# Enabling the automated identification and analysis of meter and rhyme in Russian verse

Elise Thorsen (enthorsen@gmail.com)
David J. Birnbaum (djbpitt@gmail.com)
http://poetry.obdurodon.org

DH2015: Global Digital Humanities
University of Western Sydney, 2015-07-01

## Outline

- Assumptions
- What we study and why
- How we use computation tools to do it

## Assumptions

- Russian quantitative verse studies are worth doing
  - Andrej Belyj, 1910, *Simvolizm*; Jurij Tynjanov, 1924, *Problema stixotvornogo jazyka*; Viktor Žirmunskij 1925, *Vvedenie v metriku. Teorija stixa*; Kiril Taranovski, 1953, *Ruski dvodelni ritmovi*; Boris Ejxenbaum, 1969, *O poèzii*; Mixail Gasparov, 1984, *Očerk istorii russkogo stixa. Metrika, ritmika, rifma, strofika*
  - Vladimir Nabokov, 1964, *Notes on prosody*; J. Thomas Shaw, 1993, *Pushkin's poetics of the unexpected: The nonrhymed lines in the rhymed poetry and the rhymed lines in the nonrhymed poetry*, Ian K. Lilly, 1995, *The dynamics of Russian verse*
  - Handbooks and textbooks: Boris Unbegaun 1956, Barry Scherr 1986, Michael Wachtel 2004
  - Generative poetics: Morris Halle, Bruce Hayes, Paul Kiparsky
  - Names to watch: James Bailey, Nila Friedberg, Emily Klenin, Barry Scherr, J. Thomas Shaw, Marina Tarlinskaja
- Target corpus is generally regular syllabotonic verse: stanzas, lines, feet, meter, rhyme

## Lexical stress vs metrical ictus

No longer mourn for me when I am dead
Than you shall hear the surly sullen bell
Give warning to the world that I am fled
From this vile world with vilest worms to dwell:
[Shakespeare, Sonnet 71, iambic pentameter]

o x | o x | o o | o o | o x
o x | o x | o x | o x | o x
o x | o o | o x | o o | o x
o x | x x | o x | o x | o x

## Lexical stress vs metrical ictus

- Pyrrhic (o o), spondee (x x), trochaic (x o) substitutions in iambic (o x) verse
- Metrical variation
  - Preserves meter, while preventing poetry from becoming "sing-song"
  - Establishes associations among words and lines
  - Modulates the tempo
  - Draws attention to important moments
  - Adapts international meter to local linguistic properties (stress system, word length)

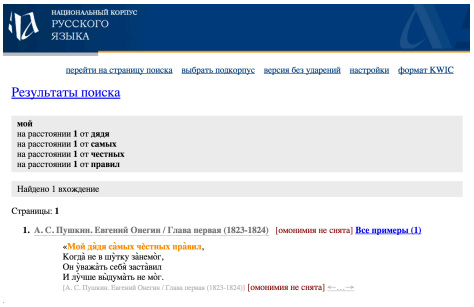
## Meter and language: orthography

- In English
  - The relationship between vowel letters and vowel sounds (syllables) is not one to one
- In Russian
  - Every vowel is syllabic
  - No silent vowels (cf. English Adelaide)
  - No representation of single vowel sounds by sequences of vowel letters (cf. Eng. Adelaide)
- Which means
  - Vowel *letters* in Russian are surrogates for *syllables*

## Meter and language: stress

- English
  - Long words often have secondary stress
- Russian
  - Secondary stress only in compound words: трёхэтажный trëxètažnyj 'three-story'
  - Otherwise Russian words, no matter how long, have only primary stress: достопримечательность dostoprimečatel'nost' '(tourist) attraction'

## Implicit meter and actual stress




## Meter and language: word length

- Average word length in Shakespeare Sonnet 71 is 3.8 letters
  - Lots of short words
- Average word length in first stanza of Pushkin's *Eugene Onegin* in Russian is 9.5 letters
  - Lots of long words
- Neither English nor Russian fits binary meter naturally

## Meter and language: verse convention

- Russian
  - Strong sense of line
  - Strong sense of foot
  - Strong syllabotonic orientation
- English
  - Stronger role for tonic organization

