

# Соотношение цифрового и гуманитарного в ДН

Борис Орехов

Высшая школа экономики, Пушкинский Дом

*nevmenandr@gmail.com*

8 октября 2024

- 1 Традиция и цифра
- 2 Digital humanities как дисциплина
- 3 Как науки соотносятся между собой
- 4 Моделирование
- 5 Digital humanities как междисциплинарность?
- 6 Объективность?
- 7 Как возможна формализация?
- 8 Предварительные итоги

- 1 Традиция и цифра
- 2 Digital humanities как дисциплина
- 3 Как науки соотносятся между собой
- 4 Моделирование
- 5 Digital humanities как междисциплинарность?
- 6 Объективность?
- 7 Как возможна формализация?
- 8 Предварительные итоги

- Чтобы перейти от традиционных идиографических описаний к применению количественных методов, необходимо перестроить сам взгляд на материал, чтобы он позволял видеть общее в уникальном.
- Например, не ставя под сомнение единственность в своем роде каждого отдельного стихотворения, ученый может сосредоточиться на некотором формальном приеме, истории его применения в рамках определенной художественной традиции.
- В таком случае прием уже может иметь количественное измерение, оно может быть описано как закономерность.

Так действуют применяющие количественные методы стиховеды, подсчитывающие частотность использования поэтических размеров:

«Обследуются не типы, а массы, не единичные явления, а общие состояния. Рассматриваются многие тысячи стихов данного размера, рассматриваются в различных разрезах, и — “большие числа” торжествуют над случайностью, выстраивая закон».

- 1 Традиция и цифра
- 2 Digital humanities как дисциплина**
- 3 Как науки соотносятся между собой
- 4 Моделирование
- 5 Digital humanities как междисциплинарность?
- 6 Объективность?
- 7 Как возможна формализация?
- 8 Предварительные итоги

- Такой подход, естественно, вызывает закономерный скепсис в глазах традиционных филологов и искусствоведов.
- Подсчет частотности вступает для них в противоречие с привычными формами работы с материалом.
- Поэтому неверно было бы разделять историю гуманитарных наук на доцифровой и цифровой периоды: одновременно с активным использованием компьютерных методик существуют и более привычные подходы.

- Тем не менее, вряд ли можно найти гуманитарную специальность, которую бы сегодня так или иначе не затронул «цифровой поворот».
- Любое гуманитарное исследование сегодня основано на спонтанной или систематической, выборочной или сплошной оцифровке
  - текстов,
  - документов,
  - изображений
  - других объектов историко-культурного наследия
- Это делает эти объекты более доступными для исследователя, то есть снижает порог входа для ученого.

- Если в доцифровую эру на пути у филолога, историка, искусствоведа могла стоять недоступность текста, живописного полотна или документа, то в современной ситуации наличие цифровой копии вовлекает в исследовательский процесс все большее число специалистов.
- Академик Е. Э. Бертельс, составляя историю персидско-таджикской литературы, писал:  
«ссылаться на книгу, которая имеется только, скажем, в библиотеке Института востоковедения в Ленинграде и которую нельзя найти даже и в Москве, или на рукопись, находящуюся в частной библиотеке, было бы просто недобросовестно».
- Оцифровка стала одной из важных ежедневных практик ремесла гуманитария.
- Оцифровка тоже имеет свою теорию.

- Статус наиболее авторитетного в сфере цифрового литературоведения удалось получить Стэнфордской литературной лаборатории (памфлет).
- В исторических исследованиях широкую известность приобрели Центр истории и новых медиа имени Роя Розенцвейга (RRCHNM), Люксембургский центр современной и цифровой истории (C2DH).
- В России соответствующие институты имеются в Москве (МГУ, НИУ ВШЭ), Санкт-Петербурге (ИТМО, Институт русской литературы РАН), Барнауле (АлтГУ), Екатеринбурге (УрФУ), Калининграде (БФУ), Красноярске (СФУ), Перми (НИУ ВШЭ-Пермь), Ростове-на-Дону (ЮФУ), Томске (ТГУ).

- Digital scholarship in the humanities (с 2012 года, с 1986 года выходил под названием Literary and Linguistic Computing),
- Digital humanities quarterly (с 2007 года),
- Историческая информатика (с 2012),
- отчасти Journal of Cultural analytics

Новые:

- International Journal of Digital Humanities (с 2019 года),
- Квантитативная филология (с 2021 года),
- Journal of Digital History (с 2021 года),
- Digital Orientalia (с 2021 года),
- Journal of Computational Literary Studies (с 2023 года),
- Цифровые гуманитарные исследования (с 2024 года).

