On a Program for Composing Verse

Boris Katz (Moscow)

Translated from Russian by N. Schipakina, N. Tarakanova, N. Fedorova, N. Monfort, and B. Katz

An algorithm for "composing" verse using a computer is introduced. The corresponding program is described. Examples of poems "composed" by the computer 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, indeed. Speaking prosaically, a poem consists of grammar + meter + rhyme + semantics. Our experiments show that if a poem possesses meter and rhyme, then 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 composed by the program, using only the first three components of versification, surpassed all our expectations.

1. Lexicon

The computer is loaded with a lexicon containing several hundred words of various number, gender, and tense. Let us assume that each line of a poem includes one subject, one predicate, several modifiers, and several adverbial and/or prepositional phrases. (Any of these parts of the sentence are optional and may be omitted.)

Accordingly, the lexicon 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. Note that the program uses only intransitive verbs to make the process of creating verse easier. Adverbs and prepositional phrases are in the fourth section; they serve as modifiers (of place, time, and manner).

Most of the words in the lexicon described here are taken from a collection of poems by Osip Mandelstam, *Kamen (Stone,* Petrograd: Giperborei, 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 in the lexicon is accompanied by information about meter, rhyme, and grammar. Metrical information is in the form of two numbers – the number of syllables in a word before the stressed syllable S^{before} and the number of syllables after the stress S^{after} (so the total number of syllables in a word $S = S^{before} + S^{after} + 1$).

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

2. Grammatical Information

Each word of the lexicon is accompanied by information about its possible function in a sentence (as a subject, predicate, modifier, or adverbial/prepositional phrase). In addition, information about gender, number, and tense is also specified.

The gender of a word takes on 4 different values – masculine, feminine, neuter, and unassigned (we will sometimes refer to the unassigned value as null-value). For instance, the gender of verbs in present tense is unassigned: he flies, she flies, it lies (он летит, она летит, оно летит).

The number of a word can be singular, plural, or unassigned. The word's tense takes on four values: past, present, future, and unassigned.

Note that for words from the first and second sections of the lexicon (nouns, adjectives, pronouns) the tense is unassigned; when such a word is plural, its gender is also unassigned. In addition, gender takes on null-value for all verbs of present and future tense and for all plural verbs.

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

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). Our computer has not yet been taught to compose poetry 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 assumed to be rhyming if it's possible to match their last syllables using Tables 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, тс (ts, ts)
13	л (l)
14	м (m)
15	н (n)
16	р (r)

17	х (kh)
18	ч /ь/, щ (ch, 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 assigned a set of numbers (*k, m, n*) in the following way: *k* is the number of the stressed vowel from Table 1. *m* is the number of the consonant (following the stressed vowel) from Table 2[1] (numbers from 06 to 18) or the number of the combination of consonants from Table 3. *n* is the type of ending of the given word.

During the initial experiments, the system distinguished 4 categories of unstressed word endings. The first category included endings consisting of a vowel *а*, *я* (ya), *о*, *ё* (yo) (паутина / pautina) possibly followed by consonants. The second category consisted of endings which start with the vowels *е, и, у, ю* (улей / uley). Category 3 corresponds to endings containing the vowel *ы* (точеный/tocheniy). In the fourth category there are words where the soft sign *ь* appears before a vowel in an unstressed ending (очертанья / ochertan′ya).

[1] If a word doesn't have a consonant after a stressed vowel (краснеет / krasnejet), *m* is set to 0.

Table 3: Combinations of consonants[2]

28	нк (nk)	45	рн (rn)
29	сл (sl), стл (stl), тл (tl)	46	жн (zhn), шн (shn)
30	ст (st)	47	чн (chn)
31	рств (rstv)	48	кж (kzh)
32	лк (lk), льк (l′k), лг (lg)	49	бк (bk), пк (pk)
33	зв (zv)	50	лс (ls), лз (lz)
34	тр (tr), втр (vtr), др (dr)	51	кл (kl), лкл (lkl)
35	рт (rt), рд (rd)	52	пл (pl)
36	нн (nn)	53	пн (pn)
37	дн (dn), тн (tn)	54	ск (sk), зк (zk), вск (vsk)
38	вн (vn)	55	тв (tv)
39	зн (zn), сн (sn), рзн (rzn), здн (zdn), стн (stn)	56	бр (br)
40	ртс (rts), рдтс (rdts), льтс (l′ts)	57	лч (lch), льч (l′ch), льж (l′z)
41	лн (ln), льн (l′n)	58	кт (kt)
42	вш (vsh)	59	рк (rk)
43	мн (mn)	60	вк (vk), вг (vg), фк (fk)
44	тк (tk), дк (dk), гк (gk)	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. Thus, in the final version of the program, the computer distinguishes only two categories of the unstressed endings of a word.

