

# Честный разговор на неудобную тему:

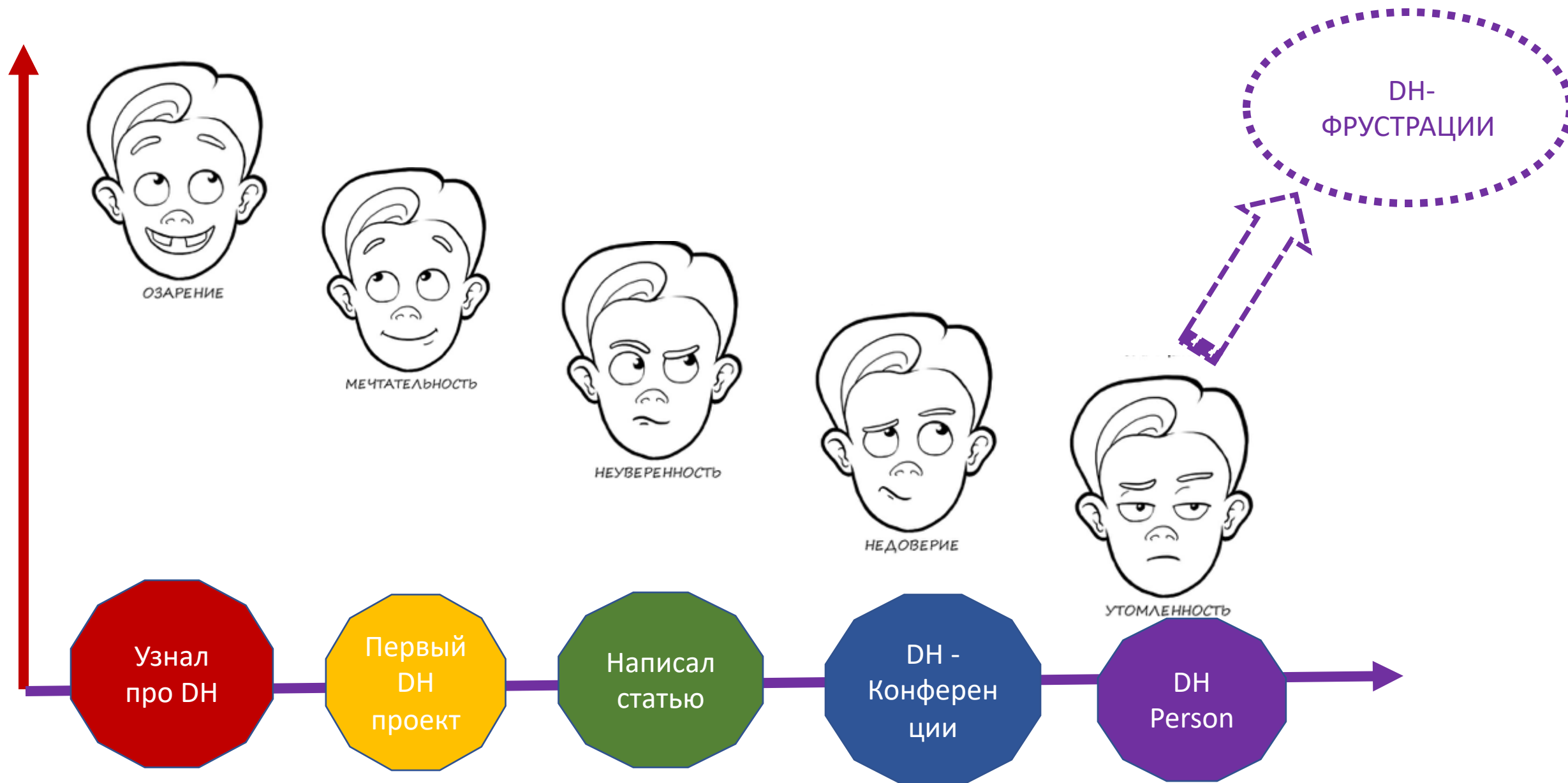
что нас раздражает в ДН и как не потерять  
самообладания



**#dhvoronovo 2019**

**Анастасия Бонч-Осмоловская**

# Типовой анамнез цифрового гуманитария



# Эволюция фрустрации

- Digital Humanities – новая гуманитарная наука 21 века
- Мы оцифруем все
- Это будет очень красиво
- И у нас тоже будут большие данные



УТОМЛЕННОСТЬ

1. Нескончаемая тягомотина самокопания
2. Недоделанность как идеология
3. Шарлатанские вау-эффекты
4. Бессмысленные игры в доказательную медицину



ОЗАРЕНИЕ

Попытки  
преодолеть кризис  
упираются в п.1



УТОМЛЕННОСТЬ

# Тягомотина самокопания

Фрустрация 1

817 определений

## What Is Digital Humanities?

DH is where the traditional fields of the humanities meet the technological advances, and increased accessibility, made possible by computers. It enables researchers to combine traditional methods with the processing power of computers.

