Kelih, E. et al. (eds.), Issues in Quantitative Linguistics, Vol. 2, 117-128

Complexity of the Vai script revisited: A frequency study of the syllabary

Andrij Rovenchak
Charles Riley
Tombekai Sherman

0. Introduction

The present paper is a continuation of the quantitative studies of writing systems, in particular in the domain of African indigenous scripts.

We analyze the statistical behavior of the script complexity defined according to the composition method suggested by Altmann (2004). The analysis of the Vai script complexity was made in a recent paper (Rovenchak, Mačutek, Riley 2009). To recall briefly, the idea of this approach is to decompose a letter into some simplest components (a point is given the weight 1, a straight line is given the weight 2, and an arc not exceeding 180 degrees has the weight of 3 units). The connections of such components are: a crossing (like in X) with weight 3, a crisp (like in T, V or L) with weight 2, and the continuous connection (like in O or S) with weight 1. For the Vai script, we also suggested that filled areas are given the weight of 2. With this method, a number of scripts had been analyzed: Latin (Altmann 2004), Cyrillic (Buk, Mačutek, Rovenchak 2008), several types of runes (Mačutek 2008), Nko (Rovenchak, Vydrin 2010).

The uniformity hypothesis for the distribution of complexity was confirmed for all the scripts but the Vai syllabary. The failure in the latter case can be caused by the fact that syllabaries require some modification of this hypothesis as all the other scripts analyzed so far are alphabets.

Here, it is worth mentioning the definitions of the discussed script types (Daniels 1990). In alphabets, a character (letter, symbol) denotes mostly one sound, either a consonant or a vowel. Some scripts do not mark all the vowels (e.g., in Arabic, only long vowels are written), and this type of scripts can be referred to as *abjad*. In a syllabary, a character denotes mostly a combination of the “consonant + vowel” type. The best known examples are Japanese *kana*, namely *hiragana* and *katakana* or the Cypriot syllabary. Slightly different approach is used in the Indic scripts derived from Brahmi, where special modifiers are added to change an inherent vowel associated with a particular symbol, hence such scripts are referred to as alphasyllabaries. Such an approach is also known for the Amharic script *fidel* having a complicated genealogy. This very script gave the name for this script type, *abugida*, from the standard letter order. Note that, for instance, Japanese *kana*, the Cypriot script, and native Cherokee script are pure syllabaries as the shapes of symbols with different vowels are unrelated. The same applies to the Vai script, an indigenous writing system of the Vai people (Liberia) originated in the 1820s.

After its invention, the Vai script continued in use through the remainder of the 19th century and into the 20th. Archives of manuscripts were kept at the villages of Jondu and Bandakolo, but both were burned in raids by the neighboring Golas around the turn of the century (Dalby 1967). Four manuscripts surviving from the nineteenth century are known to exist or were copied. In 1913, a 180-page manuscript by Boima Kiakpombo was produced, kept in diary form. Momolu Massaquoi, consul to Hamburg, produced translations in Vai script of some religious texts and began to collaborate with August Klingenheben, a German linguist. Klingenheben’s involvement with Vai culminated in a collaboration on a standardized form of the script, worked out at the University of Liberia with elders from several towns of the Vai country in 1962. Jangaba Johnson, Bai T. Moore, and Mohamed Nyei also were active throughout the latter half of the 20th century in working with Vai. In 2003, with the assistance of SIL and Lutheran Bible Translators, a New Testament in Vai was produced, with Tombekai Sherman serving as chair of the translation committee. Sherman continued to produce texts in Vai and, with Mohamed Nyei, S. Jabaru Carlon and others, contributed toward the standardization of the Vai script into Unicode.

1. Vai texts

We have analyzed four Vai texts. They are:

1) *The Universal Declaration of Human Rights* (UDHR);
2) *Vai Proverbs and Rhymes* (VPR);
3) *Our Village* (OV);
4) *Who Were the Vai?* (WWV).

The Universal Declaration of Human Rights (UDHR) was written under the authority of a standing body of the United Nations, the Commission on Human Rights, consisting of Eleanor Roosevelt (chair), with contributions from John Peters Humphrey, René Cassin, Charles Habib Malik, Chang Peng-chun, and other members of the Commission. On December 10, 1948, it was adopted by the UN General Assembly with no dissent, and only eight abstentions. Its translation into Vai was prepared by Tombekai Sherman.

Vai Proverbs and Rhymes were compiled in 2008 by Tombekai Sherman, the list contains over 200 proverbs supplied with some introductory information and comments.

