LINGUISTIC-MATHEMATICAL STATISTICS IN RECENT ROMANIAN POETRY

Florentin Smarandache
University of New Mexico
200 College Road
Gallup, NM 87301, USA
E-mail: smarand@unm.edu

"Mathematics is logical enough to be able to detect the internal logics of poetry and crazy enough not to lag behind the poetic ineffable" (Solomon Marcus).

The author of this article aims a statistical investigation of a recently published volume of poetry [3], which will make possible some more general conclusions on the evolution of poetry in the XXth century (either the literary current hermetism, surrealism or any other). Certain modifications in the structure of poetry, occurred in its evolution from classicism to modernism, are also presented. Men of letters have never agreed with mathematics and, especially, with its interference in art. Let us quote one of them: "Remarque que, a mon avis, tout literature est grotesque…(…) La seule excuse de l'écrivain c'est de se rendre compte qu'il joue, que la littérature est un jeu" (Eugène Ionesco). Well, if literature is a game why could not be subjected to mathematical investigation?

The book chosen for this study (see [3]) contains 44 poems (from which the first and the last are sort of poems essays on Romanian poetry). It comprises over 250 sentences, over 700 verses, over 2,500 words and over 11,700 letters (not sounds).

MORPHOLOGICAL ASPECTS

1. The frequency of words depending on the grammatical category they belong to.

1. Nouns	35.592%	
2.Verbs (predicat.moods)	13.079%	"Empty" words
3.Adjectives	6.183%	40.271%
4.Adverbs	4.829%	
"Full" words	59.729%	

1. The "full" words category includes – according to the author – nouns, verbs (predicative moods only), adjectives and adverbs. The "empty" words category includes verbs (i,e, infinitives, gerunds, poet participles, supines), numerals, articles, pronouns, conjunctions, prepositions and interjections. The same terminology was also used by Solomon Marcus in his "Poetica matematica" published by Ed. Academiei, Bucharest, 1970 (it was translated in German and published by Athenäum, Frankfurt-am-Mein, 1973).

2. The average distribution of "full" words[1] per verses (lines), sentences, poems

a) 1.255	nouns/line	
b) 0.461	verbs (p.m)/line	
c) 0.218	adjectives/line	
d) 0.172	adverbs/line	
e) 3.464	nouns/sentence	
f) 1.273	verbs (p.m)/sentence	
g) 0.602	adjectives/sentence	
h) 0.475	adverbs/sentence	
i) 20.393	nouns/poem	
j) 7.492	verbs (p.m)/poem	
k) 3.543	adjectives/poem	
l) 2.792	adverbs/poem	

We may conclude:

CONJECTURE 1. In the recent Romanian poetry the percentage of adjectives is, on average, under that of the total of words.

CONJECTURE 2. The percentage of verbs (predicative moods) is., on average, under 15% of the total of the total words.

In support of conjectures 1 and 2 we also mention:

- only one in six nouns is modified by an adjective, i.e. the role of the adjective diminishes and there are poems with no adjectives (see [3], pp. 9, 12, 20);

- on average, there is one verb in a predicative mood in more than two lines, i.e. the role of the verbal predicate decreases and there are poems with no verbal predicates (see [3], p. 20);

(From classicism to modernism both adjectives and verbal predicates gradually but constantly regressed).

- the poetry of the young poets is characterized by economy of words and, implicitly, by the avoidance of the overused words; the adjectives were favored by the romantics and the young poets feel the necessity to "renew" poetry;

- this renewal and effort to avoid the trivial may be also helped by elimination of adjectives. The strict use of adjectives or verbal predicates is also accounted for by the characteristics of the two main literary currents of our century.

a) hermetism – appeared after World War I – consists, mainly in the hyper intellectualization of language and its codification; an adjective (i.e. an explanation concerning an object) or the predicative mood of a verb (strict definition of the grammatical tense) may diminish the degree of ambiguity, generalization or abstraction intended by the poet.

b) Surrealism – literary of vanguard – aimed at detecting the irrational, the unconscious, the dream; because of its precise definite character, the adjective makes the reader "plunge" into the so carefully avoided real world.

CONJECTURE 3. In the recent Romanian poetry percentage of "full" words is over 55% of the total words.

Unlike in the spoken language in which the percentage of "full" and "empty" words is equal (see [1]) in poetry the percentage of "full" words is greater. This is due to the fact that poetry is essence, it is dense, concentrated. The percentage of "full" words and the "density" of a literary work are directly proportional.

As a conclusion to the three conjectures we may say that:

- in its evolution from classicism to modernism the percentage of nouns increased, while that of verbs decreased, less adverbs are used, on the other hand, because of the smaller number of verbs. In all, however, the percentage of "full" words increased.

