

USSR Academy of Sciences  
Automatics and Telemechanics  
2  
Moscow 1978  
Modeling Behavior and Intelligence  
UDC 681.3.06:82-1

On a Program for Composing Verse  
Boris Katz  
(Moscow)

A particular algorithm for “composing” verse with electronic computational help is demonstrated. The corresponding program is described. Illustrations of poems so “composed” are given.

The aim of this research is to determine which minimal means will make it possible to achieve the illusion of thoughtful versification. It was determined that it is possible to achieve such illusion by very minimal means. To put it simply, a poem consists of grammar + metrics + rhyme + semantics. If a poem possesses metrics and rhyme, it happens that semantics = grammar + the imagination of the reader. This is a very interesting point, but it relates more to the field of psychology. The quality of poems made by the program, using of only the first three components of versification, was outstanding.

#### 1. Thesaurus

The computer is loaded with a thesaurus consisting of several hundred words of various number, gender, and tense. Let us suppose that one line of a poem includes one subject, one predicate, some attributes, and some adverbial modifiers. (Any of these parts of the sentence may be omitted.) According to this, the thesaurus consists of four sections. Nouns and pronouns are in the first section. They act as subjects and can be of any gender and number. Adjectives and possessive pronouns are in the second section. They act as attributes and can be of any number and gender as well. Verbs are in the third section. They act as predicates and can be of different tenses, numbers, and genders. It should be remarked that the program uses only intransitive verbs to make the process of creating verse easier. Adverbs, and nouns with prepositions, are in the fourth section; they serve as adverbial modifiers (of place, time, and manner).

The thesaurus so described consists of words taken from a collection of poems by Osip Mandelstam, *Kamen* (*Stone*, Petrograd: Hyperborey, 1916). Experiments with other lexicons showed that the particular vocabulary selected did not affect the quality of verses that much – only the “mood” and the “themes” were changed. Each word is accompanied by information about metrics, rhyme, and grammar. Metrical information is labelled in the following way: – the number of syllables in a word before the stress  $S^{\text{before}}$  and the number of syllables after the stress  $S^{\text{after}}$  (so the total amount of syllables in a word  $S = S^{\text{before}} + S^{\text{after}} + 1$  syllable).

Information about grammar and rhyme is analyzed in the following parts of this paper.

## 2. Grammatical Information

Each word of the thesaurus is accompanied by information about its possible function in a sentence (as a subject, predicate, attribute, or adverbial modifier). Moreover, information about gender, number, and tense is also given.

The gender of a word may be of four kinds – masculine, feminine, neuter, and irrelevant (later on we will identify any irrelevant setting, for any criterion, as zero). For instance, the gender of verbs in present tense is irrelevant: he flies, she flies, it flies (он летит, она летит, оно летит). The number of a word can be singular, plural, or irrelevant. Tense of a word can be past, present, future, and zero. We should remark that for a word from the first and second sections (nouns, adjectives, and pronouns) the tense is irrelevant, in case these words are in plural, gender is also irrelevant. For all verbs in present and future tense and for all verbs in plural gender is zero.

For words in the fourth section which serve in a sentence as modifiers, the categories of gender, number, and tense are irrelevant. The exceptions to this are modifiers of time; in this case the tense of a word can be nonzero.

## 3. Rhyme

Information about rhyme is represented for words with stress on the last syllable (masculine rhyme) and with stress on the penultimate syllable (feminine rhyme). Poetry cannot yet be computationally produced with dactylic endings, where after the last stressed syllable there are two unstressed ones.

a. Masculine rhyme. Two words with stresses on their last syllables are thought to be rhymed if it is possible to match their last syllables using table 1 and 2.