- Важными для определения границ научного поля становятся конференции, центральное место среди которых занимает ежегодная созываемая ассоциацией организаций цифровых гуманитарных наук (ADHO).
- Состав секций и круглых столов этой конференции лучше всего описывает ключевые направления, в которых работают представители цифровых гуманитарных наук:
  - самоопределение digital humanities
  - проблемы оцифровки, организации электронных изданий и публикации данных
  - методы: обработка естественного языка, сетевой анализ, компьютерное зрение

- В рамках деятельности ДН-исследователя может быть создан:
  - набор данных,
  - онлайн-ресурс,
  - информационная система.
- Проектный подход в том числе означает и соавторство, участие в исследовании нескольких авторов, каждый из которых вносит свой вклад, соразмерный его компетенциям.
- Обычная для цифровых гуманитарных наук картина — это наличие нескольких авторов у одной публикации.
- У абсолютного большинства статей на гуманитарную тематику только один автор.

- И, наконец, с точки зрения профессионального сообщества, DH — это полезный бренд, позволяющий обращаться за финансированием и административной поддержкой для вполне классических гуманитарных дисциплин.
- Скептическое отношение традиционных гуманитариев, подмечающих
  - методологические слабости молодой компьютерной области,
  - легковесность стоящих за цифровыми исследованиями концепций,
  - неоднозначность продуцируемых выводов.



<https://www.youtube.com/watch?v=RfRZvsTQzYA>

- 1 Традиция и цифра
- 2 Digital humanities как дисциплина
- 3 Как науки соотносятся между собой**
- 4 Моделирование
- 5 Digital humanities как междисциплинарность?
- 6 Объективность?
- 7 Как возможна формализация?
- 8 Предварительные итоги

# Шутка про разнообразие определений

<http://whatisdigitalhumanities.com/>: 817 определений.

Большинство определений там не являются строгими определениями.  
Одно из 817: *working with computers to process research questions in the humanities* (Elisa E. Beshero-Bondar)

Ключевые слова: humanities и research questions

- Есть стереотип о двух культурах.
- Чему противопоставлены гуманитарные науки?
- Естественным и социальным.
- Три культуры.
- Деление обусловлено исторически.

# Интуиция против классификации

- Мы часто исходим из интуитивных представлений
- Интуиции так или иначе исходят из современного положения вещей
- На самом деле классификации историчны
- К сложившейся системе наук подталкивает устройство мира вокруг нас?
- «Вот есть материя, и вот поэтому существует физика. Вот кислота взаимодействует с металлом, и вот поэтому существует химия».
- Классификация — это никогда не отражение мира, это отражение нашего способа думать про некоторый предмет.
- А способ думать про что-то в свою очередь отражает разные обстоятельства, в которых этот способ появился.

- понятие «геологической революции» возникает после Великой французской революции
- социальные практики переносятся на наше осмысление того, как устроены геологические феномены
- не существует какой-то правильной формы науки, единственно адекватной миру вокруг нас
- не существует единственно правильной, предвечной, данной небом классификации наук
- гуманитарные науки — результат сложных процессов, которые, как нельзя исключать, могли бы пойти иначе

Альберт Великий (XIII в.) посвящал трактаты

- металлургии,
- именам ангелов
- лекарствам,
- как ухаживать за скотом
- как организовывать торговое предприятие

# Что такое наука?

- Наиболее убедительное описание (для меня) науки в том, что это производство высказываний на определенном концептуальном языке.
- Язык в данном случае — это метафора.
- В естественном языке есть слова, а в языке конкретной науки есть концепты, выраженные в терминах.
- Они обуславливают осмысленность тем, на которые в этой науке производятся тексты.

# Пример: что такое колониализм?

- Для человека со стороны это грабеж колоний.
- Для ученого это иерархическое мышление.
- Мир вещей и мир концептов
- Мир вещей не полностью соответствует миру концептов.
- Концепты подвижны и дают разную оптику для понимания мира вещей.
- У каждой науки (и даже поднауки) свой концептуальный язык.
- Свой, особенный, способ видения мира вещей.

- Скажем, в рамках языка физики вопрос о цвете обложки книги бессмысленный.
- Просто нет концептов, которые бы смогли такую проблему описать.
- В рамках языка книговедения такой вопрос наоборот приобретает осмысленность.

