds, syllables, words, or meanings into readable symbols.

A set or collection of graphs that adheres to a structured format and enables capturing linguistic units (be it spoken or signed) is a script (or shaped markings). The encoding information in the graphs set could manifest as tachygraph, or phonograph, or logograph (or morphograph), or kineto-graph or cipher, or undeciphered typeface or combination of the above. A script could have multiple orthographies and within a particular block or cluster of graphs the direction could be different to the general direction of text (reading or writing), thus it can be non-sequential in nature.

Given a set or collection of graphs isn't deciphered to be encoding language or isn't known to encode language owning to obsolescence of that graph set, then if it can be proved beyond reasonable doubt through pattern recognition, computational and digital humanities methods that the graph set indeed encodes a form or part of a language then such a set or collection of graphs should be considered as script as well. One should look for evidence of syntactic structure (including graphotactics) namely compositionality and recursion to validate if a set or collection of graphs is likely to encode language.

A set of graphs need not be developed over several years or centuries organically and they could be designed or co-created by field or domain experts in a short period of time and such neographies should be subject to the Litmus Test for conformance.

For extinct languages if the graph set has attestation by means of aforementioned techniques or using multi-scriptal or multi-lingual texts decipherment, then that graph set should be subject to the Litmus Test for conformance.

A set of graphs can certainly encode a constructed or auxiliary or artificial or artistic or fictional or cryptophasia or liturgical language or languages en route to revival, then that graph set should be subject to the Litmus Test for conformance.

A script could be used to express many languages and that script may not encode complete phonology of a language. Pasigraphy isn’t a script rather a notation system.

The Shavian script encodes the complete English phonology. The Adinkra symbols is a script used to encode the Akan languages.

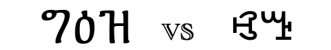
Font

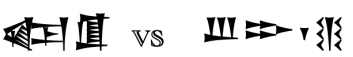
A **font**[2] is the complete set of characters in a given style, weight, and size of a typeface. It is the practical manifestation of a typeface, containing all the letters, numerals, punctuation, and other symbols required for written communication.

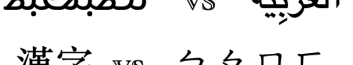
Litmus Test to determine a set or collection of graphs is a font variant (including calligraphic types) of an existing script or a new script

The sets being compared should be not very dissimilar and intentionally not bound by space and time, in other words comparison shouldn't be made between Devanagari vs. Katakana, or Devanagari vs. Ge'ez, or Katakana vs. Vai, rather Devanagari vs. Nandinagari, or Katakana vs. Man'yōgana.

देवनागरी vs カタカナ







漢字 vs ㄅㄆㄇㄈ

The set of graphs being compared should not be a composition of part aesthetically appearance such as Blackletter (Fraktur) and part aesthetically appearing like Times New Roman. Given if two sets have a similar number of graphs, encode the same information and only vary with any of these weight, size, serifs or style properties i.e. essentially visually, then the set is unequivocally a font variant of the existing script.

SERIFS vs Sans serifs

Arial vs Arial

Calibri vs Calibri

Helvetica vs Helvetica

Impact vs *Monaco*

Here are the proposed parameters to determine if a given graph set (that conforms to being a script) is a font variety of an existing script or a new script by itself.

The weightage for criteria 7, 8, 9 & 10 all passing is more than criteria 1 to 6 all passing.

1. Does this new set of graphs have aesthetically coherent distinct shapes that represent linguistic units as opposed to the existing graph set ?
   Yes (**PASS**) / No (FAIL)
2. Does this new set of graphs require a different or modified set of orthographic rules in order to represent the linguistic units, namely the set of graphs are linked or joined or spaced or ordered differently as opposed to the existing graph set ?
   Yes (**PASS**) / No (FAIL)
3. Does this new set of graphs use a new way of encoding the linguistic units, namely the existing graph set uses a segmental form whilst the newer uses a syllabic or even pictographic ?
   Yes (**PASS**) / No (FAIL)
4. Does this new set of graphs require different directionality or text layout rules as opposed to the existing graph set ?
   Yes (**PASS**) / No (FAIL)
5. Does this new set of graphs inculcate distinct cultural, or visual and / or structural identity (like tactility) as opposed to the existing graph set ?
   Yes (**PASS**) / No (FAIL)
6. Does this new set of graphs address needs of a language or group of languages that hasn't been dealt with through the existing graph set ?
   Yes (**PASS**) / No (FAIL)
7. Has there been a multitude of forms of a graph which encodes the same information owing to separation by space and time ?
   No (**PASS**) / Yes (FAIL)
8. Does this new set of graphs address the contextual usage which the existing graph set did not meet ?
   Yes (**PASS**) / No (FAIL)
9. Has there been a compendium produced (like inscriptions or books or teaching material) for the new set of graphs ?
   Yes (**PASS**) / No (FAIL)
10. Are members of a community using or have used it for their required niche purpose during their contemporary times ?
    Yes (**PASS**) / No (FAIL)

To elaborate the functionality of the Litmus test here are a few demonstrations. Consider the graph set for standardised Brahmi (Ashokan era) and if one intends to compare the graph set for the *unified* graphs (for eastern, western, northern and central Asian variants) from the Gupta dynasty period. Following is the table of comparison of the graph sets.

