

A STATISTICAL COMPARISON OF VERBS AND NOUNS IN ROGLAI

MAXWELL COBBEY

Word counts were made of four Roglai texts,¹ distinguishing between (1) noun-words, (2) verb-words, and (3) all other words. All words were classified contextually as to grammatical class, so that, for instance, noun-words included adjectives, classifiers, quantifiers, etc., in addition to nouns and pronouns, whenever they functioned endocentrically with nouns, or as noun substitutes. However, in another place an adjective might be counted as a verb-word because of its verbal function in that context.

1. WORD AND CLUSTER COUNTS

Besides this word count of grammatical types, a count was also made of the number of clusters of each of the three types, a cluster being defined as a consecutive string of one or more words of the same grammatical type. These counts of word types and cluster types are given in Table I.

TABLE I
Word and Cluster Counts

Name of Text	Word Count				Cluster Count			
	Nouns	Verbs	Other	Total	Nouns	Verbs	Other	Total
History	1512	753	394	2659	603	601	305	1509
Monkey	156	137	57	350	92	91	42	225
Eagle	124	78	45	247	68	42	42	152
Feast	325	209	66	600	173	153	63	389
Totals	2117	1177	562	3856	936	887	452	2275

By making calculations based on the counts listed in Table I, various comparisons were made (a) between the four Røglai texts, and (b) between the grammatical types.

2. RELATIVE NUMBER OF WORDS WITHIN TEXTS²

Noun-words are more numerous than any other type in all the texts; and two of the texts each consist of more than half noun-words, as seen in Table II. I feel that these two texts, History and Feast, are by far the two best texts from a literary standpoint.

TABLE II
Relative Number of Words within Texts

Text	Noun	Verb	Other	Total
History	0.57	0.28	0.15	1.00
Monkey	0.45	0.39	0.16	1.00
Eagle	0.50	0.32	0.18	1.00
Feast	0.54	0.35	0.11	1.00
Average ³	0.55	0.31	0.15	1.00

3. RELATIVE NUMBER OF CLUSTERS WITHIN TEXTS

A comparison of the number of clusters of each grammatical type within the texts (Table III) shows that for two of the texts, the History and the Monkey, there are about the same number of noun and verb clusters. These two texts contain more action, whereas the Eagle and the Feast are more descriptive, and consequently show more noun clusters than verb clusters.

TABLE III
Relative Number of Clusters within Texts

Text	Noun	Verb	Other	Total
History	0.40	0.40	0.20	1.00
Monkey	0.41	0.40	0.19	1.00
Eagle	0.45	0.28	0.28	1.00
Feast	0.45	0.39	0.16	1.00
Average ³	0.41	0.39	0.20	1.00

4. CLUSTER LENGTHS

From Table IV below it is seen that noun clusters are longer on the average than verb clusters. The two best literary texts (History and Feast) are seen to have the greatest difference between noun cluster length and verb cluster length.

TABLE IV
Words per Cluster

Text	Noun Cluster	Verb Cluster	Other Cluster	Average
History	2.51	1.25	1.29	1.76
Monkey	1.70	1.51	1.36	1.56
Eagle	1.82	1.86	1.07	1.62
Feast	1.88	1.37	1.04	1.54
Average ³	2.26	1.33	1.24	1.69

5. RELATIVE CLUSTER LENGTHS WITHIN TEXTS

The relative cluster length of each grammatical type within a text can be seen in Table V, which is obtained by dividing the cluster lengths (of Table IV) by the average cluster length of a text (the last column of Table IV). Here it is seen that the two best literary texts have the widest range of relative cluster length between grammatical types.

TABLE V
Relative Cluster Lengths within Texts

Text	Noun	Verb	Other	Range ⁴
History	1.42	0.71	0.73	0.71
Monkey	1.09	0.97	0.87	0.22
Eagle	1.12	1.14	0.66	0.48
Feast	1.22	0.89	0.68	0.54
Average ³	1.33	0.78	0.73	0.60

6. RELATIVE CLUSTER LENGTHS BETWEEN TEXTS

A comparison of cluster lengths between texts for a given grammatical type was made by dividing the cluster length of each text by the average cluster length for the type, as in Table VI. Verbs are seen to have the widest range of relative cluster lengths between the texts.