*Sara Kerr*

---

**NB:** Refresh the page to get a new definition. Quotes were pulled from participants from the [Day of DH](#) between 2009-2014. As of January 2015, the database contains 817 rows and randomly selects a quote each time the page is loaded. If you want to do something cool with the data, I am providing a download for the CSV I compiled [here](#).

# ДН от противоположного:

- не наука
- не инструменты
- это люди
- это секта?

I still think digital humanities is hard to define, but apparently, being hard to define doesn't prevent human institutions from enduring and growing.

“DH” doesn't have to be defined abstractly at all. It could be understood as an institutional achievement that happens to exist on some campuses



Ted Underwood

# Три столпа ДН

- Принципиально новый подход к **сохранению гуманитарного наследия**
- Принципиально новый подход к **доступности гуманитарного знания**
- Принципиально новые **методы получения гуманитарного знания**





# Недодделанность как идеология

Фрустрация 2

# Проекты по цифровому сохранению: худшие ожидания

- Цифровое не значит **надежное**
- Цифровое не значит **полное**
- Цифровое не значит **быстрое**
- Цифровое не значит **дешевое**

Практика «застолбить площадку» vs.  
традиционное консервативное накопление

# What future for digital scholarly editions? From Haute Couture to Prêt-à-Porter (May, 2019)

The main problem is that creating such an infrastructure requires a degree of standardization and implies a loss of originality in the edition itself: it basically transforms the digital edition into an edition produced and published digitally.

We shape our tools, and thereafter our tools shape us (Marshall McLuhan)



Elena Pierazzo



# Шарлатанские вау-эффекты

Фрустрация 3

# Шарлатанство как полезный концепт

Джон Ансворт. Чем является и чем не является гуманитарная информатика (статья 2002 г.)

- Все проекты по гуманитарной информатике отличаются некоторым шарлатанством
- Шарлатаны продают как интеллектуальную непротиворечивость и логическую связность ( то, что предполагает сложный труд, с большим количеством шагов и данных) поверхностную работу, произведенную мгновенно, с последующими шаблонными выводами.
- Самое важное – это интерактивность, т.е. возможность для пользователей генерировать собственные научные вопросы. Если интерактивности нет, то такой проект – чистое мошенничество.

# Революция доступности знания

Радикальное снижение порога  
входа и доступа

- Онлайн курсы
- Музеи
- Эдьютейнмент и сторителлинг
- Краудсорсинг
- Свободное время