Our Village is a narrative written by Tombekai Sherman.

Who Were the Vai? is an article published in the *Journal of African History* by Adam Jones (1981). Its translation into the Vai language, using the Vai script, was undertaken by Sherman between 2007 and 2009 with the permission of Cambridge University Press.

The abovementioned texts are homogeneous with respect to the orthography which in an important issue for the study of texts written in indigenous scripts.

2. Preliminary notes

Our previous study of the complexity of the Vai script revealed that the distribution of complexity values does not conform to the uniformity hypothesis. Comparing to alphabets, the set of characters in syllabaries is typically several times larger. The size of this set significantly depends on the phonotactics of a particular language and on the approach used to map the phonetic structure onto the written representation. For instance, the Cherokee syllabary has 85 characters, the Vai syllabary has over 200 characters, but the modern Yi script contains about 800 signs as the tonal distinction are implemented as separate shapes.

We suggest dividing the symbols in the syllabary into two parts, *core* (containing the most frequent characters) and *periphery* (correspondingly, all the remaining characters). There is no unique way to define the number of characters in the core, and we propose to use the notion of *h*-point (Popescu et al. 2009) in order to separate the syllabary.

The h-point definition. Given a set of data sorted in a descending order with respect to absolute frequencies, assign the rank $r = 1$ to the most frequent item, then rank $r = 2$ to the next most frequent, etc. Let the frequency of the r-ranked item be $f(r)$. The *h*-point is defined as the solution of the equation

(1) $r_h = f(r_h)$,

so that all the items with frequencies higher than $f(r_h)$ have ranks lower than r_h and vice versa.

This definition is directly related to the so-called h-index suggested by Hirsch (2005) to measure the output of a scientist. A similar quantity was proposed several decades earlier by Sir Arthur Eddington for the estimation of the achievements of long-distance cyclists (Barrow 2002, p. 83; E Numbers 2008).

It is possible that in some sample no frequency satisfies Eq. (1). In such a case, the value of the *h*-point can be defined by simple interpolation between the neighboring values corresponding to ranks r_1 and r_2 as follows:

$$\text{(2)}\quad r_h = \left\lceil \frac{r_2 f(r_1) - r_1 f(r_2)}{r_2 - r_1 + f(r_1) - f(r_2)} \right\rceil,$$

As both rank and absolute frequency are integers, we apply the ceiling function $\lceil x \rceil$ equal to the next integer after *x*.

When applied to the word frequencies, the *h*-point is believed to divide the word list into mostly synsemantic (auxiliary) and autosemantic (full-sense words) branches (Popescu et al. 2009). The relation between these parts of a dictionary seems however more complicated and probably cannot be defined solely by the *h*-point (cf. Buk 2010). Still, the *h*-point can be seen as separating a rank–frequency distribution into two regimes, and this very property we use in our work.

In the present work, we analyze the frequency distribution of syllabic characters of the Vai script and propose that the *core part* of the syllabary is composed of the characters having frequencies above and equal to the *h*-point. Note that such a definition means that different cores would be obtained from different samples, and only a large corpus of texts would allow stating some *gold-standard core*.

3. Frequency data and core part of the syllabary

We have compiled frequency lists for characters for all the Vai texts mentioned in the previous section. The frequency lists are given in Table 1.

Table 1
Frequency of the syllabic characters in Vai texts

r	OV				UDHR				VPR				WWV			
	Vai	f_r	C_r	PhT	Vai	f_r	C_r	PhT	Vai	f_r	C_r	PhT	Vai	f_r	C_r	PhT
1	ꕉ	365	16	a	ꕉ	361	16	a	ꔤ	330	24	i	ꖃ	190	13	lɔ
2	ꘋ	257	19	ŋ	ꖸ	265	20	nu	ꘋ	302	19	ŋ	ꕞ	164	8	la
3	ꖸ	252	20	nu	ꕌ	259	25	ha	ꕉ	233	16	a	ꕮ	161	17	ma
4	ꖷ	247	11	mu	ꖷ	218	11	mu	ꕌ	216	25	ha	ꘋ	156	19	ŋ
5	ꔤ	214	24	i	ꔤ	212	24	i	ꕮ	205	17	ma	ꔤ	116	24	i
6	ꕮ	206	17	ma	ꘋ	201	19	ŋ	ꗡ	195	26	e	ꗛ	111	23	ko
7	ꕌ	199	25	ha	ꕮ	190	17	ma	ꕎ	190	13	wa	ꔀ	74	32	ɛ