Августин (IV–V в.): «Смешно, когда мы видим, что языческие боги в силу разнообразных людских выдумок представлены распределившими между собой знания, подобно мелочным откупщикам налогов или подобно ремесленникам в квартале серебряных дел мастеров, где один сосудик, чтобы он вышел совершенным, проходит через руки многих мастеров, хотя его мог бы закончить один мастер, но превосходный. Впрочем, иначе, казалось, нельзя было пособить массе ремесленников, как только тем, что отдельные лица должны были изучать быстро и легко отдельные части производства, а таким образом исключалась необходимость, чтобы все медленно и с трудом достигали совершенства в производстве в его целом».

- Разные языки могут быть в пределах одной науки.
- Интервенции и заимствования концептов и терминов — термины при этом переопределяются и переосмысливаются. Как и в естественных языках.
- Но невозможно говорить одновременно на разных языках.
- В естественных языках мы видим либо переключение между языками, либо порождение нового — пиджина.
- Поэтому не бывает междисциплинарности.

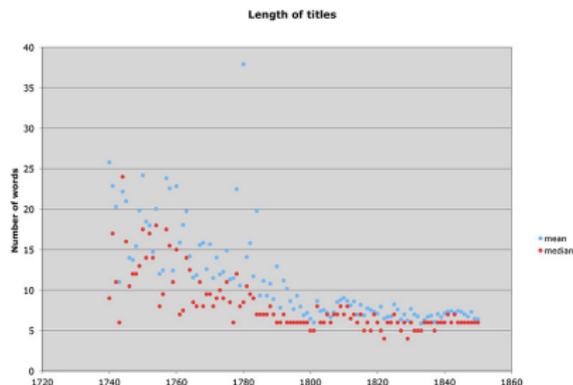
- Ваш вопрос либо осмыслен с точки зрения филологии, либо с точки зрения математики
- цепи Маркова, на материале «Онегина»:
- для понимания Онегина с точки зрения литературоведения это ничего не дает
- Colour evolution of Betelgeuse and Antares over two millennia, derived from historical records, as a new constraint on mass and age
- <https://theeffectbook.net/ch-ResearchQuestions.html>

Представим, что за человеком, который стоит на крыше, наблюдают три специалиста:

- у физика свой интерес (ускорение свободного падения),
- у медика свой (разрыв внутренних органов, травмы, несовместимые с жизнью),
- у социологов свой (что привело его на край крыши — и действительно, у одного из классиков социологии, Эмиля Дюркгейма есть работа о социальных основаниях самоубийства)

- Создает ли цифра новый язык описания или опирается на традиционный гуманитарный?
- Что у нас исследовательский вопрос?
- Как объясняется результат?
- У литературы есть собственные процессы, не связанные с социальным, например, усталость от формы.
- «Кризисы и выравнивания» у Гаспарова в истории рифмы.





- Если вам будут говорить, что Моретти литературовед, не верьте
- Его взгляд на литературу социологический, таков его исследовательский вопрос и объяснительная модель

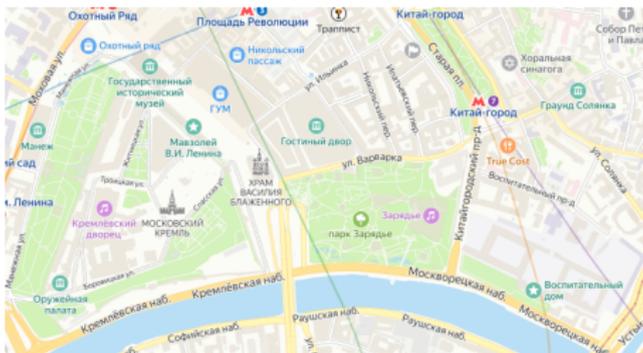
«Так называемый формальный метод образовался не в результате создания особой методологической системы, а в процессе борьбы за самостоятельность и конкретность литературной науки»

В предисловии к книге «Молодой Толстой» (1922 г.) он говорит, что охотнее назвал бы его морфологическим «в отличие от других (психологического, социологического и т. д.), при которых предметом исследования служит не само художественное произведение, а то, “отражением” чего является оно по мнению исследователя».

