

ИСТОРИЯ МАТЕМАТИЧЕСКОЙ ЛИНГВИСТИКИ

Сергей
Владимирович
Лесникóв

к. филолог. н. ,
доцент
<http://ГИЗАУРУС.РФ>
<http://МФРЯ.РФ>

serg@lsw.ru



Данная работа выполняется при финансовой поддержке
РФФИ по исследовательскому проекту № 11-07-00733
**«Гипертекстовый информационно-поисковый тезаурус
«Метаязык науки» (структура; математическое,
лингвистическое и программное обеспечения; разделы
лингвистика, математика, экономика)» (научный
руководитель С.В. Лесников)»**

[ЛСВ.РФ, ГИЗАУРУС.РФ, МФРЯ.РФ, agora.guru.ru/conf_iitnio-2011,
www.lsw.ru, www.mmik.ru, moodle.syktsu.ru,
subscribe.ru/catalog/linguistics.kto, subscribe.ru/catalog/linguistics.rusling,
www.famous-scientists.ru/2965, subscribe.ru/science.humanity.hypervault,
umk.lsw.ru].

Ключевые слова: *гизаурус, лингвистика, метаязык, поиск, тезаурус.*

Математическую лингвистику определяют как математическую дисциплину, объектом приложения которой являются естественные языки или как раздел языкоznания, занимающийся изучением возможностей применения математических методов к исследованию языков. В математической лингвистике акцент делается на абстрактные модели, позволяющие отображать феномены языка, а в компьютерной лингвистике – на прикладные методы описания и обработки языка для компьютерных систем.

На стыке и в развитии лингвистик (прикладной, **математической** и структурной) разрабатываются и развиваются: "автоматическая", "вычислительная", "инженерная", "интерпретирующая", "квантитативная", "кибернетическая", "компьютерная", "статистическая", "стратификационная" ЛИНГВИСТИКИ.

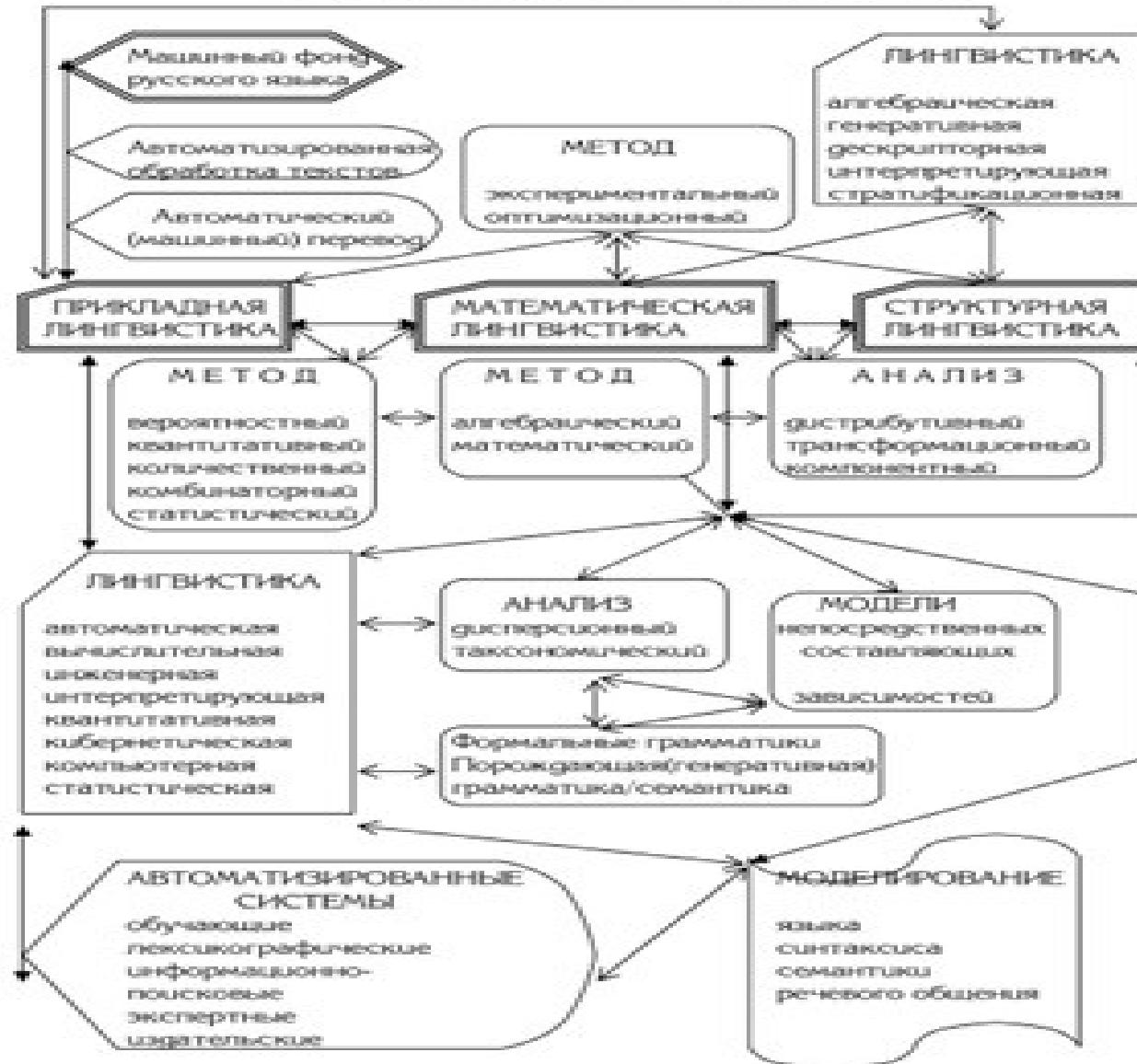
Отдельно можно выделить следующие направления современной лингвистики, в которых предполагается использование ЭВМ:

- "Автоматическая/автоматизированная обработка текстов";
- "Автоматический перевод (машинный перевод)";
- "Дистрибутивный анализ"; — "Дисперсионный анализ";
- "Таксономический анализ", а также разрабатываемые в рамках языкоznания: "вероятностные", "квантитативные", "количественные", "комбинаторные", "статистические" МЕТОДЫ.

В математической лингвистике определенное влияние в целом и в частностях на процесс автоматизации лингвистических исследований своими идеями, методами, теоретическими и практическими подходами и формальными моделями оказали следующие лингвистики: "генеративная", "дескриптивная", "интерпретирующая", а также "теория формальных грамматик", "порождающая (генеративная) грамматика/семантика" и М Е Т О ДЫ:

"алгебраические", "математические",
"непосредственных составляющих", "оптимизационные",
"трансформационные", "экспериментальные".