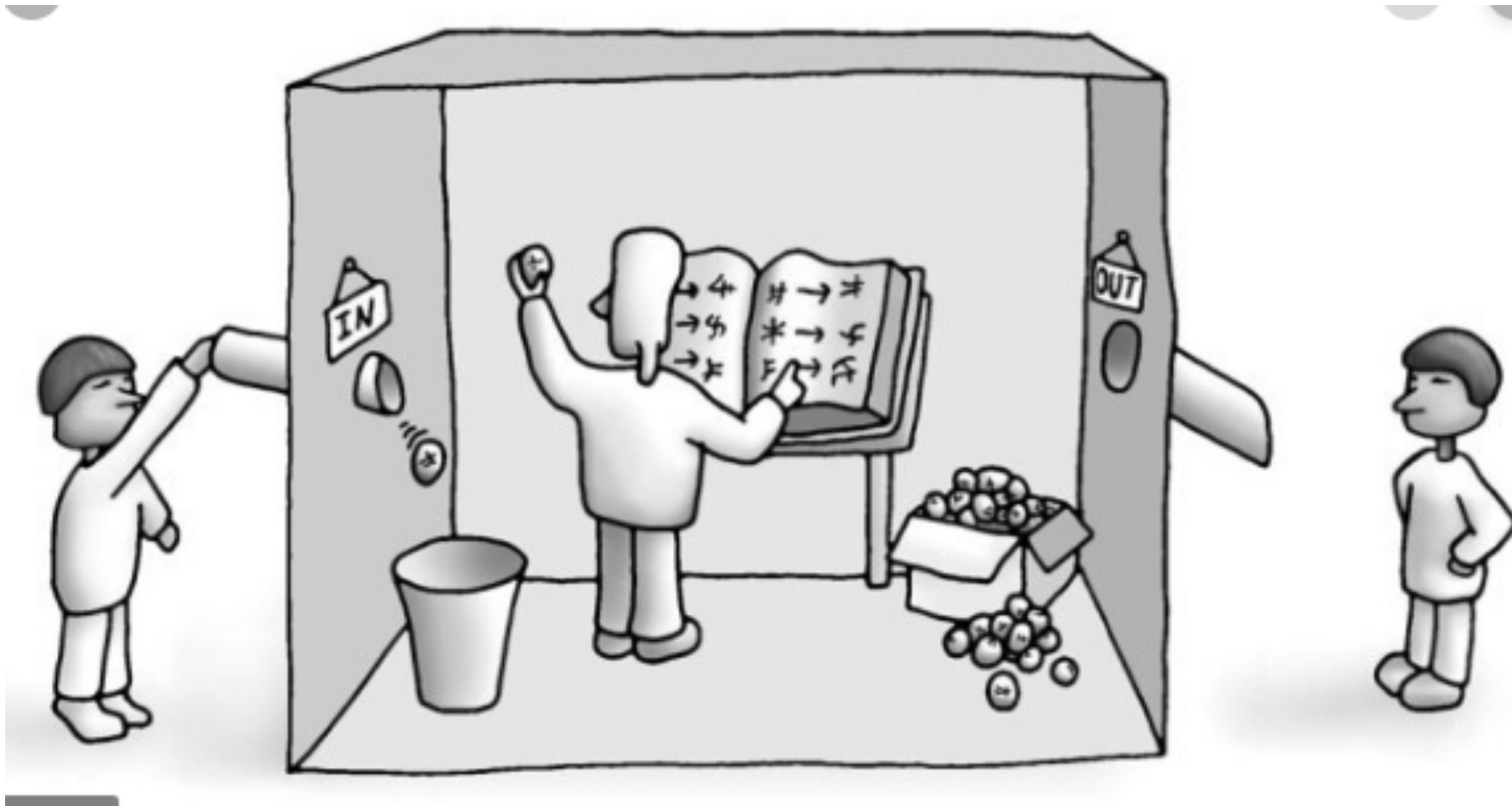


Gonzo data science:

Sunspring – фильм по сценарию нейронной сети

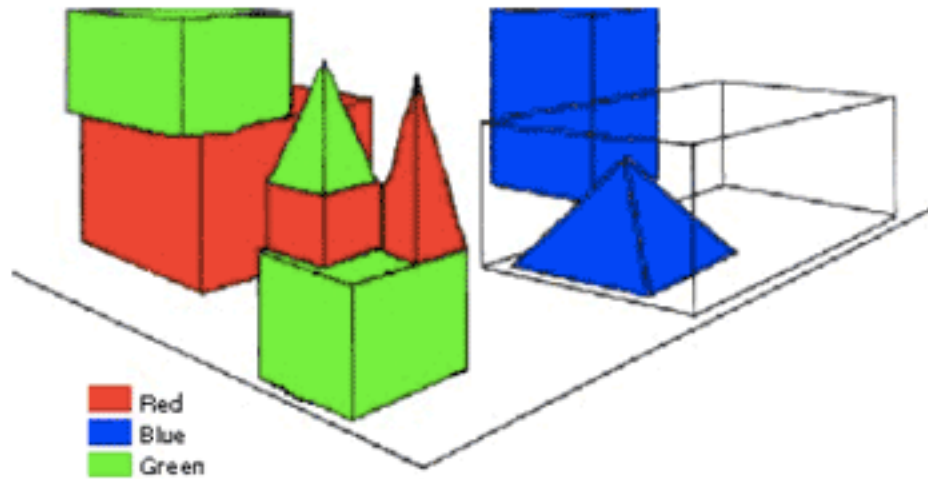


# Опять китайская комната



Что мы моделируем? Нужна ли модель без познания?

