

# Semantic halo of a meter: A keyword-based approach

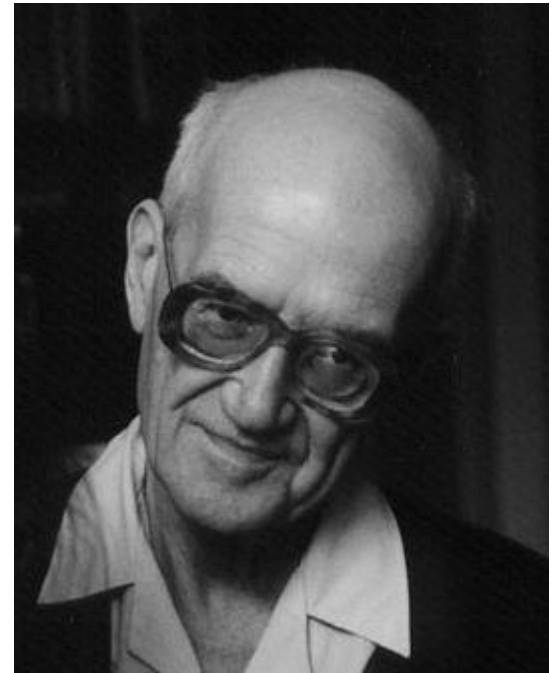
Alexander Piperski  
RSUH / HSE

 **DIALOGUE 2017**

31.05.2017

# Semantic halo of a meter

- Mikhail Gasparov  
(1935–2005)
- “Meter and meaning”  
(1999)



# Gasparov on semantic halo

- Есть детский вопрос, который стесняются задавать литературоведы: почему поэт, начиная стихотворение, берет для него именно такой-то размер, а не иной?

(Gasparov 1999)

# Gasparov on semantic halo

- Первый напрашивающийся ответ будет такой: значит, в звучании каждого размера есть что-то, от природы имеющее ту или иную содержательную окраску.
- Второй напрашивающийся ответ будет такой: значит, в звучании каждого размера есть что-то, по привычке (а не от природы) имеющее ту или иную содержательную окраску.

(Gasparov 1999)

# Михаил Лермонтов. Выхожу один я на дорогу...

*Выхожу один я на дорогу;  
Сквозь туман кремнистый путь блестит;  
Ночь тиха. Пустыня внемлет богу,  
И звезда с звездою говорит.*

*В небесах торжественно и чудно!  
Спит земля в сияньи голубом...  
Что же мне так больно и так трудно?  
Жду ль чего? жалею ли о чем?*

...

# Motifs in X5жм (Gasparov 1999)

- Way
- Night
- Love
- Landscape
- Life & Death

# Фёдор Тютчев. Накануне годовщины 4 августа 1864 г.

*Вот бреду я вдоль большой дороги  
В тихом свете гаснущего дня...  
Тяжело мне, замирают ноги...  
Друг мой милый, видишь ли меня?  
Всё темней, темнее над землею —  
Улетел последний отблеск дня...  
Вот тот мир, где жили мы с тобою,  
Ангел мой, ты видишь ли меня?  
Завтра день молитвы и печали,  
Завтра память рокового дня...  
Ангел мой, где б души ни витали,  
Ангел мой, ты видишь ли меня?*

# Александр Блок. Осенняя воля

*Выхожу я в путь, открытый взорам,  
Ветер гнет упругие кусты,  
Битый камень лег по косогорам,  
Желтой глины скудные пласты.*

*Разгулялась осень в мокрых долах,  
Обнажила кладбища земли,  
Но густых рябин в проезжих селах  
Красный цвет зареет издали.*

...



# Semantic halo and corpus linguistics

- If a meter is associated with a certain topic, we expect poems in this meter to contain words related to this topic
- **Keyword extraction**

# Keyword extraction in corpus linguistics

- Keywords = words that are more frequent in the focus corpus than in the reference corpus
- Numerous measures of keyness proposed:
  - $\chi^2$
  - log-likelihood
  - ...