Two words with stress on the penultimate syllable are considered to rhyme if the corresponding triples (*k, m, n*) match.

For an example, let's take a look at a well-known stanza[3]:

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

Here for the words "сроков / srokov" and "упрека / upryoka" we have: from Table 1 k=2, from Table 2 m=8, and, based on the vowels in the endings, n=1.

4. Task Specification

Before starting to work the computer receives a task specification indicating the number of lines in every stanza of the poem to be produced. In addition, the specification determines which ending (masculine or feminine) each line of the poem should have and what line it should rhyme with. To be more precise, let us denote the feminine rhymes with capital letters and the masculine rhymes with lowercase ones. Then a stanza of four lines can be, for example, specified with a plain rhyme scheme (AAbb), alternate (AbAb), or enclosed (AbbA). In addition, every line in the task specification indicates the number of syllables S for every line of the poem. If we wanted, for example, to generate the "Onegin" stanza[4], we should define: aBaB ccDD eFFe GG and note that

[2] The Russian language of course allows for other consonant combinations; Table 3 only lists those appearing in the selected lexicon. To expand the lexicon with additional combinations, one only needs to assign any new combination its own number.
[3] From Boris Pasternak's epic poem *Lieutenant Schmidt* (1926).
[4] Onegin stanza refers to the verse form used by Alexander Pushkin in *Eugine Onegin*.

the lines with feminine rhyme there contain nine syllables, and the lines with masculine rhyme contain eight syllables.

5. Starting the Process: Searching for Rhymes

Despite the fact that our computer writes poems in Russian, it composes every line from right to left. It begins by generating the last word in each line of one stanza, where the number of lines and the manner of rhyming is defined by the task specification. This is done in the following manner. Depending on the ending (masculine or feminine) provided by the task specification, a word is taken at random from the dictionary with the stress on the last or penultimate syllable. A word that rhymes with it is selected (searching from a random location), and both are put in the "final position" in their respective lines. Note that rhyming a word with itself is prohibited, so if a rhyming word is not found for a word taken at random from the dictionary, it will be abandoned, another random word will be chosen, and the search for a rhyming word will resume. In such a way, the final positions of every line of the stanza will eventually be filled out.

All other words in each line of the poem are inserted sequentially from the left. They are also taken at random from the dictionary but they can be approved or rejected depending upon metrical and grammatical reasons, which will be described next.

6. Metrical Analysis

Let S_1^{before} and S_1^{after} be the number of syllables before and after the stressed syllable in the word most recently added at the left (during the first step, because the line is being created from right to left, this word is the very last word in the line). The next step is to verify whether a random word with parameters (S_2^{before}, S_2^{after}) is appropriate for the poem being composed. First, we need to make sure that after adding ($S_2^{before} + S_2^{after} + 1$), the number of syllables in the resultant phrase won't exceed the total number of syllables S specified for the current line. Otherwise, this word will be rejected, and another word taken at random will be considered. Let's now examine the value $S_1^{before} + S_2^{after}$, which indicates the number of syllables between the two neighboring stresses. If the computer is in the process of writing two-syllable verse (iambic, trochaic), then the word being considered is approved only when the sum $S_1^{before} + S_2^{after}$ is odd; after that, the word undergoes a subsequent grammar check. If a ternary meter (dactylic, amphibrachic, anapestic) is specified, then the value of this expression should be equal to $3k$-1.

If these conditions fail, the word is removed from consideration, and the computer proceeds to the analysis of new random words. Monosyllabic words can be stressed or unstressed depending on the situation. Our experiments indicate that if a monosyllabic word consists of more than three letters, then in most cases the word should be put in the stressed location for the sake of euphony. For example[5]:

> И падали <u>два</u> башмачка со стуком на <u>пол</u>
> И <u>воск</u> слезами с ночника на платье капал

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

7. Grammatical Analysis

As was mentioned earlier, every line of a poem written by the machine consists of no more than one subject, one predicate, as well as modifiers and adverbial and prepositional phrases. Grammatical analysis is necessary in order to ensure agreement in gender, number and tense. To

[5] From Boris Pasternak's poem "*Winter Night"* (1946).

describe this process we will introduce the grammatical attributes of a line alongside the grammatical attributes of a word. We will call them “line gender,” “line number,” and “line tense”. These attributes will take on the same values as the corresponding word attributes. At the beginning of the verse generation process the values of line attributes are set to zero.

Let’s remember that the composition of each line starts with the selection of the last word. Then the gender, number, and tense of this word are assigned to the corresponding attributes of the line being composed. Now the validation of each subsequent word – each candidate for inclusion in the line of verse – is accomplished by comparing the three attributes of the line with the three corresponding attributes of the word under consideration. In particular, each non-null value of the attributes of the word should match the non-null value of the attributes of the line. A null value of an attribute of a candidate word is always allowed. Consider now a case where some non-null values of the line match the corresponding values of the candidate word, but one of the attributes of the line – for instance, gender – is still unassigned, while the gender of the analyzed candidate word is not null, for example masculine. In this case, the word is approved and the null-value of line gender is replaced by the corresponding value (masculine) of the approved word.

