

# Семинар НУЛ учебных корпусов

26.01.2021

# План

- План семинаров
- Наш корпус
- Оценка синтаксической сложности текста
- Проверка инструмента на точность
- Результаты анализа маркеров, практическое применение

# План семинаров

<b>февраль</b>	<b>Совместный семинар НУЛ и Центра цифровых трансформаций. Сравнительный анализ использования маркеров научного стиля в учебных и экспертных корпусах по разным дисциплинам (Смирнова Е.А., Туляков Д.С., Авраменко И.А.)</b>
<b>март</b>	<b>?</b>
<b>апрель (начало)</b>	<b>Совместный семинар НУЛ и Центра цифровых трансформаций. Тема? (стажер исследователь)</b>
<b>май</b>	<b>Совместный семинар НУЛ и Центра цифровых трансформаций. Способы измерения лексической сложности текста (Яркова В.В.)</b>

# Наш корпус

- Нужно его где-то разместить, добавить описание и **функцию выбора дисциплины**, как здесь <https://micusp.elicorpora.info/>
- Конкурс на лучшее название

Предложение: CRUSP(P) – Corpus of Russian Undergraduate Students' Project Proposals/Papers

- Проверить тексты еще раз

# Оценка сложности текста

2 варианта:

- 1. Поработать с Inspector:

- Отобразить метрики для статей и RPs (диплом И. Пантелеевой <https://docs.google.com/document/d/1NjL516g8bG1yf4Je642DNAs0UusYXHHQ0I3LEpiYWS0/edit#> )

- Добавить в Inspector возможность выбора типа текста

- Прогнать экспертные корпуса через Inspector, посмотреть данные

- Дописать фидбек для статей с учетом полученных данных

- Документация [https://github.com/lcl-hse/lab\\_inspector](https://github.com/lcl-hse/lab_inspector)

# Оценка сложности текста

- 2. Сделать дополнение к нашему инструменту на Gate по оценке сложности:
- Добавить маркеры клаузуальной и фразовой сложности (Staples et al., 2016): *wh*-complement clauses, *that*-clause complements in predicates, attributive adjectives, premodifying nouns and *of*-genitives
- Добавить комментарии для пользователей

# Тексты на проверку

- По 2 текста learner & expert, 4 эксперта, каждый текст проверяется дважды, т.е. по 2 текста на эксперта
- 1- AL, AE 2- AL, BE 3 –BL, AE 4 – BL, BE
- Предложения по маркерам, которые вызывают сомнения

Например, generic one в learner

# Результаты анализа маркеров

## **Underused by learners**

- Participle II plus noun
- noun + ing clause
- noun plus relative clause
- predicative adjectives
- specific adverbs
- degree adverbs
- linking adverbials
- Ing-cl-s contr by adj pred
- concessive clauses
- preposition + which

## **Overused by learners**

- coordination tags
- subj pred to clause
- extraposed to-clauses
- Purp&conc adv-ls



# Работа над маркерами

- Общая информация о проекте и о сборе данных
- функция маркера в академическом дискурсе
- упражнения на основе корпусных данных (error correction, transformation, translation, etc.)
  
- Вопросы для обсуждения:
  - Один семинар или серия (3 семинара)?
  - Название?
  - Корпусы здесь  
<https://drive.google.com/drive/folders/1f0x0W1p0vOkCov68n4xCRqnUxp0O01nM?usp=sharing>

# Further research

- Полученные данные нужно использовать для статей и конференций
- Конференция Corpus Linguistics 2021  
<https://easychair.org/cfp/CL2021>
- Deadline – February, 12