# Add-N keyword extraction

- Kilgarriff (2009)
- Simplest idea:  $K(w) = f_{fc}(w) / f_{rc}(w)$ 
  - Unsuitable if  $f_{rc}(w) = 0$
  - Strongly favors effect size over statistical significance
- $K = (f_{fc}(w) + N) / (f_{rc}(w) + N)$ 
  - No problem with  $f_{rc}(w) = 0$
  - The choice of N determines the balance between effect size and statistical significance

# Pasternak vs. ruTenTen, N = 0.1

lemma	<i>Pasternak</i>		<i>Russian Web 2011 (ruTenTen11)</i>		Score
	frequency	frequency/mill ?	frequency	frequency/mill	
Спекторский	<u>6</u>	60.4	<u>322</u>	0.0	514.1
отдалены	<u>5</u>	50.3	<u>183</u>	0.0	458.2
чтение	<u>5</u>	50.3	<u>349</u>	0.0	423.3
Недышащей	<u>4</u>	40.2	<u>5</u>	0.0	402.4
впотьмах	<u>12</u>	120.7	<u>3,694</u>	0.2	400.0
оконница	<u>5</u>	50.3	<u>517</u>	0.0	393.0
Клубясь	<u>4</u>	40.2	<u>55</u>	0.0	391.7
Постнику	<u>4</u>	40.2	<u>60</u>	0.0	390.6
узнаны	<u>5</u>	50.3	<u>628</u>	0.0	375.2
смятеньи	<u>4</u>	40.2	<u>248</u>	0.0	355.3
Листвы	<u>4</u>	40.2	<u>360</u>	0.0	337.1
вьюга	<u>34</u>	342.1	<u>17,702</u>	1.0	320.3
зга	<u>9</u>	90.6	<u>3,464</u>	0.2	313.2
зданья	<u>4</u>	40.2	<u>610</u>	0.0	302.5

# Pasternak vs. ruTenTen, N = 1

*Pasternak*

*Russian Web 2011  
(ruTenTen11)*

lemma	frequency	frequency/mill ?	frequency	frequency/mill	Score
вьюга	<a href="#">34</a>	342.1	<a href="#">17,702</a>	1.0	174.3
ль	<a href="#">52</a>	523.2	<a href="#">50,048</a>	2.7	140.2
мгла	<a href="#">42</a>	422.6	<a href="#">42,814</a>	2.3	126.7
зарев	<a href="#">21</a>	211.3	<a href="#">13,492</a>	0.7	122.1
высь	<a href="#">21</a>	211.3	<a href="#">13,502</a>	0.7	122.1
стужа	<a href="#">25</a>	251.5	<a href="#">21,287</a>	1.2	116.7
ширь	<a href="#">17</a>	171.1	<a href="#">9,320</a>	0.5	114.0
даль	<a href="#">53</a>	533.3	<a href="#">72,007</a>	3.9	108.2
средь	<a href="#">31</a>	311.9	<a href="#">35,528</a>	1.9	106.3
ненастье	<a href="#">22</a>	221.4	<a href="#">20,359</a>	1.1	105.2
впотьмах	<a href="#">12</a>	120.7	<a href="#">3,694</a>	0.2	101.3
небосвод	<a href="#">18</a>	181.1	<a href="#">21,387</a>	1.2	83.9
быль	<a href="#">20</a>	201.2	<a href="#">26,363</a>	1.4	82.8
тишь	<a href="#">19</a>	191.2	<a href="#">24,753</a>	1.4	81.6

# Pasternak vs. ruTenTen, N = 10

*Pasternak*

*Russian Web 2011  
(ruTenTen11)*

lemma	frequency	frequency/mill ?	frequency	frequency/mill	Score
ль	<a href="#">52</a>	523.2	<a href="#">50,048</a>	2.7	41.9
даль	<a href="#">53</a>	533.3	<a href="#">72,007</a>	3.9	39.0
туча	<a href="#">69</a>	694.3	<a href="#">162,058</a>	8.9	37.3
мгла	<a href="#">42</a>	422.6	<a href="#">42,814</a>	2.3	35.1
заря	<a href="#">64</a>	644.0	<a href="#">169,152</a>	9.3	34.0
вьюга	<a href="#">34</a>	342.1	<a href="#">17,702</a>	1.0	32.1
тоска	<a href="#">62</a>	623.8	<a href="#">205,110</a>	11.2	29.9
среди	<a href="#">31</a>	311.9	<a href="#">35,528</a>	1.9	27.0
вихрь	<a href="#">35</a>	352.2	<a href="#">81,843</a>	4.5	25.0
гул	<a href="#">33</a>	332.0	<a href="#">77,891</a>	4.3	24.0
стужа	<a href="#">25</a>	251.5	<a href="#">21,287</a>	1.2	23.4
снег	<a href="#">149</a>	1499.2	<a href="#">1,018,848</a>	55.7	23.0
гроза	<a href="#">37</a>	372.3	<a href="#">138,592</a>	7.6	21.7
ливень	<a href="#">31</a>	311.9	<a href="#">88,214</a>	4.8	21.7