TABLE VI
Relative Cluster Length Between Texts

Text	Noun	Verb	Other	Average ³
History	1.11	0.94	1.04	1.04
Monkey	0.75	1.13	1.09	0.92
Eagle	0.81	1.40	0.86	0.96
Feast	0.83	1.03	0.84	0.91
Range ⁴	0.36	0.46	0.25	0.13

7. DENSITY OF GRAMMATICAL TYPES

An empirical density coefficient was calculated for each cluster type within each text. This was obtained by multiplying the entries of Table V by the entries of Table II. The greater the relative cluster length of a grammatical type, the greater its coefficient of density. And the greater the relative number of words of that type in a text, the greater the coefficient of density.⁵

TABLE VII
Density of Grammatical Types within Texts

Text	Noun	Verb	Other
History	0.81	0.20	0.09
Monkey	0.49	0.38	0.14
Eagle	0.56	0.36	0.12
Feast	0.66	0.31	0.07
Average ³	0.73	0.24	0.11

8. LITERARINESS

As mentioned previously, the two best literary texts have the greatest differences between average noun cluster length and verb cluster length. Another characteristic of some good literary texts is relative shortness of clusters, due to the grammatical types being intermixed in semi-poetical types of sentences. These two somewhat opposing characteristics of cluster length are combined in an empirical formula for literariness, Table VIII, which gives the absolute difference between noun and verb cluster length divided by the average length of clusters in that text. The History and Feast texts come out with the highest coefficients in literariness.

TABLE VIII
Coefficient of Literariness

History	0.71
Monkey	0.12
Eagle	0.02
Feast	0.33
Word-weighted average	0.53
Non-weighted average	0.30

9. A SAMPLE ENGLISH COMPARISON

Similar counts and calculations were made of an article in English for comparison, using an article from the *Reader's Digest*, October 1975: 73-7.

	Noun	Verb	Other	Total
I. Word counts	1033	346	344	1723
Cluster counts	537	198	285	1020
II. Rel. no. of words	0.60	0.20	0.20	
III. Rel. no. of clusters	0.53	0.19	0.28	
IV. Cluster length	1.92	1.75	1.21	
V. Rel. cluster length	1.14	1.03	0.71	
VI. (only one text)				
VII. Density of gram. types	0.68	0.21	0.14	
VIII. Literariness coeff.				0.10

From II and III above, the English text is seen to have a greater relative number of both noun words and of noun clusters than the Rŏglai texts. And the English text has a smaller relative number of verb words and verb clusters than any of the Rŏglai texts.

Tables IV and V reveal that both the cluster length and the relative cluster length of the English text fall within the range of the Rŏglai texts. The noun and verb density coefficients (VII) of the English text also fall within the range of the Rŏglai texts.

The coefficient of 'literariness' of the English text is lower than three of the Rŏglai texts. We defined 'literariness' solely for Rŏglai as having short clusters and a great difference between noun and verb cluster length. Good literary style in English clearly has different characteristics from Rŏglai.

N O T E S

1. Roglai is a language belonging to the Coastal Chamic branch of Austronesian, found in south Vietnam inland from Nhatrang to Phanthiet.
2. The relative number of nouns in the History text, for instance, is found by dividing the number of nouns in the History text, 1512 (Table I), by the total number of words in the History text, 2659 (Table I): which gives 0.57.
3. These averages are not obtained by averaging the figures given in this chart, but must be obtained from the original data of Table I; for instance, for the nouns of Table II $2117 \div 3856 = 0.55$.
4. The range, for instance, of relative cluster length in the History text is $1.42 - 0.71 = 0.71$; for the Monkey text $1.09 - 0.87 = 0.22$.
5. This may also be stated: within each text the density coefficients of the respective grammatical types vary in direct proportion to both the relative lengths of clusters of each type and the relative numbers of words of each type. It may be calculated directly for each grammatical type by the formula:

$$\text{Density coefficient}_A = \frac{(\text{Words}_A)^2}{(\text{Words}_T)^2} \times \frac{\text{Clusters}_T}{\text{Clusters}_A}$$

where A = the given grammatical type

T = the total words (or clusters) of all types.