r	OV				UDHR				VPR				WWV			
	Vai	f_r	C_r	PhT	Vai	f_r	C_r	PhT	Vai	f_r	C_r	PhT	Vai	f_r	C_r	PhT
8	ꕎ	188	13	wa	ꖒ	166	12	mɔ	ꕞ	169	8	la	ꕎ	73	13	wa
9	ꖃ	162	13	lɔ	ꕊ	159	14	an	ꕩ	162	23	ya	ꔌ	71	18	fɛ
10	ꔀ	141	32	ɛ	ꖃ	145	13	lɔ	ꖒ	141	12	mɔ	ꖷ	65	20	nu
11	ꕞ	138	8	la	ꕞ	144	8	la	ꖃ	119	13	lɔ	ꕯ	62	10	na
12	ꔞ	135	18	kɛ	ꕎ	143	13	wa	ꗶ	117	20	le	ꕌ	59	25	ha
13	ꗀ	106	26	e	ꕱ	132	28	ɔ	ꖶ	109	11	mu	ꗀ	57	26	e
14	ꕊ	103	14	an	ꗚ	119	23	ko	ꖳ	104	9	ku	ꕊ	56	14	an
15	ꔆ	103	5	ɓɛ	ꕩ	117	23	ya	ꔳ	100	15	ti	ꕩ	52	23	ya
16	ꖒ	95	12	mɔ	ꔧ	110	26	hin	ꔞ	96	18	kɛ	ꕉ	51	16	a
17	ꕩ	93	23	ya	ꔀ	107	32	ɛ	ꗨ	96	11	ɓe	ꕱ	50	28	ɔ
18	ꗚ	89	23	ko	ꖳ	106	9	ku	ꕯ	87	10	na	ꕪ	50	7	ka
19	ꕯ	85	10	na	ꔞ	103	18	kɛ	ꕊ	86	14	an	ꔞ	48	18	kɛ
20	ꗶ	84	20	le	ꕢ	100	33	sa	ꗚ	86	23	ko	ꔻ	46	33	si
21	ꕑ	77	24	ɓa	ꔳ	96	15	ti	ꔧ	85	26	hin	ꖳ	40	9	ku
22	ꕡ	76	25	nɗa	ꔆ	92	5	ɓɛ	ꔀ	81	32	ɛ	ꖏ	40	26	kɔ
23	ꕱ	71	28	ɔ	ꕯ	79	10	na	ꔆ	81	5	ɓɛ	ꕃ	39	9	ki
24	ꖆ	67	24	nɗɔ	ꗀ	76	26	e	ꔝ	79	28	yɛ	ꗎ	30	20	lo
25	ꔳ	66	15	ti	ꕰ	67	26	ɲa	ꕪ	77	7	ka	ꕙ	29	28	va
26	ꕧ	65	20	ja	ꗎ	65	20	lo	ꖷ	70	20	nu	ꔒ	28	6	lɛ
27	ꕪ	64	7	ka	ꕑ	63	24	ɓa	ꖾ	66	22	wo	ꖾ	28	22	wo
28	ꕢ	63	33	sa	ꕧ	62	20	ja	ꕱ	64	28	ɔ	ꔳ	27	15	ti
29	ꔧ	62	26	hin	ꔰ	62	11	gbi	ꗎ	63	20	lo				
30	ꖾ	59	22	wo	ꕿ	61	14	tɔ	ꕧ	61	20	ja				
31	ꔁ	56	27	ɛn	ꕚ	60	14	ta	ꕑ	56	24	ɓa				

r	OV				UDHR				VPR				WWV			
	Vai	f_r	C_r	PhT	Vai	f_r	C_r	PhT	Vai	f_r	C_r	PhT	Vai	f_r	C_r	PhT
32	[illegible]	55	11	ɓe	[illegible]	59	22	sɔ	[illegible]	56	25	nɗa				
33	[illegible]	47	22	sɔ	[illegible]	59	7	ka	[illegible]	56	26	kɔ				
34	[illegible]	44	30	bi	[illegible]	54	28	yɛ	[illegible]	46	22	o				
35	[illegible]	43	8	mɛ	[illegible]	54	6	lɛ	[illegible]	46	26	ɲa				
36	[illegible]	42	9	ku	[illegible]	53	27	ɛn	[illegible]	45	6	to				
37	[illegible]	39	14	ta	[illegible]	52	20	le	[illegible]	44	33	sa				
38	[illegible]	38	11	gbi	[illegible]	48	18	fɛ	[illegible]	44	7	lu				
39					[illegible]	45	19	ɓo	[illegible]	44	22	sɔ				
40					[illegible]	37	18	nɗe	[illegible]	43	27	ɛn				
41									[illegible]	43	12	kpa				
42									[illegible]	42	13	ji				
r_h	38				40				42				28			
% till r_h	79.1				85.1				79.4				63.6			
Total	5432				5407				5587				3102			