# Pasternak vs. ruTenTen, N = 3000

*Pasternak*

*Russian Web 2011  
(ruTenTen11)*

lemma	frequency	frequency/mill ?	frequency	frequency/mill	Score
как	<a href="#">1,500</a>	15092.7	<a href="#">82,759,638</a>	4527.2	2.4
ты	<a href="#">478</a>	4809.5	<a href="#">17,935,027</a>	981.1	2.0
ночь	<a href="#">265</a>	2666.4	<a href="#">3,146,284</a>	172.1	1.8
я	<a href="#">982</a>	9880.7	<a href="#">85,595,981</a>	4682.4	1.7
и	<a href="#">4,268</a>	42943.7	<a href="#">503,894,565</a>	27564.6	1.5
снег	<a href="#">149</a>	1499.2	<a href="#">1,018,848</a>	55.7	1.5
весь	<a href="#">453</a>	4558.0	<a href="#">39,448,601</a>	2158.0	1.5
небо	<a href="#">150</a>	1509.3	<a href="#">1,603,100</a>	87.7	1.5
твой	<a href="#">147</a>	1479.1	<a href="#">2,999,805</a>	164.1	1.4
день	<a href="#">267</a>	2686.5	<a href="#">19,914,270</a>	1089.4	1.4
под	<a href="#">215</a>	2163.3	<a href="#">14,201,345</a>	776.9	1.4
за	<a href="#">468</a>	4708.9	<a href="#">50,033,436</a>	2737.0	1.3
где	<a href="#">192</a>	1931.9	<a href="#">12,799,720</a>	700.2	1.3
мой	<a href="#">199</a>	2002.3	<a href="#">13,826,484</a>	756.4	1.3

# Data

- Poetic subcorpus of the Russian National Corpus (RNC)
- 78,868 metrically annotated poetic texts, 11m tokens
- Only texts containing no more than 500 words were used for this study
- **76,559 poems, 8m tokens** (including punctuation marks), lemmatized with MyStem



# Example of lemmatization

*Поздно. Уж мы обогнули стену,  
Мы проскочили сквозь рощу пальм,  
Через Неву, через Нил и Сену  
Мы прогремели по трем мостам.*

(Николай Гумилёв. Заблудившийся  
трамвай)

*поздно уж мы огибать стена  
мы проскочить сквозь роща пальма  
через нева через нил и сено  
мы прогреметь по три мост*

# Keyword extraction

- $K = (f_{fc}(w) + N) / (f_{rc}(w) + N)$
- $N = 3000$
- Focus corpora: poems in 20 most frequent metrical schemes ( $\geq 50,000$  tokens)
- Reference corpora: all the remaining texts for each of the focus corpora

# Focus corpora

#	Meter	Tokens	#	Meter	Tokens
1.	Я4жм	695,663	11.	Ан3жм	90,658
2.	Х4жм	331,449	12.	Дк3жм	86,034
3.	Я4ж,м	292,724	13.	Я6жжмм	77,957
4.	Я5жм	291,540	14.	Х4ж,м	72,217
5.	Я4мж	219,314	15.	Я3жм	69,813
6.	Я5мж	173,737	16.	Я6мж	65,649
7.	Я6жм	144,418	17.	Я6ж,м	64,109
8.	Я5ж,м	134,355	18.	Х4ж	60,204
9.	Х5жм	125,119	19.	Я4м	55,349
10.	Аф3ж,м	100,252	20.	Я6ммжж	53,793

# Keyword extraction

- $K = (f_{fc}(w) + N) / (f_{rc}(w) + N)$
- $N = 3000$
- Focus corpora: poems in 20 most frequent metrical schemes ( $\geq 50,000$  tokens)
- Reference corpora: all the remaining texts for each of the focus corpora
- Keywords with  $K \geq 1.1$
- Mean: 30 keywords per metrical scheme

# Metrical preferences in various time periods

- **Ябммжжж**
- 106 keywords
- *сей, зреть, отрада, внимать, муза, сатира, счастье, величество, ...*
- Mostly poems from the 18<sup>th</sup> century

# Александр Сумароков. Эпитафия

*Два брата здесь лежат: один во весь свой век  
Был честный, а притом несчастный человек.  
Другой с бездельствами век прожил неразлучно  
И жил по саму смерть свою благополучно.  
Не воздан праведник, без казни умер плут, —  
Конечно, будет нам еще по смерти суд.*

# Linguistics of verse

- Clause boundaries typically coincide with verse boundaries (*enjambements* are possible, but rare)
- Preferences for subordinating conjunctions determined by foot type:
  - Я: *когда, хотя, который*
  - Х: *если, чтобы*