Рисунок 1. Структура взаимосвязей лингвистик



Термины: автоматическая, вычислительная, инженерная, интерпретирующая, квантиативная, кибернетическая, компьютерная **математическая** и статистическая лингвистики, на наш взгляд, синонимичны и различаются в основном лишь авторством — кто ввел в научный оборот тот или иной термин, претендуя при этом на новый взгляд на соответствующие проблемы и пути их разрешения. В определенной степени различие этих терминов можно частично объяснить также спором по основным теоретическим положениям научных школ и направлений (для примера можно назвать по географическому принципу без конкретизации по персоналиям: Москва, Санкт-Петербург, Тарту, Сибирь, Киев, Минск), хотя и в рамках этих школ не существует однозначности, а ученые руководствуются принципом: каждый ученый вправе иметь и отстаивать свою гипотезу и точку зрения.

системе имеются формально неразрешимые предложения (утверждения, теоремы и т.п.), которые нельзя ни доказать, ни опровергнуть. Суть 1 теоремы сводится к тому, что для любого класса формул (классовая формула – формула с одной переменной), подчиненного некоторым заранее определенным условиям, обязательно найдется формула относительно тривиального вида и такая формула, что ни она, ни её отрицание принципиально невыводимы из этого класса. При этом, т.к. в основе рассматриваемой Гёдели формальной системы были арифметические аксиомы Пеано, то эта 1 теорема Гёделя стала интерпретироваться и соответственно называться “теорема о неполноте формальной арифметики”. В.А. Успенский в своей работе “Теорема Гёделя о неполноте” рассматривает как синтаксическую (устраняя по возможности упоминания об истинности суждений), так и семантическую (для языка арифметики речь идет об истинности и ложности – смысле выражений) формулировки теоремы.

2 теорема Гёделя, являясь следствием первой теоремы показывает, что, если формальная система, содержащая арифметику, непротиворечива, то в данной формальной системе нельзя доказать её собственную непротиворечивость.

Лофти Заде фактически расширил математический аппарат для рассмотрения и изучения ситуаций, настолько сложных и/или неопределенных, что не поддающихся точному (однозначному) описанию, введя в научный оборот понятие «нечёткого» множества и лингвистической переменной. С середины шестидесятых годов Заде фактически возглавил группу ученых из разных стран, которые, как и он, считали возможным преодоление противоречий между строгими математическими моделями (и методами, рассуждениями) и подходами (принципами) решения задач гуманитариями, традиционно отличающихся чрезвычайно низким уровнем (и даже полным отсутствием) достигнутой формализации именно благодаря средствам нечеткой математики.

«Лингвистической» переменной Л. А. Заде называет «переменную, значениями которой являются слова или предложения естественного или искусственного языка. Например, Возраст – лингвистическая переменная, если она принимает лингвистические, а не числовые значения, т.е. значения молодой, не молодой, очень молодой, вполне молодой, старый, не очень старый и не очень молодой и т.п., а не 20, 21, 22 и т.д.»

Л.А.Заде подчеркивал, что уделял «основное внимание развитию понятий, лежащих в основе так называемого *лингвистического подхода* к анализу сложных или плохо определенных систем и процессов принятия решений.

Существенные отличия этого подхода от обычных количественных методов системного анализа рождают много вопросов и проблем, которые являются новыми по своей природе и решение которых требует поэтому большого объема дополнительных исследований и экспериментирования.

Это относится, в частности, к основным аспектам понятия лингвистической переменной, на котором мы лишь коротко остановились в нашем изложении: лингвистическим приближениям, представлению лингвистических неопределенностей, нечисловым базовым переменным, λ - и β -взаимодействиям, нечетким теоремам, лингвистическим вероятностным распределениям, нечетким блок-схемам и др.».

Марвин Минский. Система фреймов для представления знаний.

По мнению Марвина Минского, «основные структурные элементы, образующие фундамент для развертывания процессов восприятия, хранения информации, мышления и разработки языковых форм общения, должны быть более крупными и иметь более четкую структуру; их фактическое и процедуральное содержание следует более тесно увязывать друг с другом с тем, чтобы получить возможность объяснить феномен силы и "быстродействия" человеческого мышления». И именно поэтому Минский в своей работе делает попытку создать единую и стройную теорию, отправным моментом для которой «служит тот факт, что человек, пытаясь познать новую для себя ситуацию или по-новому взглянуть на уже привычные вещи, выбирает из своей памяти некоторую структуру данных (образ), называемую нами фреймом, с таким расчетом, чтобы путем изменения в ней отдельных деталей сделать ее пригодной для понимания более широкого класса явлений или процессов.

В 50-е годы XX века перед учеными встали новые задачи, связанные с громадным увеличением потока информации в ходе происходящей научно-технической революции. Поскольку большая часть информации записывается при помощи естественных языков, то возникает проблема автоматического (математического, компьютерного) решения задач, связанных с языком. С этого времени начинается интенсивное обсуждение следующих задач:

- а) создание ЭВМ, автоматически анализирующих тексты на естественных языках;
- б) создание информационно-логических устройств, в компактной форме запоминающих информацию и быстро выдающих ее по первому требованию;
- в) развитие способов передачи речи по каналам телефонной, телеграфной, радио связи;
- г) создание устройств, воспринимающих устную речь и воспроизводящих ее в письменной форме («автоматические машинистки»);
- д) машинный перевод с одного языка на другой.

Ученые стремились обсуждать эти задачи в контакте с представителями других наук, напр., математики, логики, теории информации, электроники, акустики, физиологии высшей нервной деятельности. Языковеды и математики для решения подобных задач развивали новые методы и представления, допускающие формулировку этих задач на языке машин, т. е. языке формальном - языке математики.

В сентябре 1956 г. на филологическом факультете МГУ начал работать семинар, посвященный некоторым применению математических методов в лингвистике. Основные проблемы были предложены акад. А.Н. Колмогоровым, влияние которого на семинаре, хотя сам он в нем не участвовал, ощущалось очень сильно, ибо руководители семинара: В.А. Успенский и Р.Л. Добрушин были непосредственными учениками Колмогорова.

Отметим работы: - «О последовательности построения системы языка» (1958) П.С. Кузнецов, где впервые была выдвинута идея рекуррентного построения системы языка с возвращением от высших уровней к низшим; - «Значение математических методов в лингвистике» Р.Л. Добрушин, где впервые поставлен вопрос взаимоотношения новых и классических методов; - «О некоторых аналогиях к проблемам и методам современного теоретического языкознания в трудах древнеиндийских грамматиков» В.Н. Топоров; - «Активная и пассивная грамматика Л.В. Щербы» И.И. Ревзин; - «Объективное исследование смысловых связей» А.Р. Лурия, О.С. Виноградова; - «Лингвистические вопросы стихотворного перевода» Вяч.Вс. Иванова и его же «Теорема Геделя и лингвистические парадоксы»; - «Теория вероятности и установление языкового родства» А.Б. Долгопольский (см. Тезисы, 1958). На Совещании по математической лингвистике в Ленинграде (Тезисы, 1959) обсуждалась концепция о сочетании теоретико-множественных (комбинаторных) и статистических представлений при моделировании структуры языка (Днепров 1960).