Legend: r — rank in the frequency list; f_r — absolute frequency of the character; C_r — complexity of the character; PhT — phonetic value of the character (vowel nasalization is marked by /n/; according to the conventions typical for African linguistics, /y/ denotes IPA [j] and /j/ denotes IPA [ʤ]); % till r_h — the percentage of the characters from the core syllabary relative to all occurring characters (given in the Total row).

All the studied texts but WWV have almost equal number of syllabic characters. As expected (cf. Popescu et al. 2009:19), this reveals close estimations for the *h*-point of all three texts and a smaller value for the shorter WWV text. The same applies to the percentage of text covered by the characters from the core.

From the frequency point of view, the proportion of 80% roughly corresponds to an English text lacking ‘j’, ‘x’, ‘q’, and ‘z’ (cf. Lewand 2000 or any other statistics on English letter frequencies). For the illustration, in the remaining part of this paragraph the te#t is typed without these letters substituting them with a single number sign ‘#’. It is easily seen that such a modification does not cause any problems in understanding. An obvious reason is the low fre#uency of the mentioned letters (less than 0.2 per cent). Even #ointly, the four least-fre#uent letters occur less than the next, ranked 22nd letter ‘k’. Note however that omitting a letter in an English te#t introduces probably less confusion than omitting a syllabic sign in a Vai one as words in Vai contain 2–3 syllables on average.

The cumulative list of characters from all the four texts consists of 55 symbols (occurring at least in one core part). These characters are listed in Table 2 ordered according to the Unicode values.

Table 2
Cumulative list of characters from the core parts of the syllabary from different texts

Vai	*C*	PhT	Vai	*C*	PhT	Vai	*C*	PhT	Vai	*C*	PhT	Vai	*C*	PhT
[illegible]	26	e	[illegible]	9	ki	[illegible]	33	sa	[illegible]	20	lo	[illegible]	26	kɔ
[illegible]	11	ɓe	[illegible]	16	a	[illegible]	20	ja	[illegible]	23	ko	[illegible]	12	mɔ
[illegible]	20	le	[illegible]	14	an	[illegible]	23	ya	[illegible]	7	lu	[illegible]	32	ɛ
[illegible]	18	nɗe	[illegible]	25	ha	[illegible]	7	ka	[illegible]	9	ku	[illegible]	27	ɛn
[illegible]	24	i	[illegible]	13	wa	[illegible]	17	ma	[illegible]	11	mu	[illegible]	5	ɓɛ
[illegible]	26	hin	[illegible]	24	ɓa	[illegible]	10	na	[illegible]	20	nu	[illegible]	18	fɛ
[illegible]	30	bi	[illegible]	28	va	[illegible]	26	ɲa	[illegible]	28	ɔ	[illegible]	6	lɛ
[illegible]	11	gbi	[illegible]	12	kpa	[illegible]	22	o	[illegible]	14	tɔ	[illegible]	28	yɛ
[illegible]	15	ti	[illegible]	14	ta	[illegible]	22	wo	[illegible]	13	lɔ	[illegible]	18	kɛ
[illegible]	13	ji	[illegible]	8	la	[illegible]	19	ɓo	[illegible]	24	nɗɔ	[illegible]	8	mɛ
[illegible]	33	si	[illegible]	25	nɗa	[illegible]	6	to	[illegible]	22	sɔ	[illegible]	19	ŋ

From the point of view of a literate native speaker in the Vai script, one does not need to know a whole set of characters to be able to read and write using the syllabary. When literacy programs were intensified among Vai people, it took between two and three months for most people to be able to read most of the literature that was produced at that time since there were not so many characters involved.