## "The old woman of Berkeley"

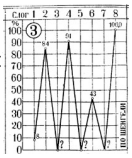
**Robert Southey (1774–1842; 1799)**

| | | |
|---|---|---|
| The ra\|ven croaked \| as she sate \| at her meal, \| | 2 2 3 3 | a |
| And the Old \| Woman knew \| what he said; \| | 3 3 3 | b |
| And she \| grew pale \| at the Ra\|ven's tale, \| | 2 2 3 2 | c |
| And sick\|ened, and went \| to her bed. \| | 2 3 3 | b |

**Vasilij Andreevič Žukovskij (1783–1852; 1814/1831)**

| | | |
|---|---|---|
| На кро\|вле во\|рон ди\|ко про\|кричал —\| | 2 2 2 2 2 | a |
| Стару\|шка слы\|шит и \| бледнеет. | 2 2 2 2+ | B |
| Понят\|но ей, \| что во\|рон тот \|сказал: \| | 2 2 2 2 2 | a |
| Слегла \| в постель, \| дрожит, \|хладеет. | 2 2 2 2+ | B |

## What quantitative metrics tells us about Russian verse




- Final stress must always be realized
- "Law of regressive accentual dissimilat (Taranovski)
  - Pre-final foot is weakest
  - Iambic tetrameter: 2 3 1 4
  - Iambic pentameter: 3 2 4 1 5
- Pattern holds over 18th, 19th, 20th centuries (Friedberg), but with changes
- No such regularity in English (Tarlinskaja)

[image from: N. V. Lapšina, I. K. Romanovič, and B. I. Jarxo, *Metričeskij spravočnik k stixotvorenijam A. S. Puškina*, Moscow: Academia, 1934, p. 134bis. http//feb-web.ru/feb/pushkin/critics/jar/jar-005-.htm]

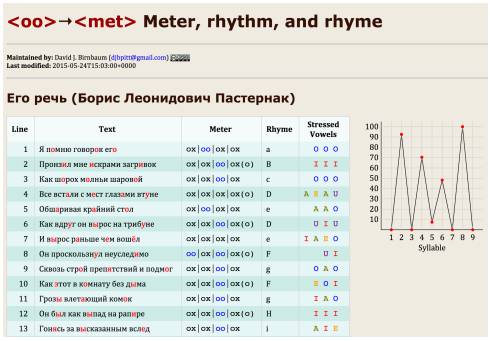
## Methodology

- Knowledge-based
- Output
  - Goals: meter, rhyme (and more?)
  - Sample output
- Assumptions about the input
- Linguistic prerequisites
  - Stress
  - Pronunciation
- Determining meter
- Determining rhyme

## What the system should tell us

- For individual poems:
  - Identify which syllables are stressed linguistically
  - Identify metrical structures and ambient meter
  - Identify deviations from the ambient meter
  - Identify rhyme schemes
  - (Other formal regularities?)
- Corpus level
  - Historical patterns (authors, periods, movements)
  - Relationships between form and meaning
    - E.g., semantic halo

## Sample browsing output



**<oo>→<met> Meter, rhythm, and rhyme**

Maintained by: David J. Birnbaum (djbpitt@gmail.com)
Last modified: 2015-05-24T15:03:00+0000

**Его речь (Борис Леонидович Пастернак)**

| Line | Text | Meter | Rhyme | Stressed Vowels |
|---|---|---|---|---|
| 1 | Я помню говорок его | ox \| oo \| ox \| ox | a | O O O |
| 2 | Произнёс мне искрами загривок | ox \| ox \| oo \| ox(o) | B | I I I |
| 3 | Как шорох молнии шаровой | ox \| ox \| oo \| ox | c | O O O |
| 4 | Все встали с мест глазами втуне | ox \| ox \| ox \| ox(o) | D | A E A U |
| 5 | Обшаривая крайний стол | ox \| oo \| ox \| ox | e | A A O |
| 6 | Как вдруг он вырос на трибуне | ox \| ox \| oo \| ox(o) | D | U I U |
| 7 | И вырос раньше чем вошёл | ox \| ox \| ox \| ox | e | I A E O |
| 8 | Он проскользнул неуследимо | oo \| ox \| oo \| ox(o) | F | U I |
| 9 | Сквозь строй препятствий и подмог | ox \| ox \| oo \| ox | g | O A O |
| 10 | Как этот в комнату без дыма | ox \| ox \| oo \| ox(o) | F | E O I |
| 11 | Грозы влетающий комок | ox \| ox \| oo \| ox | g | I A O |
| 12 | Он был как выпад на рапире | ox \| ox \| oo \| ox(o) | H | I I I |
| 13 | Гонясь за высказанным вслед | ox \| ox \| oo \| ox | i | A I E |



