



---

# ЧТО СЛОЖНЕЕ: ТОЧНЫЕ ИЛИ ГУМАНИТАРНЫЕ НАУКИ? КОРПУСНЫЙ АНАЛИЗ СТАТЕЙ ИЗ МЕЖДУНАРОДНЫХ ЖУРНАЛОВ

Смирнова Е.А., доцент департамента ин.яз,  
научный сотрудник НУЛ учебных корпусов

[easmirnova@hse.ru](mailto:easmirnova@hse.ru)

---

---

# ИЗУЧЕНИЕ ЯЗЫКОВОЙ СЛОЖНОСТИ

Сложность текстов носителей и неносителей языка (Hinkel, 2003; Ai & Lu, 2013; Lambert & Nakamura, 2019)

Связь сложности текста с уровнем владения языком и типом задания (Crossley & McNamara, 2012; Biber et al., 2016; Casal & Leeb, 2019)

Изменение уровня сложности текста после интенсивного курса обучения языку (Crossley & McNamara, 2014; Mazgutova & Kormos, 2015)

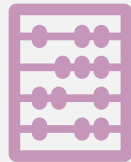
---

# ПОДХОДЫ К ОПРЕДЕЛЕНИЮ СЛОЖНОСТИ

- Использование специализированных компьютерных инструментов для определения сложности текста (например, the L2 Syntactic Complexity Analyser (L2SCA) (Lu, 2010), Coh-Metrix (Graesser et al., 2004; McNamara et al., 2014), SyB (Chen, 2018))
- Анализ синтаксических и лексических маркеров сложности текста (Gray, 2013; Nesi & Gardner, 2019)
- Комбинация двух подходов дает более полное представление о сложности текста



1 Есть ли различия в автоматически определенных показателях сложности текстов, относящихся к точным и гуманитарным наукам?



2 Различается ли частотность употребления маркеров синтаксической сложности текста в точных и гуманитарных науках? Наблюдаются ли различия между конкретными дисциплинами?

---

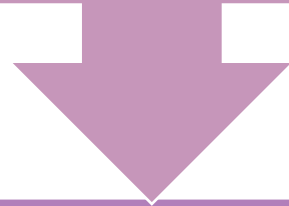
# ДААННЫЕ

Discipline	Number of texts	Number of words
Hard sciences		
Chemistry	16	97947
Physics	18	95852
Mathematics	13	98430
Engineering	17	99003
<b>Hard sciences total</b>	<b>64</b>	<b>391232</b>
Soft sciences		
Business Studies	10	95350
Linguistics	10	95603
History	10	99303
Political Science	11	93366
<b>Soft sciences total</b>	<b>41</b>	<b>383622</b>

---

# МЕТОДЫ

Тексты обработаны при помощи Syntactic Complexity Analyser - 14 показателей сложности, например, средняя длина клаузы, средняя длина предложения, соотношение количества предложений к количеству клауз и др.



Проанализировано употребление маркеров клаузальной и фразовой сложности, а именно: различных видов обстоятельственных придаточных предложений, изъяснительных придаточных, атрибутивных прилагательных, существительных, выступающих в роли определения и др.

---

# СТАТИСТИЧЕСКИЙ АНАЛИЗ

- Значимость разницы - **логарифмическое правдоподобие** (log likelihood):  $LL \geq 6.63$  означает, что результаты значимы на уровне 1% ( $p < 0.01$ );  $LL \geq 3.84$  означает, что результаты значимы на уровне 5% level ( $p < 0.05$ ) (Smith, 2009).
- Величина разницы – **размер эффекта** (%DIFF):  
сравнивается нормализованная частотность и указывается процентное соотношение различия.

---

# РЕЗУЛЬТАТЫ (1)

- В гуманитарных науках используются более длинные предложения и клаузы, а также наблюдается более высокий показатель сложности предложения
- Показатели подчинительной связи (subordination indexes) также выше в гуманитарных науках, чем в точных, а показатели сочинительной связи значительно различаются в разных дисциплинах
- Самые длинные предложения и клаузы используются в статьях по истории; тексты по химии оказались ближе всего к гуманитарным наукам в отношении длины предложения и клаузы, а также использования сложноподчиненных предложений



---

# РЕЗУЛЬТАТЫ (2)

## ➤ Маркеры клаузуальной сложности:

- Обстоятельственные придаточные предложения чаще используются в точных науках, но, в основном, за счет математики, где часто употребляются придаточные условия (if), места (where) и причины (since). Этот вывод противоречит результатам, полученным автоматически.
- В гуманитарных науках чаще употребляются обстоятельственные придаточные времени (when), образа действия (as if) и контраста (while).

## ➤ Маркеры фразовой сложности:

- Частотность существительных и существительных, использующихся в качестве определений, значительно выше в точных науках, что также противоречит результатам автоматического анализа.
- Атрибутивные прилагательные и генетивы с предлогом *of* преобладают в гуманитарных науках.

---

# ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ (1)

ИВ1: Есть ли различия в автоматически определенных показателях сложности текстов, относящихся к точным и гуманитарным наукам?

В целом, показатели сложности выше в гуманитарных науках; однако, была выявлена гетерогенность в показателях сочинительной связи, что может свидетельствовать о высокой степени дисциплинарной специфичности данного показателя.

---

---

## ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ (2)

ИВ2 Различается ли частотность употребления маркеров синтаксической сложности текста в точных и гуманитарных науках? Наблюдаются ли различия между конкретными дисциплинами?

- В целом, клаузуальная сложность более выражена в точных науках, в то время как фразовая сложность, реализуемая за счет использования атрибутивных прилагательных и генетивов с предлогом *of*, в большей степени наблюдается в гуманитарных дисциплинах. Этот вывод противоречит результатам, полученным в других исследованиях, которые выявили, что клаузуальная сложность играет более важную роль в гуманитарных науках (Gray, 2015; Staples et al., 2016; Nesi & Gardner, 2019).

---

Спасибо  
за внимание!