One can compare this cumulative list to the repertoire of characters found in the *Book of Ndɔlɛ* (known also as the *Book of Rora*), the earliest known Vai manuscript of ca. 1850 (Stewart 1972). The cumulative set of characters is about twice smaller than that of the *Book of Ndɔlɛ* with some 120 syllabic signs and sixteen ideograms. The majority of the cumulative list is found in the character repertoire of the *Book of Ndɔlɛ*. Not occurring there are most individual vowels, both oral and nasal (/e, an, o, ɔ, ɛ, ɛn/), and the syllable /ha/. Prenasalized nɗ-series are written with different shapes in the *Book of Ndɔlɛ*. All the mentioned differences can be explained, most probably, by different orthographic approaches used in the old manuscript and in modern texts (cf. Stewart 1972).

4. Uniformity tests

As we mentioned in the Introduction, previous studies revealed the uniform distribution of complexities for alphabetic scripts but not in the Vai syllabary. The uniformity hypothesis can be tested by the Wald–Wolfowitz runs test (Wald & Wolfowitz 1940; cf. also Stewart 2009, Chap. 17; Rajagopalan 2006, p. 187-188). We demonstrate it here for the cumulative set of characters from Table 2. All the relevant numbers are listed in Table 3 for four Vai texts analyzed in this work. Let $I = 55$ denote the inventory size and $R = 28$ is the range of complexities. The uniform distribution of data means that all expected frequency values equal $E = I/(R+1)$. A run is a sequence of frequencies which are either all greater than E or all smaller than E. For the case under consideration $E \simeq 1.9$ and $r = 12$, namely [1, 2,2,2,2, 1, 3,2,3,3, 1,1,1, 3,2,4, 0, 3,2,3,2,4, 1, 3, 0,1,0,1, 2]. Let $n_1 = 11$ is the number of frequencies smaller than E, $n_2 = 18$ is the number of frequencies larger than E, and $n = R + 1 = 29$.

Table 3
Distribution of complexities in Vai texts.

C \ f_C	OV	UDHR	WWV	PR	Cumulative
5	1	1	—	1	1
6	0	1	1	1	2
7	1	1	1	2	2
8	2	1	1	1	2
9	1	1	2	1	2
10	1	1	1	1	1
11	3	2	0	2	3
12	1	1	0	2	2
13	2	2	2	3	3
14	2	3	1	1	3
15	1	1	1	1	1
16	1	1	1	1	1
17	1	1	1	1	1
18	1	3	2	1	3

C \ f_C	OV	UDHR	WWV	PR	Cumulative
19	1	2	1	1	2
20	3	4	2	4	4
21	0	0	0	0	0
22	2	1	1	3	3
23	2	2	2	2	2
24	3	2	1	2	3
25	2	1	1	2	2
26	2	3	2	4	4
27	1	1	0	1	1
28	1	2	2	2	3
29	0	0	0	0	0
30	1	0	0	0	1
31	1	0	0	0	0
32	1	1	1	1	1
33	1	1	1	1	2
I	38	40	28	42	55
R	28	28	27	28	28
$E = I / (R+1)$	1.3	1.4	1.0	1.5	1.9
n_1	19	19	21	18	11
n_2	10	10	7	11	18
n	29	29	28	29	29
r (runs)	11	13	15	11	12
$E(r)$	14.1	14.1	11.5	14.7	14.7
σ_r	2.38	2.38	1.92	2.48	2.57
z	1.09	0.25	1.56	1.27	1.51

The number of runs is considered random (meaning that the distribution is uniform) if

$$z = |r - E(r)| - 0.5\sigma_r < 1.96,$$

where

$$E(r) = 1 + \frac{2n_1n_2}{n} \quad \text{and} \quad \sigma_r^2 = \frac{2n_1n_2(2n_1n_2 - n)}{n_2(n-1)}.$$

The results of the calculations are presented in Table 3. As one can see, the defined core parts in all the samples, as well as the cumulative list of characters, confirm the uniformity hypothesis.

5. Correlation between complexity and frequency in the Vai syllabary

Previously, the correlation between complexity and frequency was studied for the Nko script (Rovenchak & Vydrin 2010), and the values of the Pearson correlation coefficient $r_P = –0.39$ and the Spearman correlation coefficient $r_S = –0.20$ were obtained. We have analyzed the four Vai texts with respect to this property by calculating the respective correlation coefficients. The full frequency lists were taken into consideration, not only that of the core part. The data are shown in Table 4 in comparison with the Morse code for English.