В конце 80-х гг. XXв. активизировались исследования с применением ЭВМ в рамках проекта по созданию Машинного фонда русского языка. МФ РЯ формировался в виде нескольких систем комплексной автоматизации лингвистических исследований и предусматривал "накопление на машинных носителях всего лексического богатства русского языка, создание фонда лингвистических алгоритмов и программ, фонда полностью завершенных систем автоматического анализа и синтеза русского текста, нескольких информационно-справочных систем по языкоznанию ". Впервые мысль о МФ РЯ высказал академик А.П. Ершов в докладе "К методологии построения диалоговых систем: феномен деловой прозы" 26 сентября 1978 г. на научной конференции "Диалог-78": "Любой прогресс в области построения моделей и алгоритмов останется, однако, академическим упражнением, если не будет решена наиважнейшая задача создания Машинного фонда русского языка".

Математический анализа произведений русских поэтов Золотого и Серебряного веков.

В качестве корпуса текстов для исследования взяты произведения 34 поэтов Золотого и Серебряного века, в отношении подлинности и авторства которых сомнений нет. Золотой век русской поэзии – первая половина 19 века. Серебряный век русской поэзии – период с 1890-х годов по начало 1920-х годов. Невозможно рассматривать одно в отрыве от другого, Серебряный век без наследия Золотого, а Золотой – без продолжения его в Серебряном.

Для формализованного (автоматического, автоматизированного, алгебраического, аналитического, валентностного, вычислительного, дистрибутивного, инженерного, квантиативного, кибернетического, комплексного, компьютерного, количественного, контентного, математического, машинного, механистического, статистического, численного, факторного, цифрового...) анализа текстовой информации надо определиться с базовыми понятиями: что именно и по каким формулам считается

Количественный анализ текста предполагает расчёт ряда некоторых количественных, кванитативных, математических, статистических, численных характеристик корпуса текстов: N – объём текста, равный числу лексических единиц /ЛЕ/ (ЛЕ = словоупотребление, словоформа, текстоформа, лексема) в тексте. N_{letter} – число букв в тексте. $N_{syllable}$ – число слогов в тексте. N_{clause} – число предложений в тексте. n – число абзацев в тексте. m – число абзацев текста, в которых встретилась ЛЕ. L – число ЛЕ, которые встретились в тексте хотя бы один раз. Lf_1 – число ЛЕ, которые встретились в тексте только один раз. Lf_k – число ЛЕ, которые встретились в тексте с частотой больше одного раза. F_i – абсолютная частота ЛЕ. F^*_i – накопленная абсолютная частота ЛЕ, равная сумме частоты данной ЛЕ и всех предшествующих абсолютных частот ЛЕ. r – ранг ЛЕ. Ранг ЛЕ может измеряться следующим образом: а) по частоте встречаемости в тексте (или фрагменте) – самая частотная ЛЕ имеет ранг равный 1 и далее ранг r увеличивается по мере уменьшения частоты встречаемости ЛЕ в тексте (ЛЕ, имеющие одинаковую частоту, имеют и равные ранги); б) по длине слова (напр., число букв в ЛЕ); в) число значений ЛЕ (по толковым словарям). f_{rl} – максимальная частотность ЛЕ. L_{word} – средняя длина слова в буквах (среднее число букв в слове, подсчитанное для каждой выборки). L_{clause} – средняя длина предложений в словах (среднее число слов в предложении, подсчитанное для каждой выборки). L_i – длина i -ого предложения. $L_{syllable}$ – средняя длина слов в слогах (среднее число слогов в слове, подсчитанное для каждой выборки). $L_{syllable3}$ – среднее число "длинных" слов (более трех слогов). $L_{syllable3}i$ – число "длинных" слов в i -ом предложении. L_{empty} – количество служебных слов в предложении (среднее число союзов, предлогов и частиц в предложении для каждой выборки). $F_{n-grams}$ – частота n-грамм, биграмм, триграмм, квадрограмм (т.е. сочетания из n, 2, 3, 4 букв). F_{morph} – частота морфем (приставок, суффиксов). F_{verb} – частота употребления глаголов. $F_{interjection}$ – частота употребления междометий. $F_{adjective}$ – частота употребления прилагательных. F_{empty} – частота употребления служебных слов (предлогов, союзов, частиц). F_{noun} – частота употребления существительных. K_{cons} – консонансный коэффициент, равный отношению числа согласных к числу гласных в слове.

Энтропия – мера неопределенности. Энтропия (от греч. “entropía” – поворот, превращение), понятие, впервые введенное в термодинамике для определения меры необратимого рассеяния энергии. Энтропия широко применяется и в других областях науки. В информатике под энтропией понимается мера неопределённости или непредсказуемости информации, неопределённость появления какого-либо символа первичного алфавита. В лингвистике, например, под первичным алфавитом можно понимать традиционный алфавит, расширенный знаками препинания и специальными разделителями (пробел, разметка, непечатные невидимые на экране символы) или, полный набор ЛЕ (словоформ, текстоформ, лемм и т.п.). При отсутствии информационных потерь энтропия равна количеству информации на символ передаваемого сообщения.

Индексы дистрибуции, итерации, исключительности, предсказуемости, плотности текста характеризуют стилистические особенности текстов при помощи математико-статистических методов

Индекс итерации (индекс повторения ЛЕ в замкнутом тексте): $I_i = \frac{N}{L}$

Индекс исключительности (специфиности) лексики: $I_e = 20 \cdot \frac{Lf_1}{N}$ В поэзии бывает до 50 % “исключительных ” слов, в художественной прозе их намного меньше.

Индекс плотности текста: $I_q = \frac{Lf_k}{N}$ Этот индекс зависит от числа повторяющихся слов в тексте и длины текста. Чем богаче тематика, тем выше I_q . Чем однороднее в тематическом отношении текст, тем I_q ниже. Плотность влияет на процесс восприятия текста, делая этот текст более или менее привлекательным для аудитории и оптимизируя или, напротив, ухудшая его усвоение. Низкая плотность, обеспечиваемая повторами одной и той же идеи и обстоятельной аргументацией, представляющей одну и ту же ситуацию в различных аспектах, облегчает смысловое восприятие текста.

Индекс дистрибуции: $I_c = (f_{r1}^2 + L^2)^{\frac{1}{2}}$ Чем больше значение индекса дистрибуции, тем богаче словарь текста.