## From plain text input to rich output

- Input must be in native Russian orthography
  - Native Russian orthography almost never marks stress
- Meter
  - Meter depends on stress
- Rhyme
  - Rhyme depends on pronunciation
  - Pronunciation can be inferred from orthography only if stress is also known
- But if we can determine stress automatically …

## Procedure

1. Dictionary lookup
2. Metrical valence
3. Strong and weak position
4. Metrical type (binary ~ ternary)
5. Metrical subtype (foot type)
6. Line length
7. Catalexis and hypermetricality
8. (Rhyme)

## 1. Dictionary lookup

- Input is word in normal orthography
  - Mixed case, punctuation, no stress
  - Dictionary contains word forms with stress and morphological information
    - Morphological information is irrelevant for our purposes
- Output has all vowels tagged
  - Stressed
  - Unstressed
  - Unknown
    - Not in dictionary
    - Dictionary evidence is contradictory

## 2. Metrical valence

- For each vocalic position in the line
- Metrical valence
  - Stressed / (stressed + unstressed)
  - Ignore unknowns
- Varies between 0 and 1

## 3. Strong and weak position

- Compare valence of each position to preceding and following
  - Assume a 0 value if preceding or following is missing, i.e., at beginning or end of a line
- If target value is higher than both neighbors: strong
- If target value is lower than both neighbors: weak
- Otherwise: weak
  - Provisional; adjacent strong positions do not occur in common Russian meter

## 4. Metrical type (binary ~ ternary)

- Calculate how often the strong ~ weak property of a syllable matches the property two (resp. three) syllables earlier
- Count both strong/strong and weak/weak matches
- The greater number of matches determines the type
- Resolve ambiguities according to positional valence (where possible)

## 5. Metrical subtype (foot type)

- Having determined metrical type (binary ~ ternary)
- Subtype is based on last foot
  - Last stress is the only obligatory one
  - Iamb ~ trochee
  - Dactyl ~ anapest ~ amphibrach

## 6. Line length (number of feet)

- Number of strong positions = number of feet
- May be global or line-specific

| | |
|---|---|
| Во всем мне хочется дойти | ○× \| ○× \| ○× \| ○× |
| До самой сути. | ○× \| ○× (○) |
| В работе, в поисках пути, | ○× \| ○× \| ○○ \| ○× |
| В сердечной смуте. | ○× \| ○× (○) |

[Pasternak 1956]

## 7 Catalexis and hypermetricality

- Catalexis: Number of syllables
  - Is sufficient for the number of feet
  - Is not sufficient for the number of complete feet
- Hypermetricality
  - Syllables after the final stress are easily identified
  - Hypermetrical caesura: Demarcate feet based on strong position

| | |
|---|---|
| Как ветер мокрый, ты бьешься в ставни, | ○× \| ○× (○) ‖ ○× \| ○× (○) |
| Как ветер черный, поешь: ты мой! | ○× \| ○× (○) ‖ ○× \| ○× |
| Я древний хаос, я друг твой давний, | ○× \| ○× (○) ‖ ○× \| ○× (○) |
| Твой друг единый,- открой, открой! | ○× \| ○× (○) ‖ ○× \| ○× |

[Gippius, Neljubov', 1907]

## Taking stock

- We can count syllables by counting vowel letters
- If we know the place of stress
  - We get meter
  - We get most pronunciation …
    … and therefore most rhyme
- If we also know e ~ ë
  - We get the rest of pronunciation
  - We also get rhyme

## Thank you!

Elise Thorsen (enthorsen@gmail.com)
David J. Birnbaum (djbpitt@gmail.com)
http://poetry.obdurodon.org

Assisted by: Sam Depretis, Erin Harrington
Thanks to: Elisa Beshero-Bondar, Sibelan Forrester