- 1 Традиция и цифра
- 2 Digital humanities как дисциплина
- 3 Как науки соотносятся между собой
- 4 Моделирование**
- 5 Digital humanities как междисциплинарность?
- 6 Объективность?
- 7 Как возможна формализация?
- 8 Предварительные итоги

# Как «работают» количественные методы?

- Исходный материал упрощается («редукция»).
- Создается модель, отражающая в параметрах свойства исходного материала.
- Параметры должны быть выражены количественно.
- Анализируются распределения значений параметров, корреляция параметров между собой, выводятся закономерности.



# Упрощение не всегда осознаваемо

Пикассо садится в такси, таксист его узнал и говорит:

— Зачем вы так непохоже пишете?

— А похоже — это как?

Таксист достает из бумажника фотографию жены:

— Вот моя жена, здесь она на себя похожа.

— Что, она на самом деле такая маленькая и плоская?



# Опыт применения количественных методов в естественных науках

- Естественные науки — *номологические*. Они ищут закономерности в **повторяемости**.
- Это часть их концептуального языка
- Представители естественных наук сравнительно легко мирятся с редукцией. То же в лингвистике. Ср. у Соссюра: «В шахматах важны правила игры, а не материал, из которого сделаны фигуры».
- Это позволяет вывести формулы, которые увязывали бы параметры модели в одну систему:

$$E = mc^2$$

$$g = G \cdot \frac{M}{(r+h)^2}$$

- Гуманитарные науки — идиографические. Их интересует **единичное**, а не повторяемое.
- Страх перед редукцией. Не потеряем ли мы что-то важное? Идея «художественного целого».
- Представление о творческих агентах как о тех, кто стремится менять правила.

М. И. Шапир: «С успехом применяя точные методы для теоретического описания физического мира, мы бессильны построить универсальную математическую модель даже самого «простого» и «поверхностного» явления духовной культуры. Причина банальна: нельзя сконструировать математизированную картину универсума, которым управляет множество относительно равноправных демиургов; кто-нибудь из них во что бы то ни стало отменит закон, установленный другим».

# Как не потерять важное?

- Помимо истории, взаимодействия персонажей и прочих драматических элементов у Эсхила важны лейтмотивы, связывающие вместе трилогию «Орестея».
- Все трагедии связаны идеей-образом *зверь, попавший в сеть*.
- Эсхил оформляет лейтмотив разнообразно, обозначая его понятием «сеть», обогащает образами западни, пут, ткани, охватывающей тело, извивов змеи.

Что эта ткань? Капкан ли? Зверолова ль сеть?  
Иль с головою тело пеленающий  
Могильный саван? Путы ли на пленника?

Зевс, Зевс всезрящий! Будь сих дел свидетелем!  
Воззри на сырой выводок отца-орла!  
Ехидны лютой кольцами задушен он;  
Орел — змеи добыча.

Х. Ортега-и-Гассет «Дегуманизация искусства»: «Нет ничего более легкого, нежели подмечать различия между отдельными произведениями. Но подобное акцентирование различий и специфики ни к чему не приведет, если сначала не определить то общее, которое разнообразно, а порою и противоречиво утверждается во всех них. Еще старик Аристотель учил, что вещи различаются между собою в том, в чем они походят друг на друга, в том, что у них есть общего»

- 1 Традиция и цифра
- 2 Digital humanities как дисциплина
- 3 Как науки соотносятся между собой
- 4 Моделирование
- 5 Digital humanities как междисциплинарность?**
- 6 Объективность?
- 7 Как возможна формализация?
- 8 Предварительные итоги

# Установка оставаться внутри концептуального языка своей науки

Б. И. Ярхо: «Ни один математический акт не должен совершаться, пока в него не будет вложен конкретный литературоведческий смысл».