Table 4
Correlation coefficients in Vai texts

	OV	UDHR	WWV	PR	Nko	Morse code (English)
Pearson	–0.15	–0.09	–0.14	–0.09	–0.39	–0.81
Spearman	–0.20	–0.12	–0.11	–0.13	–0.20	–0.79

Small values of the coefficients suggest that the simplification of shapes is not a prevailing mechanism in the development of a script, but it still has some [marginal] role. Note the values of the correlation coefficients of the Morse code which was artificially created basing on the frequency considerations. Complexity of the Morse code was defined according to the standard duration of its elementary signals, namely 1 for the short signal ('dot') and 3 for the long signal ('dash').

Interestingly, the highest correlation is obtained for the text being a continuous narrative, namely *Our Village*. Such texts are the most natural for the analysis of frequency structure and for the correlation study in particular.

6. Conclusions

In the paper, we proposed the definition of the *core part* of a syllabary basing on the *h*-point, which is obtainable from the frequency analysis of text. The characters with frequencies higher and equal to the value of the *h*-point are considered as belonging to the core. Previously, it was suggested that some set of the most frequent characters can be tested with respect to the uniformity hypothesis for the distribution of complexities, which otherwise failed for the whole syllabary. This hypothesis was confirmed in all the considered cases for the core part of the Vai syllabary defined as above. The core parts obtained from different texts significantly overlap, and the compiled cumulative list of core characters contains 55 symbols versus 42 symbols in the largest core for an individual text.

No considerable differences of the frequency structure occur between the indigenous texts (OV and VPR) and translated ones (UDHR and WWV). This should facilitate further studies of Vai texts as many of them are translations.

Lastly, we studied the correlation between the complexity of characters and their frequency. Small but negative values of both Pearson's and Spearman's correlation coefficients signal that the simplicity of shapes is not a key feature in the script development. This issue requires a broader analysis, with more texts (in particular, from different periods of time) if speaking specifically about the Vai syllabary and with more scripts in general.

References

Altmann, G. (2004). Script complexity. *Glottometrics 8, 68-74*.

Barrow, John D. (2002). *The constants of nature*. New York: Pantheon Books.

Buk, Solomija (2010). Statystyčna struktura romanu Ivana Franka *Boryslav smijetjsja* [Statistical structure of *Boryslav Laughs*, a novel by Ivan Franko]. *Scientific Notes of Taurida V. I. Vernadsky National University. Series: Philology. Social communications. 23*[*62*](*3*), *114-118.*

Buk, S., Mačutek, J., Rovenchak, A. (2008). Some properties of the Ukrainian writing system. *Glottometrics 16, 63-79*.

Dalby, D. (1967). A survey of the indigenous scripts of Liberia and Sierra Leone: Vai, Mende, Loma, Kpelle and Bassa. *African Language Studies 8, 1–51*.

Daniels, Peter T. (1990). Fundamentals of grammatology. *Journal of the American Oriental Society, 110*(*4*), *727-731*.

E Numbers (2008). In: *The Mudflap: Hertfordshire Wheelers Monthly Newsletter* (December), *2-3*.

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceeding of the National Academy of Sciences USA 102*(*45*), *16569–16572*.

Jones, Adam (1981). Who Were the Vai? *The Journal of African History 22*(*2*), *159-178.*

Lewand, Robert Edward (2000). *Cryptographical mathematics*. Washington, DC: The Mathematical Association of America Publishing.

Mačutek, J. (2008). Runes: complexity and distinctivity. *Glottometrics 16, 1-16.*

Popescu, I.-I. et al. (2009). *Word frequency studies.* Berlin–New York: Mouton de Gruyter.

Rajagopalan, V. (2006). *Selected statistical tests.* New Age International.

Rovenchak, A., Mačutek, J., Riley, C. (2009). Distribution of complexities in the Vai script. *Glottometrics 18, 1–12*.

Rovenchak, A., Vydrin, V. (2010). Quantitative properties of the Nko writing system. In: P. Grzybek, E. Kelih, & J. Mačutek (eds.) *Text and Language:*

Structures - Functions - Interrelations. Quantitative perspectives: 171-181 Wien: Praesens.

Stewart, Gail (1972). The early Vai script found in the Book of Ndole. In: *Conference on Manding Studies 1-27*. London: School of Oriental and African Studies.

Stewart, William J. (2009). *Probability, Markov chains, queues, and simulation: the mathematical basis of performance modelling.* Princeton*:* Princeton University Press.

Wald, A., Wolfowitz, J. (1940). On a test whether two samples are from the same population. *The Annals of Mathematical Statistics 11*(*2*)*, 147-162*.