Индекс предсказуемости: $I_p = 100 - \frac{100 \cdot L f_1}{N}$ Чем ниже индекс предсказуемости текста, тем привлекательнее текст.

I_{ext} – объем экстенсивности словаря текста. Пропорционален широте лексики, разнообразию выражения.

I_f – индекс стереотипности текста. Длина интервала средней части повторяющихся ЛЕ. Если I_f больше, то главное не форма, а содержание (для беглого нестилизованного чтения, спонтанная речь). I_f меньше у художественных текстов, беллетристики.

Относительная частота ЛЕ: $f_i = \frac{F_i}{N}$

Накопленная относительная частота ЛЕ: $f_i^* = \frac{F_i^*}{N}$

Коэффициент важности ЛЕ: $K_i = \frac{F_i * m}{n * N}$

Удельная энтропия ЛЕ: $H_i = -f_i \log_2 f_i$

Накопленная энтропия текста: $H_k^* = -\sum_{i=1}^k f_i \log_2 f_i$

В результате математических расчетов значений накопленной энтропии, индексов итерации, исключительности, плотности, дистрибуции и предсказуемости, и представив результаты в графической форме, была произведена оценка общего вклада поэтов в русский язык и литературу, в культуру России. Подсчитав общий итог и выделив первых пять поэтов в рейтинге (см. диаграмму 7), мы выделили их красным цветом на диаграммах 1-6, посвященных количественным характеристикам произведений 34 поэтов.

В итоге констатируем, что

1) Блок А.А., 2) Пушкин А.С., 3) Некрасов Н.А., 4) Лермонтов М.Ю., 5) Жуковский В.А. выделились на фоне остальных.

Хотя некоторые считают, что излишнее увлечение количественными подсчетами может убить любую гармонию, но при этом без количественного анализа текстов, трудно сделать объективные и обоснованные выводы. Однако, можно сказать:

...Поверил я алгеброй гармонию. Тогда

Уже дерзнул, в науке искушенный,

Предаться неге творческой мечты...

(А.С.Пушкин, “Моцарт и Сальери”)

«Сальери:

Все говорят: нет правды на земле.
Но правды нет - и выше. Для меня
Так это ясно, как простая гамма.

...

Я сделался ремесленник: перстам
Придал послушную, сухую беглость
И верность уху. Звуки умертвив,
Музыку я разъял, как труп. Поверил
Я алгеброй гармонию. Тогда
Уже дерзнул, в науке искушенный,
Предаться неге творческой мечты.

...

Усильным, напряженным постоянством
Я наконец в искусстве безграничном
Достигнул степени высокой. Слава
Мне улынулась; я в сердцах людей
Нашел созвучия своим созданьям.

...

Где ж правота, когда священный дар,
Когда бессмертный гений - не в награду
Любви горящей, самоотверженья.
Трудов, усердия, молений послан -
А озаряет голову безумца,
Гуляки праздного?... О Моцарт, Моцарт!»

(Пушкин А.С. Моцарт и Сальери)