# Motifs in X5жм (Gasparov 1999)

- Way
- Night
- Love
- Landscape
- Life & Death



# X5жм (26 keywords)

ЭТОТ

если

ТОЛЬКО

земля 

СЛОВНО

ГОД

НОЧЬ 

мочь

такой

чтобы

что

СКОЛЬКО

над

звезда  

город

уходить 

синий

ветер 

каждый

путь  

ЗОЛОТОЙ

голубой

по

темный 

жить 

разве

 Way  Night  Love  Landscape  Life & Death

# АфЗжм (14 keywords)

*туман*

*холодный*

*на*

*и*

*далекий*

*а*

*над*

*под*

*шуметь*

*идти*

*за*

*русский*

*Россия*

*она*

# АфЗжм

- Gasparov (1999): elevated intonation, poems about Russia
- *Россия, русский* among the keywords

# АфЗжм

- Gasparov (1999): elevated intonation, poems about Russia
- *Россия, русский* among the keywords

# Владимир Набоков. Родина

*Бессмертное счастье наше  
Россией зовется в веках.*

*Мы края не видели краше,  
а были во многих краях.*

*Но где бы стезя ни бежала,  
нам русская снилась земля.*

*Изгнание, где твое жало,  
чужбина, где сила твоя?*

...

# Андрей Белый. Родине

*Рыдай, буревая стихия,  
В столбах громового огня!  
Россия, Россия, Россия, —  
Безумствуй, сжигая меня!*

...

# Рюрик Ивнев. Смольный

...

*Пусть ветер холодный и резкий  
Ревет и не хочет стихать.  
Меня научил Достоевский  
Россию мою понимать.*

*Не я ли стихами молился,  
Чтоб умер жестокий палач,  
И вот этот круг завершился,  
Россия, Россия, не плачь!*

...

# АнЗжм (25 keywords)

*ЭТОТ*

*НИ*

*простор*

*ТОЛЬКО*

*ЗОЛОТОЙ*

*ДО*

*СЛОВНО*

*ЧТОБЫ*

*НА*

*НАД*

*ПО*

*НЕ*

*ЭТО*

*дорога*

*ОТ*

*молодой*

*за*

*впереди*

*голубой*

*я*

*КТО-ТО*

*земля*

*СКОЛЬКО*

*ТЫ*

*снег*



# АнЗжм (25 keywords)

*ЭТОТ*

*ТОЛЬКО*

*СЛОВНО*

*НАД*

*ЭТО*

*МОЛОДОЙ*

*ГОЛУБОЙ*

*ЗЕМЛЯ *

*ТЫ*

*НИ*

*ЗОЛОТОЙ*

*ЧТОБЫ*

*ПО*

*ДОРОГА *

*ЗА*

*Я*

*СКОЛЬКО*

*СНЕГ *

*ПРОСТОР *

*ДО*

*НА*

*НЕ*

*ОТ*

*ВПЕРЕДИ*

*КТО-ТО*

Иван Елагин. «Всё снега, да  
снега, да метели...»

*Всё снега, да снега, да метели,  
Нелюдимый скалистый простор.  
В горностаевых мантиях ели,  
Как монархи, спускаются с гор.*

*И оленей пугливое стадо  
От дороги уходит в снега.  
Вот какое оно — Колорадо,  
И такая ж, наверно, тайга.*

# X4ж (33 keywords)

*у*

*этот*

*солнце*

*море*

*туча*

*ти*

*ка*

*песня*

*белый*

*ах*

*роза*

*к*

*небо*

*словно*

*слышать*

*птица*

*рыцарь*

*полный*

*девица*

*это*

*лес*

*если*

*с*

*народ*

*сердце*

*посмотреть*

*тройка*

*тихо*

*только*

*птичка*

*отвечать*

*выходить*

*дама*

# X4ж (33 keywords)

*у*

*этот*

*солнце* ♣

*море* ♣

*туча* ♣

*ти*

*ка*

*песня*

*белый*

*ах*

*роза*

*к*

*небо* ♣

*словно*

*слышать*

*птица*

*рыцарь*

*полный*

*девица*

*это*

*лес* ♣

*если*

*с*

*народ*

*сердце*

*посмотреть*

*тройка*

*тихо*

*только*

*птичка*

*отвечать*

*выходить*

*дама*

# Николай Гумилёв. Лето

*Лето было слишком знойно,  
Солнце жгло с небесной кручи, —  
Тяжело и беспокойно,  
Словно львы, бродили тучи.*

...

# Яков Полонский. Финский берег

*Лес да волны — берег дикий,  
А у моря домик бедный.  
Лес шумит; в сырые окна  
Светит солнца призрак бледный.*

...

# Conclusion

- Keywords can be linked to the semantic halo of a meter
- Caveats and questions for further research:
  - Historical developments are ignored
  - Identification of motifs is extremely impressionistic

Thank you  
for your attention!