3. The frequency of the nouns with and without an article.

1. Percentage of nouns with an article	- 47.884%
2. Percentage of nouns without an article	- 52.116%

CONJECTURE 4. In the recent Romanian poetry the number of nouns with an article is, on an average, smaller than the number of those without an article. With an article the noun is more definite, specified which are characteristics undesirable from the same viewpoint as that mentioned above. That is why the indefinite article is favored in modern poetry. The consequence of this preferred indefinite character of the noun enlarges the abstraction, generalization, ambiguity and, hence, the "density " of the poem. (See also the second part of assertions 1 and 2 and the statistical conjecture 3). In its evolution from classicism to modernism the number of nouns without an article used in poetry also increased.

4. The frequency of nouns depending on the grammatical case they belong to.

Nominative	Genitive	Dative	Accusative	Vocative
29.497%	19.888%	0.335%	50.056%	0.224%
2	3	4	1	5
↑C L A S S I F I C A T I O N↑				

CONJECTURE 5. In the poems under study, over 75% of the nouns are accusative or nominative.

5. Sentences, lines, words, syllables, letters – average relationships

a) 2.402	letters/syllable
b) 1.933	syllables/word
c) 4.643	letters/word
d) 3.528	words/line
e) 6.820	syllables/line

f) 16.380	letters/line	

g) 2.760	lines/sentence
h) 9.737	words/sentence

i) 18.823	syllables/sentence
j) 45.208	letters/sentence

k) 5.887	sentences/poem
l) 16.250	lines/poem
m) 57.330	words/poem
n) 110.825	syllables/poem
o) 266.175	letters/poem

Conclusion: the poems are of medium length; the lines are short while the sentences are, again, of medium length.

6. The frequency of words according to their length (in syllables)

Syllables	Percentages	Order
1	41.509%	1
2	32.069%	2
3	19.363%	3
4	5.688%	4
5	1.371%	5
6	0.000%	6

The total number of syllables in the volume is … 4,800. The frequency of words and their length (in syllables are in inverse ratio. Long words seem "less poetical".

CONJECTURE 6. In the recent Romanian poetry the percentage of words of one and two syllables is … 75%. Again, it seems that short and very short words (of one and two syllables) appear more adequate to satisfy the internal rhythm of the poem. Longer words already have their own rhythm dictated by the juxtaposition of the syllables; it is very probable that this rhythm comes into … with the rhythm imposed by the poem. Shorter words are more easily uttered; longer words seem to render the text more difficult.

7. The frequency of words according to their length (in letters)

1 letter	2	3	4	5	6	7	8	9	10	11	12	13	14
3.604 %	25.426%	8.475%	11.089%	13.347%	13.149%	13.703%	5.861%	3.129%	1.149%	0.752%	0.237%	0.079%	0.000%
Order 8	1	6	5	3	4	2	7	9	10	11	12	13	14

In the whole volume there are only two words of 13 letters and 6 of twelve. A 90% of the words consist of no more than 7 letters.

CONJECTURE 7. In the recent Romanian poetry the percentage of the two letter words is, on average, about 25% of the words. In fact, the same percentage, or even higher, is found in the ordinary language. Because of esthetic reasons in poetry there is a slight tendency of reducing the frequency of the two letter words – which are especially, prepositions and conjunctions.

8. The frequency of the letters

The order of the letter	Letter	The average % of the frequency of the letter	The average % of vowels	The average % of cons
1	E	11.994%		
2	I	10.166%		
3	A	8.406%		
4	R	7.680%		
5	N	6.407%		
6	U	6.347%		
7	T	5.792%		
8	L	5.237%		
9	C	5.143%	46.865%	
10	S	4.220%		
11	O	3.699%		
12	P	3.451%		
13	Ă	3.417%		53.135%
14	M	3.178%		
15	D	2.981%		
16	Î	2.828%		
17	V	1.435%		
18	G	1.48%		
19	B	1.358%		
20	Ș	1.281%		
21	F	1.179%		
22	Z	0.846%		
23	Ț	0.803%		
24	H	0.496%		
25	J	0.196%		
26	X	0.034%		
27	Ă	0.008%		
28-31	K	0.000%		
28-31	Q	0.000%		
28-31	Y	0.000%		
28-31	W	0.000%		

CONJECTURE 8. In the recent Romanian poetry the percentage of vowels is, on average, over 45% of the total of letters.