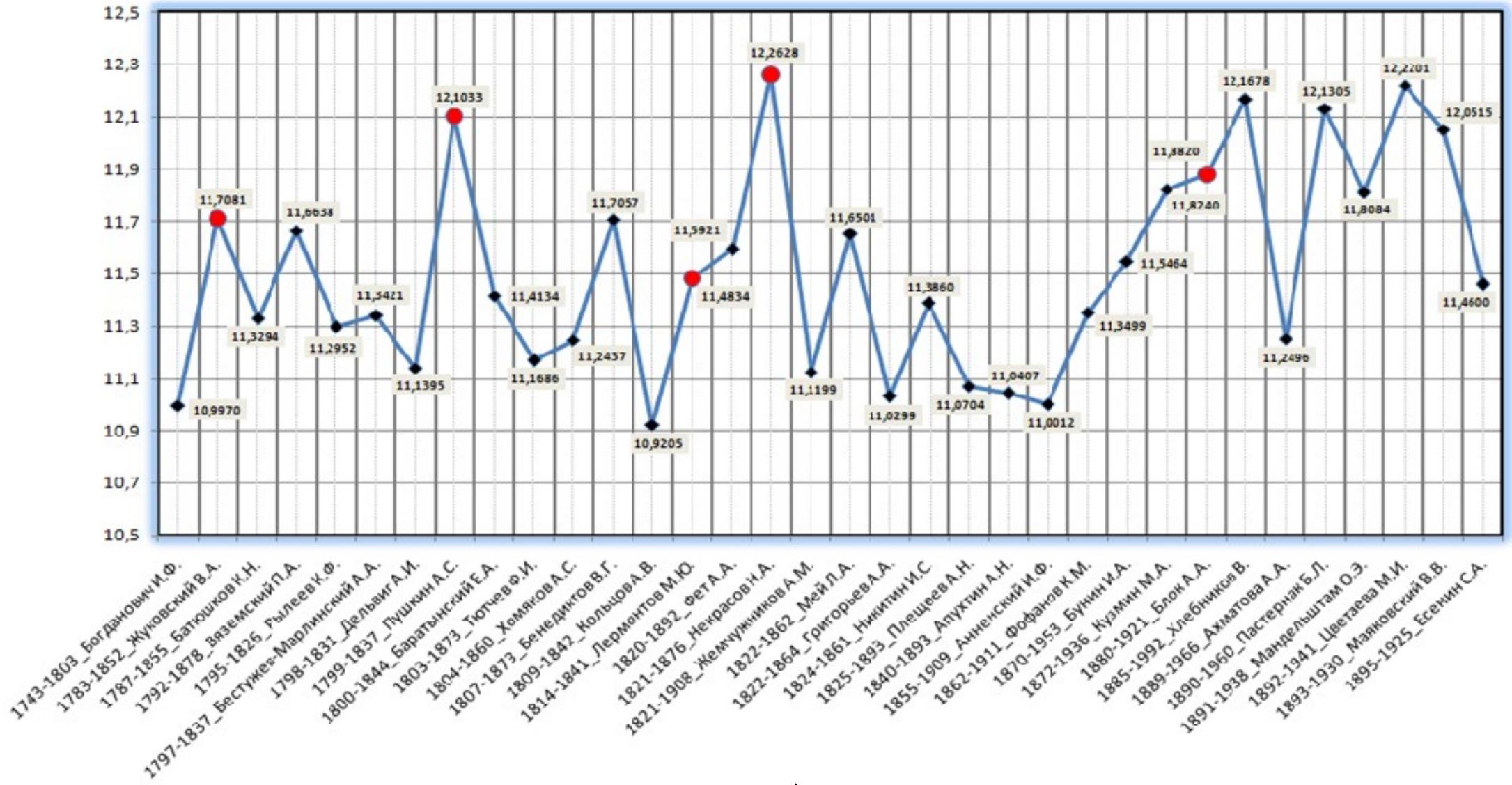
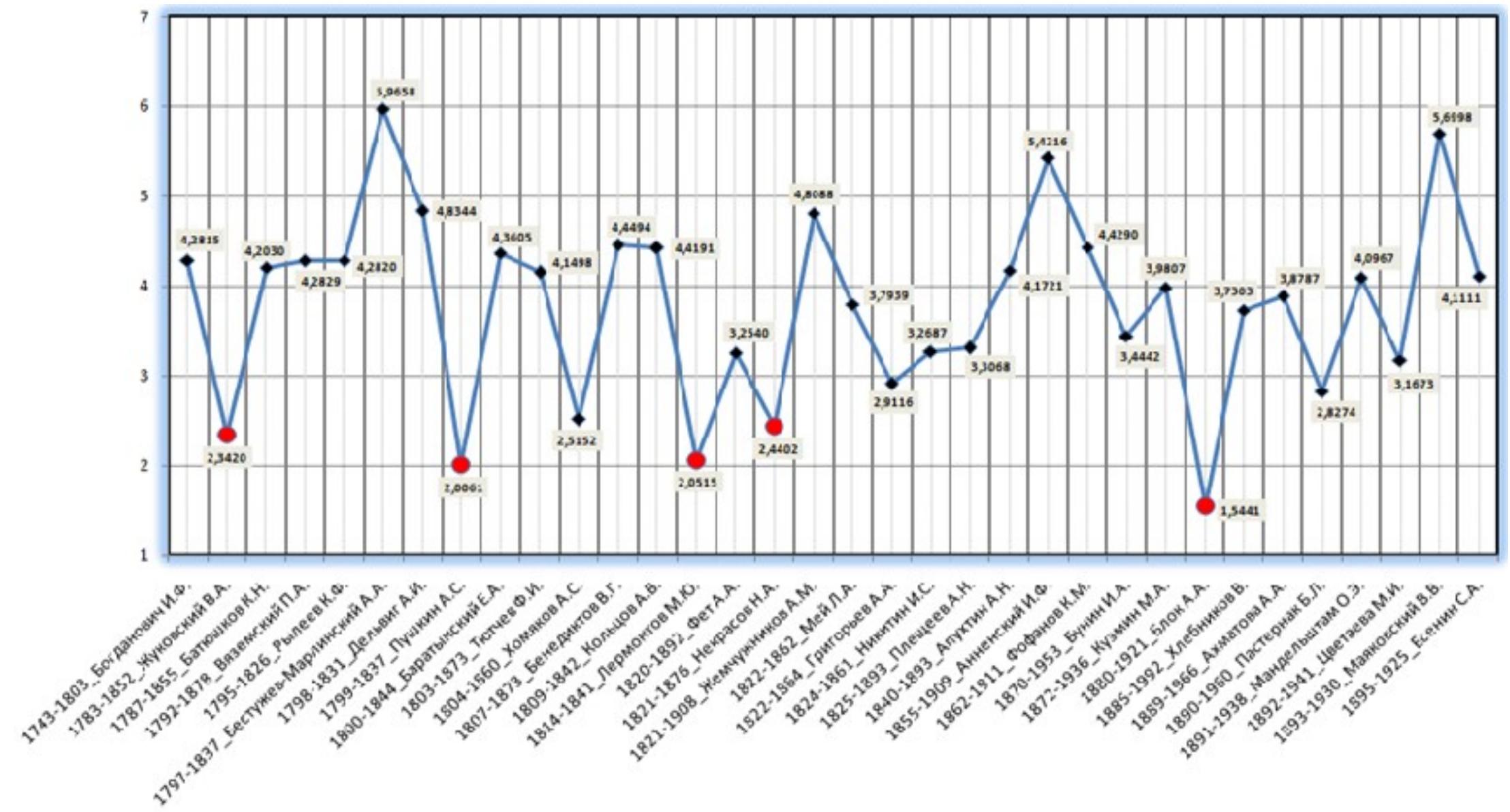
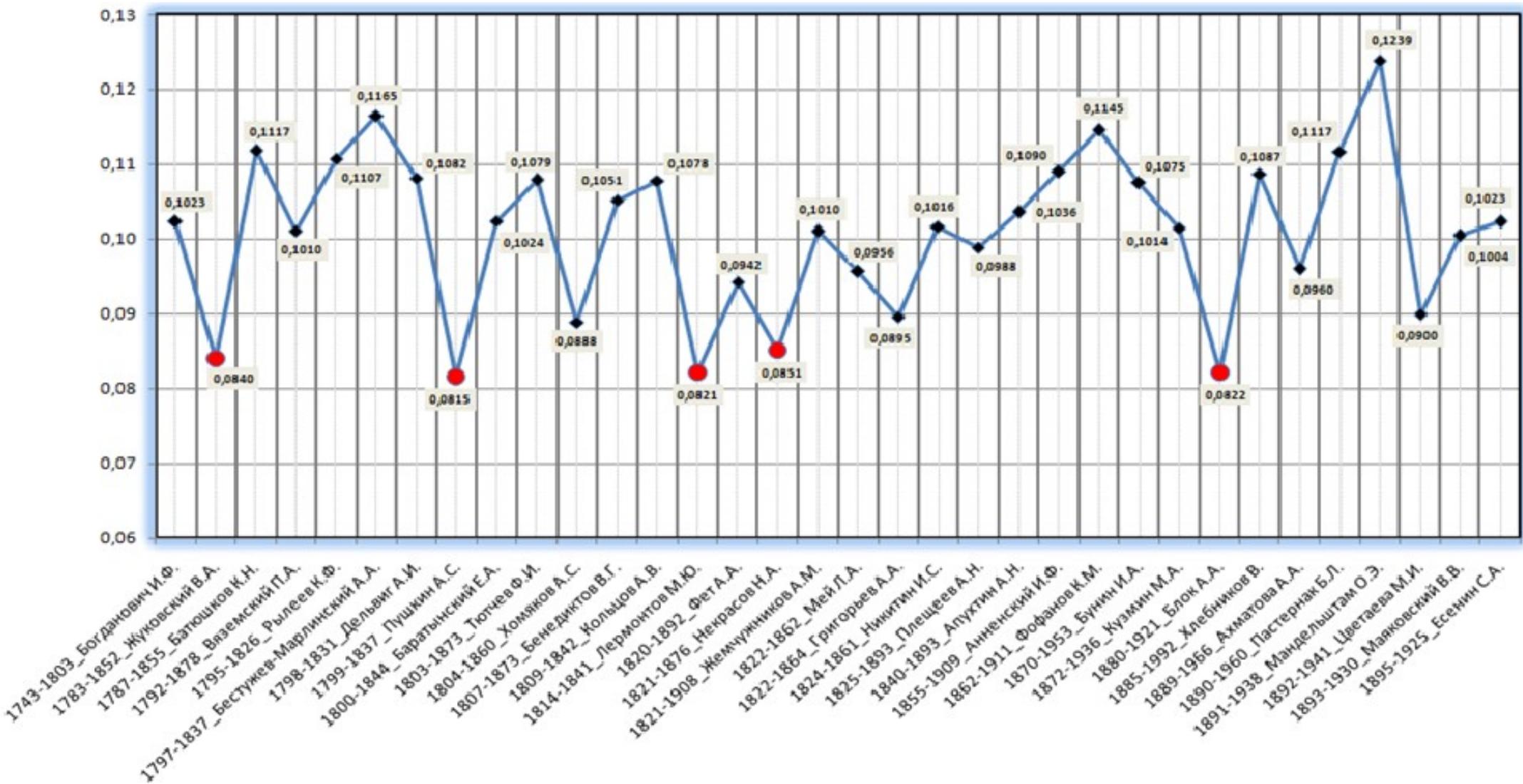
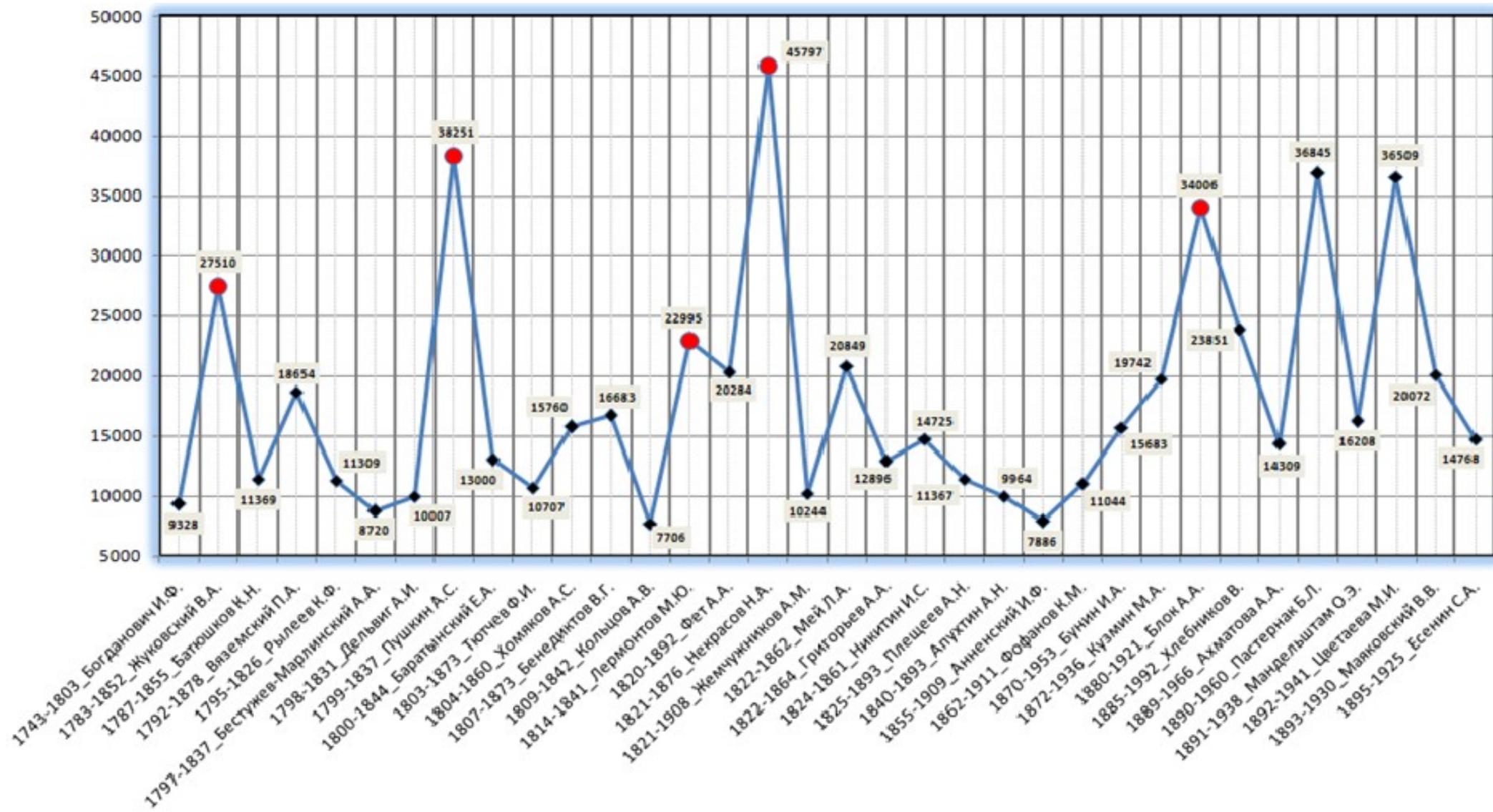


