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
  – Andrey Babiy, 1910, Стихотворения, 1924, Problemy стиховедения
  – Viktor Belyj, 1912, Стихотворения
  – Jurij Tynjanov, 1924, Problemy stikhovvedeniya
  – Viktor Zhirmuinskij, 1925, Vvedenie v strofiku
  – Kirill Tarasovskij, 1953, Russkij stikh
  – Andrei Zhelyabov, 1969, Стихотворения
  – Borís Èjxenbaum, 1969, O poèzi i
  – Mikhail Gasparov, 1984, Očer k istorii russkogo stixa
  – Vladimir Nabokov, 1964, Notes on prosody
  – James Bailey, Nila Friedberg, Emily Kleiman, 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 | 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 ‘three-story’
  - Otherwise Russian words, no matter how long, have only primary stress: достопримечательность ‘(tourist) attraction’

Implicit meter and actual stress

- 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: 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 raven 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 Raven’s tale; |  2 2 3 2  c
And sickened, and went | to her bed. |  2 3 3  b

Vasilij Andrejevič Žukovskij (1783–1852; 1814/1813)

На кро[ле во]рон дил но про[живал] —  2 2 2 2 2  a
Страш[на сълз]и | |  |  |  |  | Бл[еринет].  2 2 2 2 a
Поня[тно ей, | что во]рон тут | с[казал]: |  2 2 2 2 2  a
Следя | в постель, | | |  | хладеет.  2 2 2 2 b

What quantitative metrics tells us about Russian verse

- Final stress must always be realized
- “Law of regressive accentual dissimilation” (Taranovskiy)
  - 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)

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

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
     - lamb ~ trochee
     - Dactyl ~ anapest ~ amphibrach

6. Line length (number of feet)
   - Number of strong positions = number of feet
   - May be global or line-specific

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

[Pasternak 1956]

Во всем мне хочется дойти. ox|ox|ox|ox
Досамой сущи. ox|ox|o
Вработе, впоисках пути. ox|ox|ox|ox
Всердечной-смуте. ox|ox|o

Как ветер мокрый, ты бьешься в стали. ox|ox|ox|ox
Как ветер-черный, поешь: ты мой! ox|ox|ox|ox
Я древний хаос, в друг твой двойный. ox|ox|ox|ox
Твой другой двойный, опий, опий! ox|ox|ox|ox

[Gippius, Nelubov', 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 ~ ê
  – 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