Table 1

Vowels

01 а (a), я (ja)

02 о (o), ё (jo)

03 у (ou), ю (ju)

04 ы (y), и (i)

05 э (e), е (je)

## Table 2

### Consonants

06 б, п (b, p)

07 в, ф (v, f)

08 г, к (g, k)

09 д, т (d, t)

10 ж /ь/, ш /ь/ (zh, sh)

11 з, с (z, s)

12 ц (ts), тс (ts)

13 л (l)

14 м (m)

15 н (n)

16 р (r)

17 х (kh)

18 ч /ь/ (ch), щ (shch)

19 й (y)

20 бь, пь (b', p')

21 вь, фь (v', f')

22 дь, ть (d', t')

23 сь, зь (s', z')

24 ль (l')

25 нь (n')

26 рь (r')

27 сть (st')

b. Feminine rhyme. Every word with the stress on the penultimate syllable is compared with the set of numbers (k, m, n) in the following way: k is the number of the stressed vowel from the table 1; m is the number of the consonant from table 2\* (numbers from 06 to 18), which goes after the stressed vowel or a number corresponding to a combination of consonants from the table 3. n is the number of the ending category for the given word.

At first unstressed word endings were divided into 4 categories. In the first one there were the endings consisting of a vowel а, я (ya), о, ё(yo) (паутина / pautina) and, possibly, consonants which go after it. The second category consisted of endings which start with the vowels е, и, у, ю (улей/uley). Category 3 corresponds to words ending with vowel ы (точеньй/tocheniy). In the fourth category there are words where the soft sign (ь) appears before a vowel in an unstressed ending (очертанья / ochertan'ya).

\* If a word does not have a consonant after a stressed vowel (краснеет / krasejet), then m is set to 0.

## Table 3

### The combinations of the consonants

28 нк (nk)

29 сл (sl), стл (stl), тл (tl)

30 ст (st)

- 31 рств (rstv)
- 32 лк (lk), льк (l'k), лг (lg)
- 33 зв (zv)
- 34 тр (tr), втр (vtr), др (dr)
- 35 рт (rt), рд (rd)
- 36 нн (nn)
- 37 дн (dn), тн (tn)
- 38 вн (vn)
- 39 зн (zn), сн (sn), рзн (rzn), здн (zdn), стн (stn)
- 40 ртс (rts), рдтс (rdts), льтс (l'ts)
- 41 лн (ln), льн (l'n)
- 42 вш (vsh)
- 43 мн (mn)
- 44 тк (tk), дк (dk), гк (gk)
- 45 рн (rn)
- 46 жн (zhn), шн (shn)
- 47 чн (chn)
- 48 юж (kzh)
- 49 бк (bk), пк (pk)
- 50 лс (ls), лз (lz)
- 51 кл (kl), лкл (lkl)
- 52 пл (pl)
- 53 пн (pn)
- 54 ск (sk), зк (zk), вск (vsk)
- 55 тв (tv)
- 56 бр (br)
- 57 лч (lch), льч (l'ch), льж (l'z)
- 58 кт (kt)
- 59 рк (rk)
- 60 вк (vk), вг (vg), фк (fk)
- 61 кхн (khn)

Subsequently, however, in order to "modernize" the poem and to enrich the rhyme, the first three categories of endings were combined into a single category. This way, in the final version of the program, the machine distinguishes only two categories of the unstressed endings of a word.

Two words with stress on the penultimate syllable are considered to be rhymed if the corresponding triples (k, m, n) coincide. Consider, for instance, a well-known stanza:

Как вы, я - часть великого  
 Перемещенья сроков,  
 И я приму ваш приговор  
 Без гнева и упрека.

Here in the words “сроков / srokov” and “упрека / upryoka” we obtain from table 1  $k=2$ , from table 2  $m=8$ , and, judging by the vowels in the endings,  $n=1$ .

#### 4. The Specification

Before starting to work the machine receives a specification consisting of the same number of lines the operator wishes to be in every stanza of the poem to be produced. In the specification it is stated which ending (masculine or feminine) is present in this line and what line it rhymes with. To be more specific, we label the feminine rhymes with capital letters and the masculine rhymes with lowercase ones. Then a stanza of four lines can be, for example, defined with plain rhyme (AAbb), alternate (AbAb), and enclosed rhyme (AbbA). Moreover, in every line of the specification a quantity of syllables  $S$  is stated for the corresponding line of the poem. If we want, for example, to get the “Onegin” stanza, we should define: aBaB ccDD eFFe GG and note that in the lines with feminine rhyme there are nine syllables, but in the lines with masculine rhyme there are eight of them.