# Shrdlu Т.Винограда



**Person:** Pick up a big red block.

**Computer:** OK.

**Person:** Grasp the pyramid.

**Computer:** I don't understand which pyramid you mean.

Марксистский принцип  
**Праксис** – мы должны  
учиться, двигаясь в ту  
сторону, куда должны идти



УТОМЛЕННОСТЬ

# Бессмысленные игры в доказательную медицину

Фрустрация 4

Nan Z. Da. The computational case against computational literary studies, *Critical Inquiry*, 2019.

- cultural analytics
- literary data mining
- quantitative formalism
- literary text mining
- computational textual analysis computational criticism
- algorithmic literary studies



**what is robust is obvious and  
what is not obvious is not  
robust**

# Основные претензии Nan Z. Da

- the semantic reduction of the meaning of form and formalism to trackable units
- a study of the patterns made by trackable things
- argumentative arbitrariness
- very little explanatory power

Все исследования и методы можно свести к:

**basic measurements and statistical representations of overlapping vocabulary**

network analysis, digital mapping, linear and nonlinear regressions, topic modeling, topology, entropy—**are just fancier ways of talking about word frequency changes**

# Основные претензии Nan Z. Da

- Подсчет частотностей слов – это не моделирование
- Если мы работаем со статистикой то нужна нулевая гипотеза, более сильная, чем «слова не меняются»
- В промышленном дата майнинге всегда есть trade-off в пользу типового.

# Подход Nan Z. Da

- Разбор 15 работ
- Ошибки в разметке/в выделении признаков/в использовании методов
- Отличающиеся результаты при исправлении ошибок и репродуцировании
- Вывод о несостоятельности работ
- Вывод о несостоятельности подхода

# Ответы Nan Z. Da: Fottis Jannidis, Cultural analytics

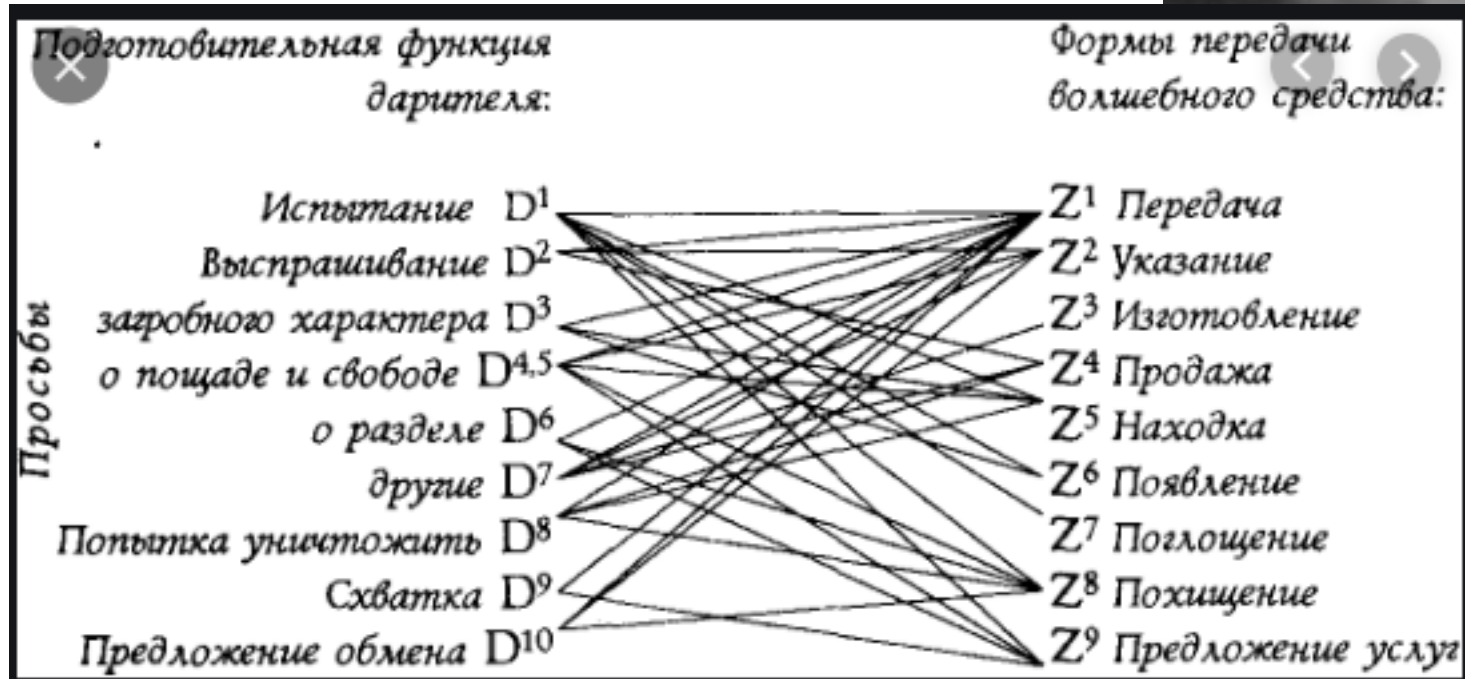
- Нельзя делать два утверждения сразу: в работах есть ошибки, статистический метод не работает.
- Частотность слов – это индикатор. Индикатор никогда не равнозначен явлению и не выражает всю его сложность
- Исследование данных – важная часть современной статистики, введенная в 1970х (John Tukey)