Criteria - Evaluation	Result	Criteria - Evaluation	Result
1 : Yes	PASS	6 : No	FAIL
2 : Yes	PASS	7 : Yes	FAIL
3 : No	FAIL	8 : Yes	PASS
4 : No	FAIL	9 : Yes	PASS
5 : No	FAIL	10 : Yes	PASS

If the one would consider only the graphs for the central Asian *variant* and perform the Litmus test against the standardised Brahmi (Ashokan era) graph set, then among the above table entries the criteria 7 would *pass* and if that *variant* alone was used for several other languages then criteria 6 would *pass* as well, thus this hypothetical graph set at this point of evaluation should be a script. Thus as per the Litmus test the inscriptions made during the Gupta dynasty i.e. the *unified* set of graphs (for eastern, western, northern and central Asian variants) are the font-variant of the Brahmi script.

Consider the graph set for the Braille & Latin alphabets with diacritics and conducting the Litmus test on these two sets performs as follows.

Criteria - Evaluation	Result	Criteria - Evaluation	Result
1 : Yes	PASS	6 : No	PASS
2 : No	FAIL	7 : Yes	PASS
3 : No	FAIL	8 : Yes	PASS
4 : No	FAIL	9 : Yes	PASS
5 : Yes	PASS	10 : Yes	PASS

Thus, as per the Litmus test the Braille graph set is a separate script.

Consider the graph set for the Ugaritic & Elamite cuneiform and conducting the Litmus test on these two sets performs as follows.

Criteria - Evaluation	Result	Criteria - Evaluation	Result
1 : Yes	PASS	6 : No	PASS
2 : No	FAIL	7 : Yes	PASS
3 : Yes	PASS	8 : Yes	PASS
4 : No	FAIL	9 : Yes	PASS
5 : No	FAIL	10 : Yes	PASS

Thus, as per the Litmus test the Ugaritic graph set is a separate script.

Consider the graph set for the Lisu alphabet & Latin alphabet and conducting the Litmus test on these two sets performs as follows.

Criteria - Evaluation	Result	Criteria - Evaluation	Result
1 : No	FAIL	6 : No	PASS
2 : No	FAIL	7 : Yes	PASS

3 : Yes	PASS	8 : Yes	PASS
4 : No	FAIL	9 : Yes	PASS
5 : Yes	PASS	10 : Yes	PASS

Thus, as per the Litmus test the Lisu graph set is a separate script.

A set or collection of graphs that doesn't conform to the Litmus test is a font. The guidelines imply that variance (font-variant or typeface or type-set of font-face) is an attribution of the graph set and the set of graphs should be a font variant of at least one script or be a script by itself. Therefore the above mentioned Litmus test can be conducted recursively for different graph sets.

A substitution cipher from a graphemic perspective could be prescribed as a font variant of an existing or a new script itself.

The Gaelic script is a font-variant of Latin script. The Kanji, Hanja are font-variant of the Hanzi. The Arabic language can be written in several forms, one can write Arabic in the standard Naskh, or Nastaliq or several of the calligraphic forms, as well as Latin with diacritics (IJMES transliteration) or Roman alphabets with numbers as one does in colloquial usage. These are different writing systems for the language and when graph sets are compiled together, they would be font variants (Naskh, Nastaliq and calligraphic forms) of the unified Arabic script in the former case or font variant (Latin with diacritics or Roman alphabets with numbers) of the unified Latin script. Similarly, the languages English or Portuguese or Swedish or Polish or Icelandic or Māori or Akan are all writing systems that employ the Roman alphabets and not scripts by themselves and essentially the graph sets for each of those languages are font variants of the unified Latin script.

The graph set that doesn’t conform to being font variant of a script nor a script, could be a writing system which encodes semasiographic data namely pasi-graph or kineto-graph, or yet undeciphered or unknown type.

References

1. Daniels, P. T. and Bright, W. (Eds.). (1996). *The World's Writing Systems*. Oxford University Press.
2. Bringhurst, R. (2012). *The Elements of Typographic Style*. Hartley & Marks Publishers.
3. Meletis, D., & Dürscheid, C. (Eds.). (2023). Writing systems and their use. De Gruyter Mouton.

Acknowledgements

I want to thank Prof. Dr. Øyvind Eide for the support and guidance for my doctoral studies and absolutely appreciate the tapestry of feedback, critique and suggestions by Dr. Jürgen Hermes, Dr. Phillipa Steele, Tim Brookes, Oreen Yousuf and Olgierd Uziembło on the current topic.