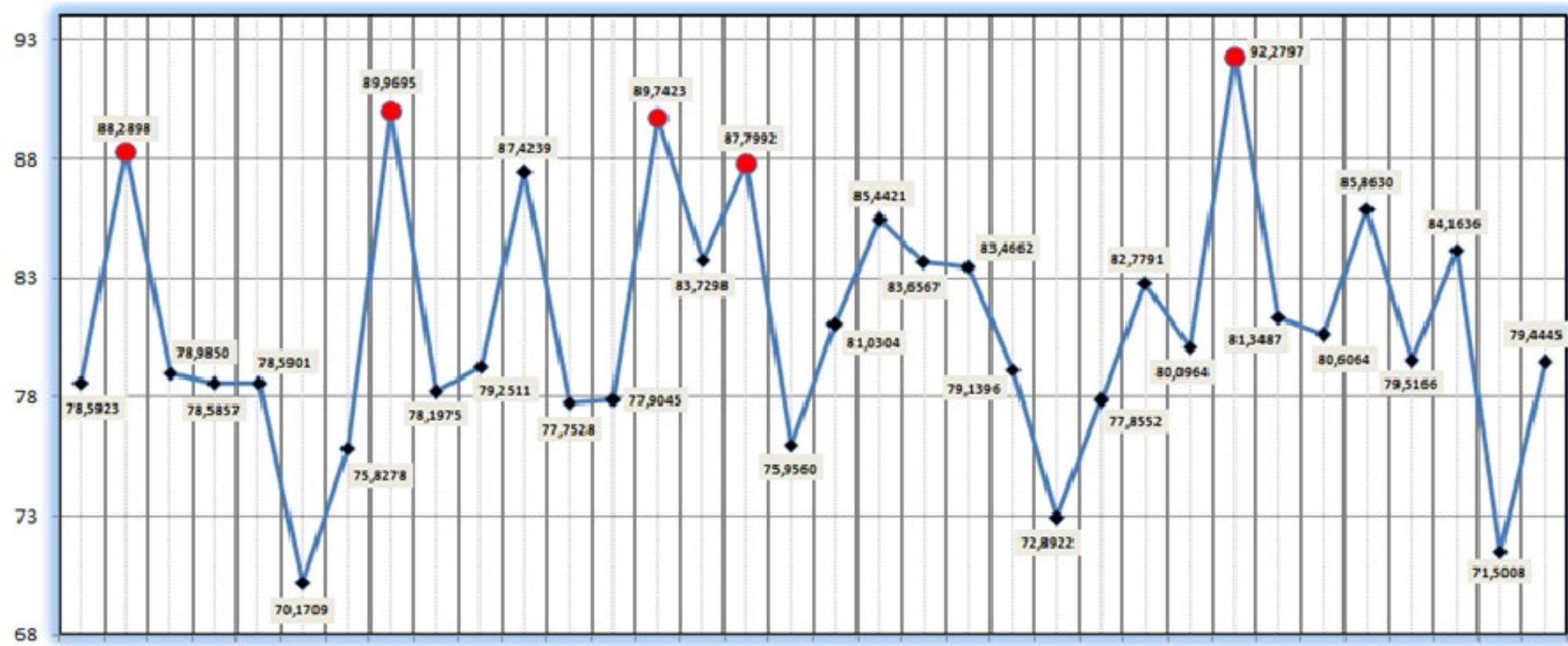
Диаграмма 1 – Накопленная энтропия H_k^* (мера неопределенности)