Tukey states "Exploratory data analysis can never be the whole story, but nothing else can serve as the foundation stone—as the first step."

# Новый тренд: event/characters tracking

- Bamman, Popat, Sheng “An annotated dataset of literary entities”, NAACL 2019
- Snigdha Chaturvedi, Mohit Iyyer, and Hal Daumé III. Unsupervised learning of evolving relationships between literary characters. In *Association for the Advancement of Artificial Intelligence 2017*
- Droog-Hayes, Wiggins, Purver “Automatic Detection of Narrative Structure for High-Level Story Representation” 5<sup>th</sup> AISB Symposium on Computational Creativity, 2018
- Sims, Park, Bamman “Literary Event Detection” ACL, 2019

# Не забываем про истоки: В.Я. Пропп

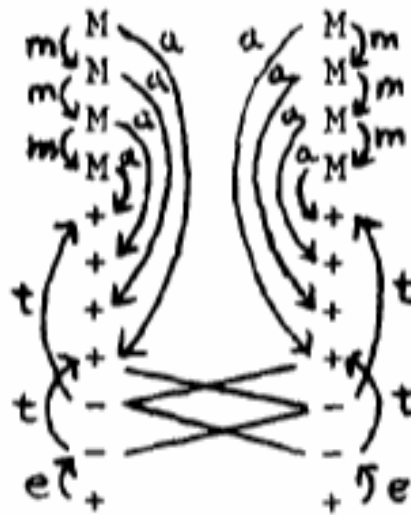


# Не забываем про истоки: Wendy Lehnert

COGNITIVE SCIENCE 4 (293-331) (1981)

## Plot Units and Narrative Summarization\*

wants to give gift  
wants gift  
wants money  
wants to sell hair  
sells hair  
gets money  
gets chain  
gives chain  
gets ornament  
regrets chain  
appreciation



wants to give gift  
wants gift  
wants money  
wants to sell watch  
sells watch  
gets money  
gets ornament  
gives ornament  
gets chain  
regrets ornament  
appreciation

Figure 34.

# Sims, Park, Bamman “Literary Event Detection” ACL, 2019

---

(1) Stephen Dedalus, displeased and sleepy, **leaned** his arms on the top of the staircase and **looked** coldly at the **shaking gurgling** face that **blessed** him, equine in its length, and at the light untousured hair, grained and hued like pale oak. (Joyce, *Ulysses*)

(2) My eyes **followed** his trim figure, richly though sombrely clad, then **fell** with a sudden **dissatisfaction** upon my own stained and frayed apparel. (Johnston, *To Have and to Hold*)

(3) He generally arrived in London (like the influenza) from the Continent, only he arrived unheralded by the Press; and his visitations set in with great severity. (Conrad, *The Secret Agent*)