Whenever a mismatch occurs between the value of any attribute of the word and that of the line, the word is rejected and the computer selects another random word.

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

Suppose however that the line is almost completed and there is a place for only one syllable, but the computer keeps rejecting the candidates one after the other. In this case, after a time limit for a search is reached, the vacant position is filled with a one-syllable interjection, conjunction or adverb. For example, an unstressed position could be filled by “и” (and) while a stressed position could be filled by “вновь” (anew). This simple mechanism improved the quality of the verse and strengthened its emotional impact.

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

Submitted for publication on March 9, 1977.

Appendix

Examples of poems composed by the computer:

1. В тумане сумерки остались,
   И мой краснеет человек,
   Над лесом ели зимовали,
   И дикий, мелкий, щедрый снег.
   И шепчется волна морская,
   Томится тусклое тепло,
   Влюбилась робкая, немая,
   И шелестит еще стекло.

2. Быстрей – как будто тень спала –
   Высокий месяц.
   И паутина умерла
   Зимой над лесом.

3. Умирающий – в смятенья.
   Вновь, как тень, огни дрожат,
   Вновь над бездною движенья –
   Где-то далека душа…
   Крик смертельный рядом, зыбкий,
   Тлели в хрустале глаза,
   Шелест мечется с улыбкой,
   Где-то в чаще небеса.

4. И позади цветок сереет,
И ясноглазые белеют,
Отточенный спокойно лед.
Как тень, прозрачно утро славит,
И зыбкий музыкант лукавит:
Прозрачно-бледная придет.

5. Ветви хрупкие светились,
Нежный танцевал, влюблен,
Краски за рекой кружились.
Осмелеет перезвон,
Серая земля взовьется –
Вновь над лесом голоса.
Боль в темнице не вернется,
Спорят сладкие глаза.

6. Уходит неизбежный мальчик черный,
Зимой летит вуаль,
И милый воздух, гробовой, притворный,
Надменная печаль.

7. Вновь шипит над лесом птица,
Гулкий грянул звук небрежно!
В чаще пустота боится,
Бедная белеет нежность.

8. И перламутровый узор
Точеный бледный
Над озером ответил взор
Печальный бедный.

9. Здесь нежная птица,
Здесь с нежностью взор,
Он также струится
Мой странный узор.
Быстрей пламенеет,
В руках тяжелеет…

10. И тайно тишина гремит
И в темном небе – устыдись!
Зеленый легкий снег горит,
Сады прозрачно обнялись.

11. Над бездною сосуд блуждал
На небе крест – вернись!
И только зыбкий воевал,
И скалы обнялись,
И трепетание бежит,
И как безумный свет!
Надменный мир благодарит
И грянул влажный бред!

12. Добрый реет шелест
Плачет пустота
Слушают качели
И поет беда
Стань покорно горе
Томно тишь летит
И прозрачно море
Тайно шелестит
И бежит земная
Незаметно тень
Медленно лесная
Славит влажный день

13. Вечером бледная в небе планета
Очаровательный рядом прилив
Вновь будет сниться над бездною лето
В жизни внимательный видит отлив

14. В бреду ступают имена,
Земная плачет глубина,
В горах, в темнице, в жизни чары.
Мой в жизни май глухой летит,
Всегда прекрасен, ритм горит,
Лад снежный веселится, старый.

15. И старый небосвод пустеет,
Уйдут закаты, тяжелы,
Быть может, в хрустале белеют
Сегодня ласково стволы?

16. Застыла тихо девушка простая,
Небрежны голоса.
И серая последняя пустая –
Полночные глаза.

17. Добрый воздух равнодушный
Добрый мир иной ненужный
Вновь печальна реет радость
Только в опьяненьи сладость

18. Вновь в кустах горят ресницы
Ветер хрупкий светлый злой
На столе желтели птицы
Взор играет за рекой
Лодка далека краснеет
На закате соловьи
Вновь высокие белеют
Стены вечером твои.

19. Миг настоящий на столе
Наполнит море,
Дрожат минуты в хрустале –
Не хочет горе…
Горят над бездною глаза
И солнечные небеса –
Поет природа.
И целый вспоминает плод
Огромный в опьяненьи лед
Люби, забота!

20. Злой музыкант, дрожа вернись,
Ломая краски!
И сила, также устыдись!
Бред сонный, вязкий…

21. В жизни скалы в опьяненьи
Мир теперь последний в чаще
И в лесах упрек горячий
Также слушает забвенье.