- 1 – Накопленная энтропия H^*k (мера неопределенности)
- 2 – Индекс итерации I_i (индекс повторения ЛЕ)
- 3 – Индекс исключительности (специфичности) лексики I_e
- 4 – Индекс плотности текста I_q
- 5 – Индекс дистрибуции I_c (богатство словаря текста)
- 6 – Индекс предсказуемости текста I_p
- 7 – Общий вклад поэтов в русский язык и литературу, культуру России

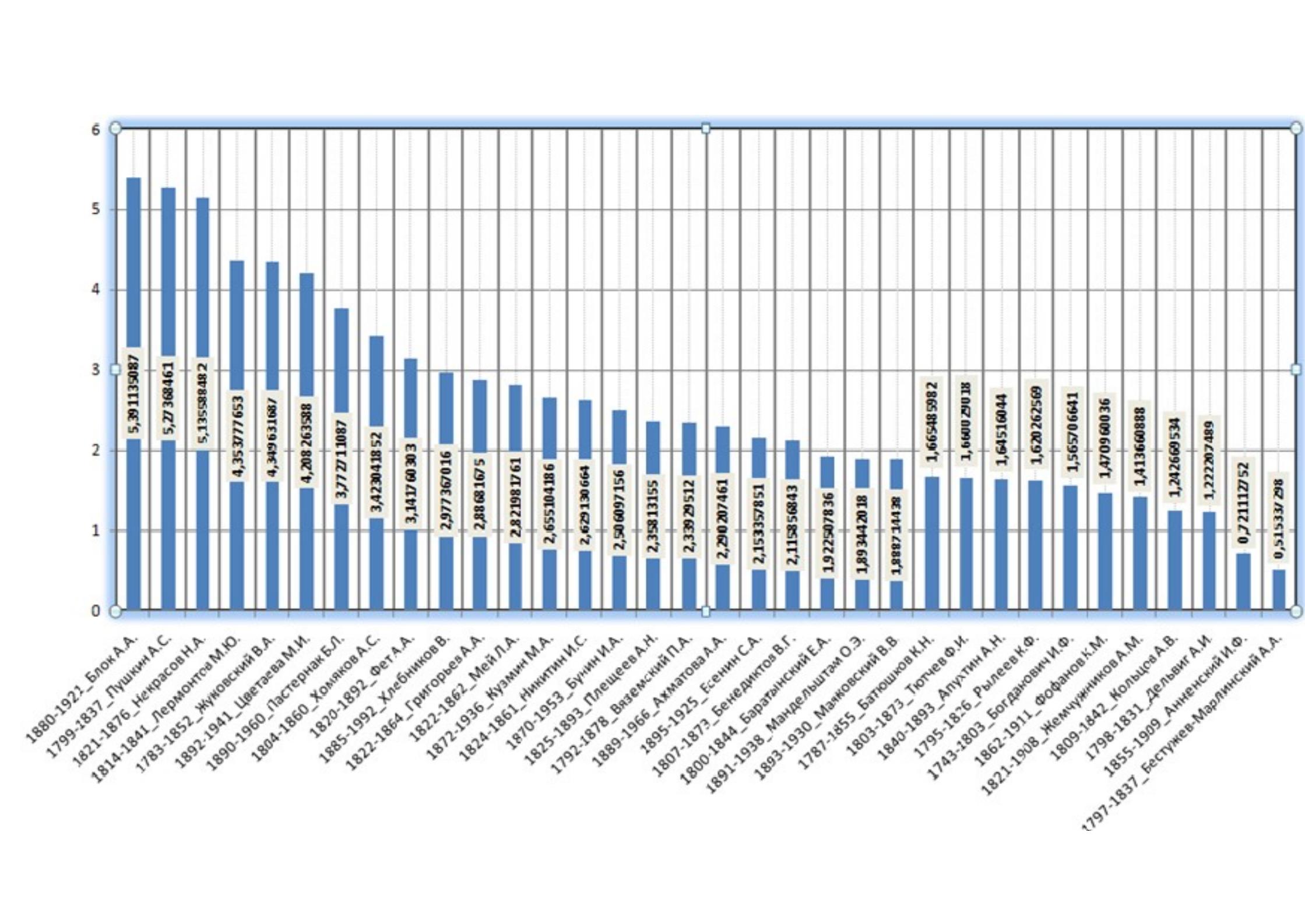








1743-1803_Богданович И.Ф.
1783-1852_Муновский В.А.
1787-1855_Батюкова К.Н.
1792-1878_Вяземский П.А.
1795-1826_Рылеев К.Ф.
1798-1831_Бестужев-Марлинский А.А.
1799-1837_Дельвиг А.И.
1800-1844_Пушкин А.С.
1803-1873_Баратынский Е.А.
1804-1860_Хомяков А.С.
1807-1873_Бенедиктов В.Г.
1814-1841_Лермонтов М.Ю.
1820-1892_Нольцов А.В.
1821-1876_Фет А.А.
1822-1864_Некрасов Н.А.
1824-1861_Григорьев А.А.
1825-1893_Никитин И.С.
1840-1893_Плещеев В.А.Н.
1846-1909_Апухтин А.Н.
1852-1911_Фофанов К.М.
1870-1953_Бунин И.А.
1880-1921_Кузмин М.А.
1885-1992_Хлебников В.
1889-1966_Ахматова А.А.
1890-1960_Пастернак Б.Л.
1891-1938_Мандельштам О.Э.
1892-1941_Цветаева М.И.
1893-1930_Маяковский В.В.
1895-1925_Есенин С.А.