---

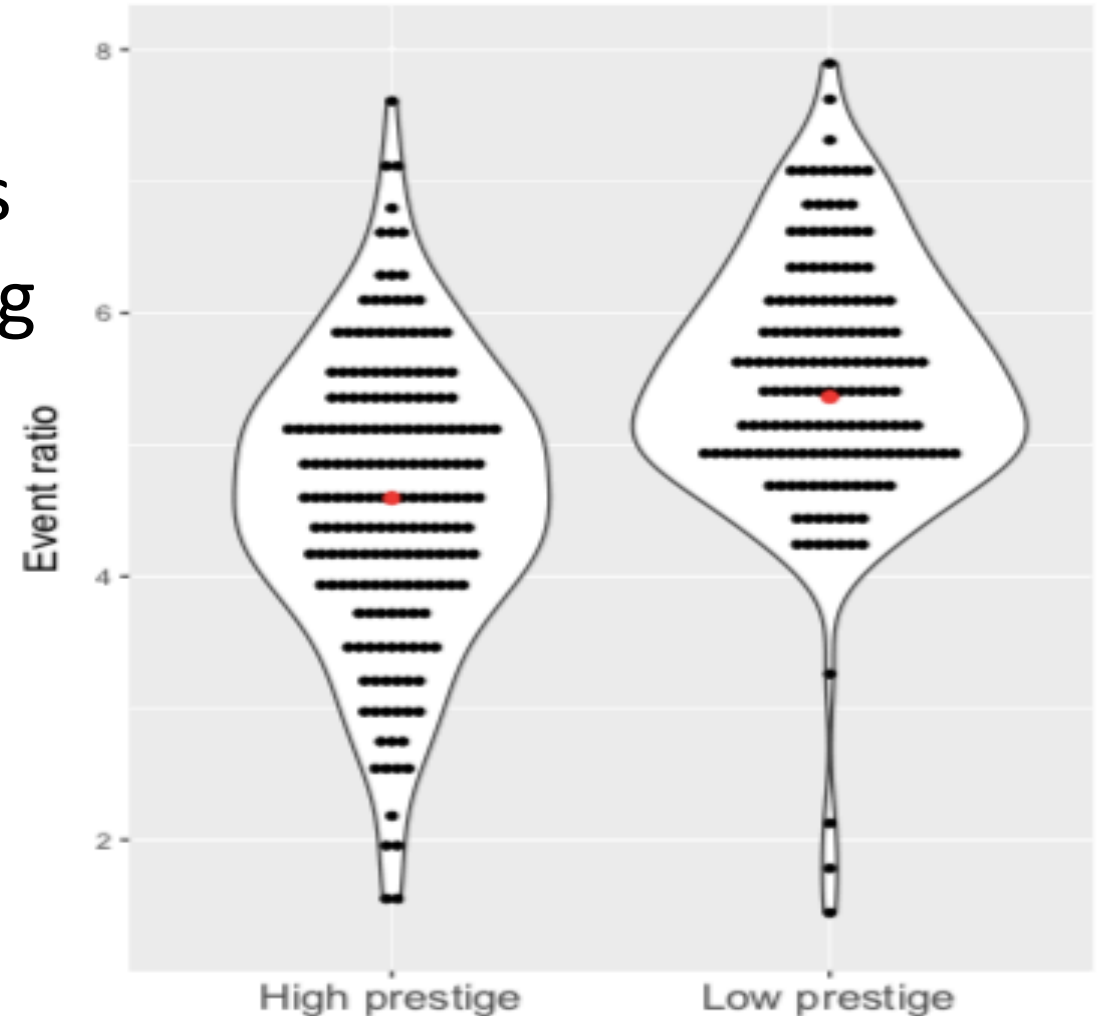
Table 1: Three annotation examples with tagged event triggers in bold and candidate triggers that would not be tagged underlined.