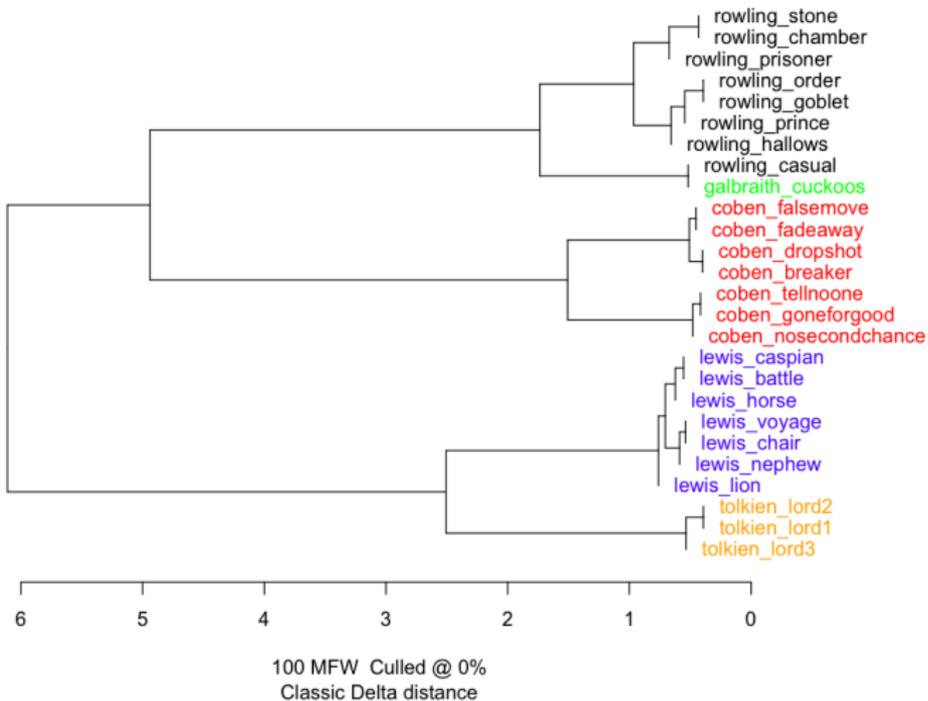
Возможно ли свое «дигитальное» литературоведение с отличной от традиционного литературоведения концептуализацией?

- 1 Традиция и цифра
- 2 Digital humanities как дисциплина
- 3 Как науки соотносятся между собой
- 4 Моделирование
- 5 Digital humanities как междисциплинарность?
- 6 Объективность?**
- 7 Как возможна формализация?
- 8 Предварительные итоги

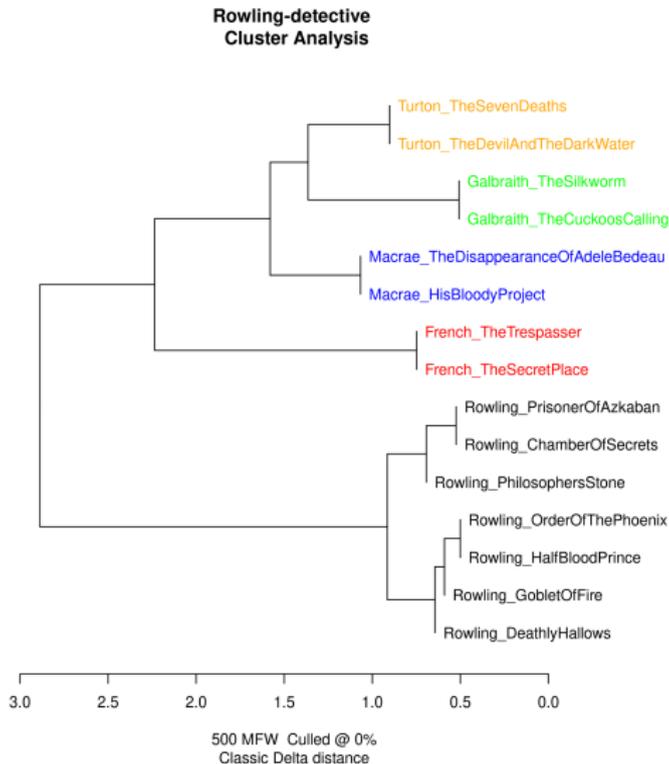
- Субъективность наступает нас на разных этапах:
  - когда мы ставим задачу,
  - когда мы выбираем методы,
  - когда мы отбираем материал,
  - когда мы интерпретируем результаты.

Даже там, где ситуация предельно упрощена до бинарности: случай Роулинг.

## StylometryPractice Cluster Analysis



# Роулинг на фоне других детективов



- Дж. Барнс «История мира в  $10\frac{1}{2}$  главах»: *История — это ведь не то, что случилось. История — это всего лишь то, что рассказывают нам историки.*
- Ф. Анкерсмит «История и тропология: взлет и падение метафоры»: *Исторические нарративы — это интерпретации прошлого.*
- Там же: *Научные теории являются неопределенными, поскольку бесконечное число теорий может служить объяснением известных данных; интерпретации же являются недоопределенными, поскольку только бесконечное число интерпретаций могло бы служить объяснением всех известных данных.*

- Чтобы интерпретировать данные, мы пользуемся метафорой (ср. «темные века») и изобретаем нарратив (или следуем нарративу).
- Выбор метафоры, как и классификация (см. разговор в прошлый раз), субъективен или (и) социально обусловлен.
- Метафора работает и до, и после подсчетов.
- Это работает не только для исторической науки, но и для истории литературы и истории искусства.