Андрющенко В.М Концепция и архитектура Машинного фонда русского языка // Машинный фонд русского языка: Идеи и суждения. М.: Наука, 1986. 200с.

Андрющенко В.М. Вычислительная лексикография. Её возможности и перспективы // ВЯ.1986.№3. С.42-53. Лингвистический энциклопедический словарь. С.397.

Арапов М.В. Квантиативная лингвистика. М.: Наука, 1988. Бектаев К.Б. Статистико-информационная типология тюркского текста. Алма-Ата: Наука КазССР, 1978. 184с

Батура Т.В. Математическая лингвистика и автоматическая обработка текстов. Новосибирск: РИЦ НГУ, 2016. 166с.

Бектаев К.Б., Пиотровский Р.Г. Математические методы в языкоznании. Ч.1. Теория вероятностей и моделирование нормы языка. Алма-Ата: КазахГУ, ЧимкентПИ, 1973. 282с. Ч.2. Математическая статистика и моделирование текста. Алма-Ата, 1974. 335с.

Беляевская М.И. Математическая лингвистика и машинные языки. М.: Знание, 1973. 23с.

Гёдель K. Gödel K. Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I // Monatshefte für MATHEMATIK und Physik mit unterstützung des Österr. Bundesministerium für unterricht herausgegeben von H.Hahn und W.Wirtinger in Wien. XXXVIII. Band. I. Heft. Leipzig und Wien. Akademische Verlagsgesellschaft M.B.H. 1931. S.173-198. (Eingelang: 17.XI.1930)

Гладкий А.В. Лекции по математической лингвистике для студентов НГУ. Новосибирск: НГУ, 1966. 190с.

Гладкий А.В., Мельчук И.А. Элементы математической лингвистики. М.: Наука, 1969. 192с.

Дешериева Т.И. Языкоzнание и математика. Алма-Ата: Наука, 1973. 84с.

Ершов А.П. К методологии построения диалоговых систем: феномен деловой прозы. Новосибирск, АН СССР, 1979. 25с.

Заде Л.А. Понятие лингвистической переменной и его применение к понятию приближенных решений // Математика. Новое в зарубежной науке. / Под ред. Н.Н.Моисеева и С.А. Орловского. М.: Мир, 1976. 167с.

Исследования по математической лингвистике, математической логике и информационным языкам / Под ред. Р.А. Бочвара и Ю.А. Шрейдера. М.: Наука, 1972. 296с.

Кабаков Ф.А., Петров И.М., Френкель В.И. Математика для лингвистов. Ч.1. Множества. Алгебра. Логика. М.: МГПИИЯ, 1973. 160с.

Кайсарова Д.В., Коцюба И.Ю. Математическая лингвистика. СПб: Университет ИТМО, 2016. 67с.

Лапшин В.А. Лекции по математической лингвистике. М.: Научный мир, 2010. 248с.

Лесников С.В. Дефиниции и формулы для компьютерного анализа текстов // Прагмалингвистика и практика речевого общения. Ростов н/Д: ЮФУ, 2009. Вып.3. С.242-247.

Лесников С.В. Лингвистические основы информатики (теория, алгоритмы и лингвистические программы). Часть 1. Сыктывкар: Изд-во Сыктывкарского гос. ун-та, 2009. 292с.

Лесников С.В. Расчет энтропии текстов русской поэзии XIX-XX веков и индексов дистрибуции, итерации, исключительности, предсказуемости, плотности // Родной язык: проблемы теории и практики преподавания. Борисоглебск: БГПУ, 2011. С.81-87.

Лесников С.В. Тезаурус как отражение системности языка // Вестник Челябинского государственного университета. Серия: филология, искусствоведение. Вып.59. №28(243). Челябинск: ЧелГУ, 2011. С.52-61.

Лесников С.В. Фреймовое конструирование тезауруса метаязыка лингвистики // Вестник Поморского университета. Серия «Гуманитарные и социальные науки». № 4. Архангельск: ПГУ, 2011. С. 84-89.

Лесников С.В., Ловягин Ю.Н. Математика: Программа курса для студентов филологического и финно-угорского факультета. Сыктывкар: СыктГУ, 1995. 27с.

Математическая лингвистика. 1. Ежегодник по структурной, прикладной и математической лингвистике. Киев: КГУ, 1973. 140с. (1270 экз., 16 статей). Структурная и математическая лингвистика. 2. 1974. 128с. (920 экз., 17 статей); 3. 1975. 120с. (1000 экз., 18 статей); 4. 1976. 112с. (1000 экз., 20 статей); 5. 1977. 129с. (1000 экз., 17 статей); 6. 1978. 119с. (1000 экз., 18 статей).

Математическая логика и математическая лингвистика: Межвузовский тематический сборник / Отв. ред. А.В.Гладкий. Калинин: КГУ, 1981. 172с.

Метаязык науки: материалы Международной научной конференции / Глав. ред. С. В. Лесников. Сыктывкар: СыктГУ, 2012. 878с.

Минский М. Фреймы для представления знаний / Пер. с англ. О.Н.Гринбаума. Под ред. д.т.н. Ф.М.Кулакова. М.: Энергия, 1979. 152с.

Пальм Р. Математическая лингвистика. Часть 1 // Труды вычислительного центра. Вып.12. Тарту: ТартуГУ, 1968. 140с.

Пиотровский Р.Г., Бектаев К.Б., Пиотровская А.А. Математическая лингвистика. М., Высшая школа, 1977. 383с.

Пиотровская А.А., Пиотровский Р.Г., Разживин К.А. Энтропия русского языка. ВЯ. 1962. N6. С.115-130.

Ревzin И.И. Современная структурная лингвистика. Проблемы и методы. М.: Наука, 1977. 264с

Тимофеева М.К. Математические методы в филологии. Новосибирск: ННИГИ, 2011. 79с.

Тузов В.А. Математическая модель языка. Л.: ЛНИВЦ АН СССР, 1980. 47с. ЛГУ, 1984. 176с.

Филиппович Ю.Н. Ч.1. Семантические исследования естественно-языкового описания.
http://www.philippovich.ru/Library/Books/Semantics_IT/gl1_2/glava1_2.htm

Хомский Н. Формальные свойства грамматик // Кибернетический сборник. Вып.2. М.: Мир, 1966. С.121-230.