# Sims, Park, Bamman “Literary Event Detection” ACL, 2019

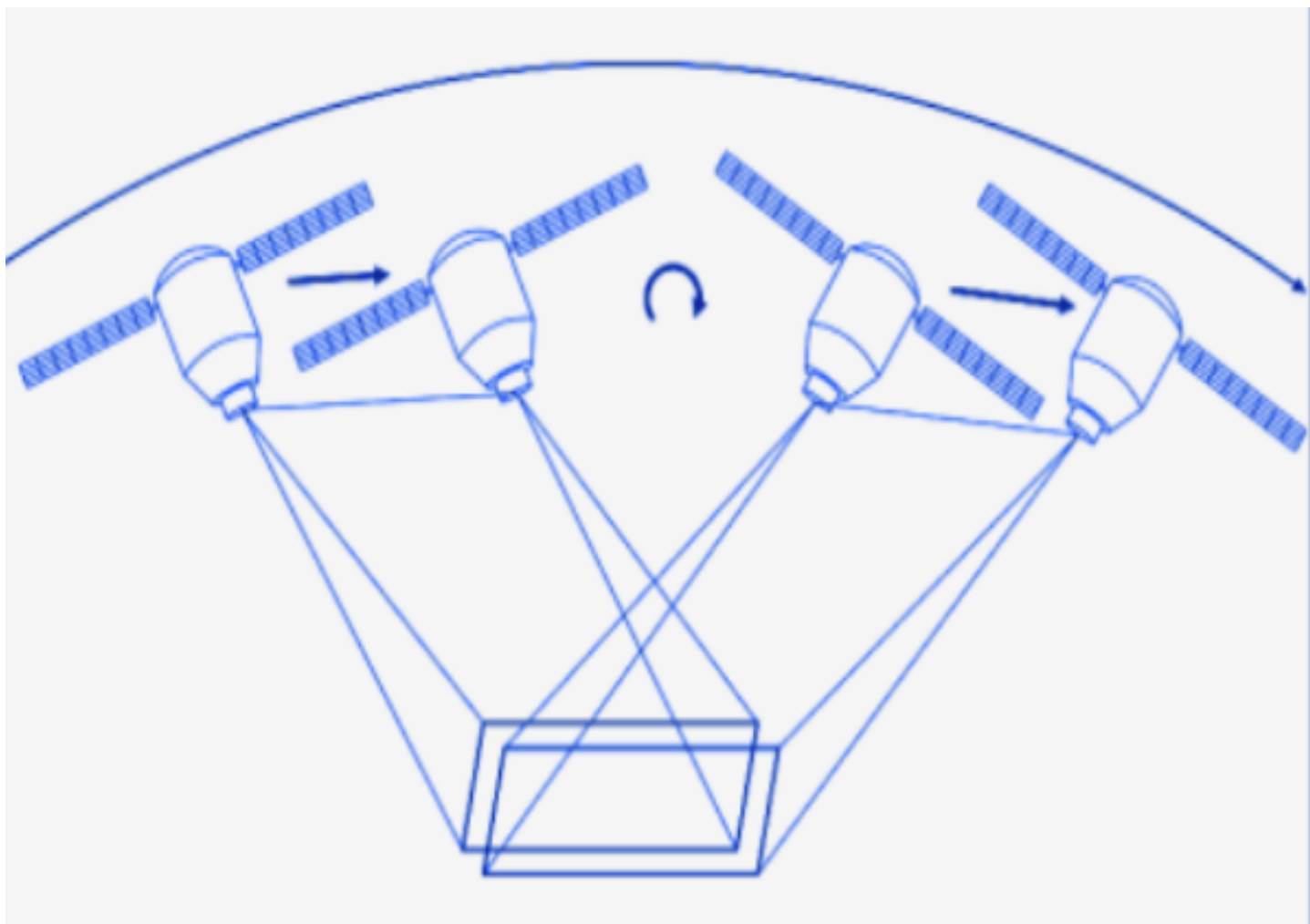
<b>Method</b>	<b>Precision</b>	<b>Recall</b>	<b>F</b>
Verbs only	17.7 [16.6-18.8]	76.2 [74.1-78.3]	28.7 [27.3-30.2]
Featurized	68.9 [66.2-71.7]	50.5 [48.0-52.9]	58.3 [56.1-60.4]
LSTM	66.6 [64.1-69.1]	60.5 [57.9-63.1]	63.4 [61.3-65.5]
BiLSTM	70.4 [67.8-72.9]	60.7 [58.0-63.4]	65.2 [63.1-67.3]
+ document context	74.2 [71.7-76.6]	58.8 [56.0-61.6]	65.6 [63.5-67.8]
+ sentence CNN	71.6 [69.1-74.1]	56.4 [53.8-59.0]	63.1 [61.0-65.1]
+ subword CNN	69.2 [66.6-71.6]	64.8 [62.2-67.3]	66.9 [64.8-68.9]
+ BERT	75.5 [73.3-77.8]	72.3 [69.7-74.8]	73.9 [72.0-75.7]
+ subword CNN	73.6 [71.2-75.8]	73.3 [70.8-75.7]	73.4 [71.5-75.2]

# Sims, Park, Bamman “Literary Event Detection” ACL, 2019

Put differently, prestigious novels don't have the same constraints as less prestigious ones in maintaining our attention through something *happening* in the narrative.



# Complex layer analysis



Б. В. ОРЕХОВ

## Башкирский стих XX века

КОРПУСНОЕ  
ИССЛЕДОВАНИЕ



Спасибо за внимание!