# Мисик М. А. Историописание «Постсовременности» (2004)

В 1995 г. на страницах Журнала Современной Истории (Лондон) <...> известный британский историк Артур Марвик в резко полемичной форме выступил в защиту истории от так называемых «постмодернистов». <Настоящие историки> отбросив «простые спекуляции» о прошлом, долго и скрупулезно изучают его следы, т.е. первичные источники, давая таким образом жизнь знанию о прошлом, поступая при этом так же, как исследователи природного мира. По его твердому убеждению, техническое мастерство историка состоит как раз в рассортировывании источников. При этом Марвик игнорирует то обстоятельство, что сам процесс рассортировывания материала и есть начало его интерпретации.

Принцип «онтологической относительности» заключается в том, что знание об объектах обусловлено теми научными теориями, которые мы используем. Он был введен американским философом, логиком и математиком Уиллардом Куайном (1908–2000).

Н. В. Брагинская: «Угроза репутации комментатора со стороны электронных средств поиска информации существует. А именно той репутации, которая строится на эрудиции, на больших мнемонических способностях, огромном многолетнем труде по накоплению знаний. Например, авторитетный филолог считает, что в таком-то тексте есть влияние аттицизма: дескать, вот это, и это, и это слово характерны для аттикизирующей стилизации. И приводит примеры. Ему верят как эксперту, на этом основании датируют памятник. Но по базе данных, где помещается практически вся греческая литература, выясняется, что названные им слова равномерно распределены по доаттикизирующему, аттикизирующему и уже не аттикизирующему периоду».

# Для чего нужны компьютеры в филологии?

Н. В. Брагинская: «Экспертные оценки часто терпят крах, мнемонический ресурс падает в цене. Однако наличие баз данных дает возможность ставить такие научные и комментаторские задачи, о которых до их появления нечего было и думать. Непосредственно базы данных, ресурсы Интернета дают доступ только к готовому знанию-справке и освобождают от рутины, а не от профессии».

- 1 Традиция и цифра
- 2 Digital humanities как дисциплина
- 3 Как науки соотносятся между собой
- 4 Моделирование
- 5 Digital humanities как междисциплинарность?
- 6 Объективность?
- 7 Как возможна формализация?**
- 8 Предварительные итоги

# Что такое формализация?

- Чтобы использовать цифровой подход (не только исследования, оцифровка тоже), мы должны формализовать материал.
- Формализация — это та же редукция.
- Сложный объект членится на **атомарные факты** (термин Л. Витгенштейна).
- Это должны быть **значимые** для языка науки факты.
- Желательно, чтобы эти факты было можно детектировать компьютерным способом.
- Они должны быть структурированными.
- Случай Проппа: формализация не обязательно означает квантификацию.

# Что можно формализовать?

- Конвенционально принятая область: стиховедение (метр, ритм, рифма, строфика).
- Язык художественной литературы (частотный словарь — это тоже модель, хотя и плоская, одномерная).
- Всё то, что имеет дело не с высокоуровневыми понятиями (стиль, смысл), а с атомарными фактами: слоги, метры, слова.
- И даже такие понятия, как персонаж, формализуется через атомарные факты, например, слова и их отношения в тексте.

# Содержание

- 1 Традиция и цифра
- 2 Digital humanities как дисциплина
- 3 Как науки соотносятся между собой
- 4 Моделирование
- 5 Digital humanities как междисциплинарность?
- 6 Объективность?
- 7 Как возможна формализация?
- 8 Предварительные итоги**

# Зачем нужны digital humanities?

- DH не столько позволяет получить новое знание, сколько измерить, калибровать то, что уже было известно, но осмыслено в качественных (не количественных) категориях.
- Измерение позволяет сравнивать (башкирский и киргизский стих).
- Наиболее осмысленно на больших объемах (великое непрочтенное), там, где наиболее выражены надличностные силы: жанр, смена эпох. И где филологи-качественники не конкуренты.
- Хуже работает на отдельных текстах.