Explanation: in the ordinary language the percentage of vowels is 42.7% (see [1]). In poetry it is greater because:

- vowels are more "musical" than consonants; therefore the words with more vowels "seem" more poetical; words with many vowels confer a special sonority to the text;

- modern poets and poetry are more preoccupied by form than by content, so that more attention is given to expression; the form may prejudice the content, because, very often, the reader is "caught" by sonority and less by essence;

- the internal rhythm of poetry, usually absent in the ordinary language, is also conditioned, partially, by a greater number of vowels;

- rhyme, when used, also favors a greater percentage of vowels. The percentage of vowels was greater in the period of classicism of poetry when the rhythm and rhyme were more frequently used. The special requirements of poetry impose a thorough filtration of the ordinary language.

Given the frequency of the letters in the Romanian language [1] in general:

1. E	5. N	9. L	13. D	17. S	21. F	25. J
2. I	6. T	10. S	14. P	18. B	22. T	26. X
3. A	7.T	11. O	15. M	19. V	23. Z	27. K
4. R	8. C	12. A	16. I	20. G	24. H	

we may calculate the deviation of this volume of verses from the ordinary language:

$$\alpha(v) = \frac{1}{27}\sum_{i=1}^{27}|\alpha(A_i)| \approx 0.741$$

where $\alpha(A_i)$ is the deviation of the letter A_i, $1 \le i \le 27$.

The informational energy, according to O. Onicescu, is

$$\varepsilon(v) = \sum_{i=1}^{27} p_i^2 \approx 0.064,$$

where p_i, $1 \le i \le 27$, is the probability that the letter p_i may appear in the volume (see [1]).

The first order entropy of the volume (according to Shannon) is:

$$H_1(v) = -\frac{1}{\log_{10}2} \cdot \sum_{i=1}^{27} p_i \log_{10} p_i \approx 4.222.$$

9. The themes of the volume are studied by determining the recurrent elements, those that seem to obsess the poet. We will call these elements "key-words" and they are, in order: nouns, verbs, adjectives. Their frequency in the volume is studied. The more frequent words are all included in common notional spheres that will "decode" the themes dealt with by the poet in the volume under study, i.e.:

Elements of the Nature

trees   leaves
water

birds
street
wind
soft-hearted
to ripe
        tender

Cosmological Elements

            sky
Moon
Sun
stars   light
night evening
hours
            time

Existence Elements

life
death
        to live
spirit   tears
to put out
to love
to lower
to deplore
    not to be

Poet's condition

poem

    to read
to dream

        to write
verses
words

These 33 key-words (together with their synonyms) confer certain pastoral note (this was noticed by Constantin Matei, the newspaper "Înainte", Craiova), cosmological (Constantin M. Popa), existentialist nuances (Aureliu Goci, "Luceafarul", Bucharest); the preoccupation of the poet for the condition of the poet and society (Ion Pachia Tatomirescu, Craiova) is also revealed by the frequent use of certain suggestive words.

Of all the words, 33 key-words together with their synonyms have the greatest frequency in the volume.

10. The frequency of words and phrases strongly deviated from the "normal", i.e. the rules of the literary language are about 1.980 of the total of words. (We mean expressions like: "state of self", "very near myself", "it is raining at plus infinite" or words like "nontime", etc. (see [3], pp. 9, 29, 40, 31).

CONJECTURE 9. In the recent Romanian poetry the percentage of words and phrases that strongly deviated from the "normal" of the ordinary language, as well as the rules of the literary language, is slightly over 1. This fact may be accounted for by:
- content seems less important; poets are more concerned with form;
- poets invent words and expressions to be able to better reveal their feelings and emotions;
- the association of antonyms may give birth to constructions that, somehow "violate" the normal;
- poetry is, in fact, destined to break the rules and rebel against the ordinary fact (if, this right is denied, any newspaper article could be called poetry).

"In art" said Voltaire, "rules are only meant to be broken".

In its evaluation from classicism to modernism the percentage of such abnormal words and constructions increased, starting, in fact from zero. Modern literary currents favor the appearance of them.

REFERENCES

[1]   Marcus, Solomon – "Poetica matematică" – Ed. Academiei, Bucharest, 1970 (translated into German, Athenäum, Frankfurt-am-Mein, 1973).
[2]   Marcus, Solomon – Edmond Nicolau – Sorin Stati – "Introducere în lingvistica matematică", Bucharest, 1966 (translated in Italian, Patron, Bologna, 1971 and in Spanish, Teide, Barcelona, 1978).
[3]   Florentin, Ovidiu – "Formule pentru spirit", Ed. Litera, Bucharest, 1981 (Translated in French, les Editions Express, Fès, Maroc, 1983); modern poems.
[4]   Smarandache, Florentin – "A mathematical linguistic approach to Rebus" – article published in "Revue roumain de linguistique", tome XXVIII, 1983, collection "Cahiers de linguistique théorique et appliquée", tome XX, 1983, No. 1, pp. 67-76.

[Editions Scientifiques, Casablanca, Morocco, 1984]