#### 5. Starting the Process: Searching for Rhymes

Despite the fact that the machine writes poems in Russian, it composes every line from right to left. Firstly, it writes the last two words in each line of one stanza, with the number of lines and the manner of rhyming defined by the specification. The procedure is as follows. With the help of a random variable, a word is taken at random from the dictionary with the accent on the last syllable or the penultimate one according to the end of the line mentioned in the specification (masculine or feminine). A rhymed word is selected for it (searching from a random location), and both of them are put in the last place in their respective lines. Note that rhyming a word with the same one is prohibited, so if there are no rhymes to any word taken at random in the dictionary, we should drop it, then take another random word and re-start the search for rhymes. This way, the ends of every line of the stanza will be filled out.

All other words in each row are assigned sequentially from the left. They also are taken at random from the dictionary (using the procedure of “a random word”), but they can be approved or rejected depending upon metrical and grammatical reasons, which will be described next.

#### 6. Metrical Analysis

Let  $S_1^{\text{before}}$ ,  $S_1^{\text{after}}$  be the number of syllables before and after the stress in the word most recently added (for the first step, as the line is built from right to left, this word is very last word in the line). We will check for a random word with the parameters ( $S_2^{\text{before}}$ ,  $S_2^{\text{after}}$ ). First of all, we'll check that after adding ( $S_2^{\text{before}} + S_2^{\text{after}} + 1$ ) that the number of syllables in the resultant phase won't exceed the total number  $S$  defined by the specification for the current line. Otherwise we should immediately reject this word, and once again turn to a random one. Let's look at the  $S_1^{\text{before}} + S_2^{\text{after}}$  meaning the number of syllables between the two neighboring stresses. If the machine writes two-syllable verse (iambic, trochaic), then the word being considered is approved when the quantity  $S_1^{\text{before}} + S_2^{\text{after}}$  is odd, and then it is sent to the subsequent grammar check. If there are ternary meters (dactylic, amphibrachic, anapestic) in the specification, then the meaning of this should be equal to  $3k-1$ .

If these conditions fail the word is dropped, and the machine goes back to the analysis of new random words. Monosyllabic words can be stressed or unstressed depending upon the situation. Experience shows that if there are more than three letters in a monosyllabic word, then in most cases the word should be put in the stressed location for the poem's euphony. For example:

И падали два башмачка со стуком на пол  
И воск слезами с ночника на платье капал

This fact about monosyllabic words is also specified to the computer.

## 7. Grammar Analysis

As mentioned earlier, every line of a poem written by the machine consists of no more than one subject, one predicate, adverbial modifiers, and objects. Grammar analysis is necessary in order to provide for agreement in gender, number and tense. To describe this process we will introduce the parameters of words alongside the parameters of a line. We will attribute to them "the gender of the line," "the number of the line," "the tense of the line.". These parameters will take the same meanings as the corresponding parameters of the words. Before starting the process all these meanings will be adjusted to zero.

One should remember that the composition of a line starts with the choice of the last word. Then the gender, number, and tense of this word mentioned in the information are assigned to the corresponding line. Now the verification of each consequent word – each candidate for the line of verse – is realized by means of comparing the three parameters of the line with the three corresponding parameters of the word. In this case non-zero meaning of the parameters of the word should correspond to non-zero meaning of the line; for instance, the gender is still irrelevant, while the gender of the analyzed word is not irrelevant, for example masculine. Then this word is approved and the zero of the gender of the line is changed to the corresponding parameter (masculine gender) of the word approved.

In the case of misfitting the parameter values, the word is rejected and the machine addresses "the random word" procedure again.

A word that successfully passes metrical and grammatical examination is accepted in the line and placed on the left. The process is continued until the line has been filled (all S syllables).

Suppose the line has been written so that there is a place for only one syllable, but the machine is rejecting one after another all the random words. In this case there is a time limit for the search, after which a vacant space is occupied by a one-syllable interjection, conjunction or adverb. For example an unstressed space can be occupied by “и” (and) while a stressed space can be taken by “ВНОВЬ” (anew). This simple construction considerably impacted the improvement of the verse and enforced its emotional and semantical connotations.

The author is grateful to Prof. E. M. Landis for numerous meetings and talks without which the work would neither have been started nor completed.