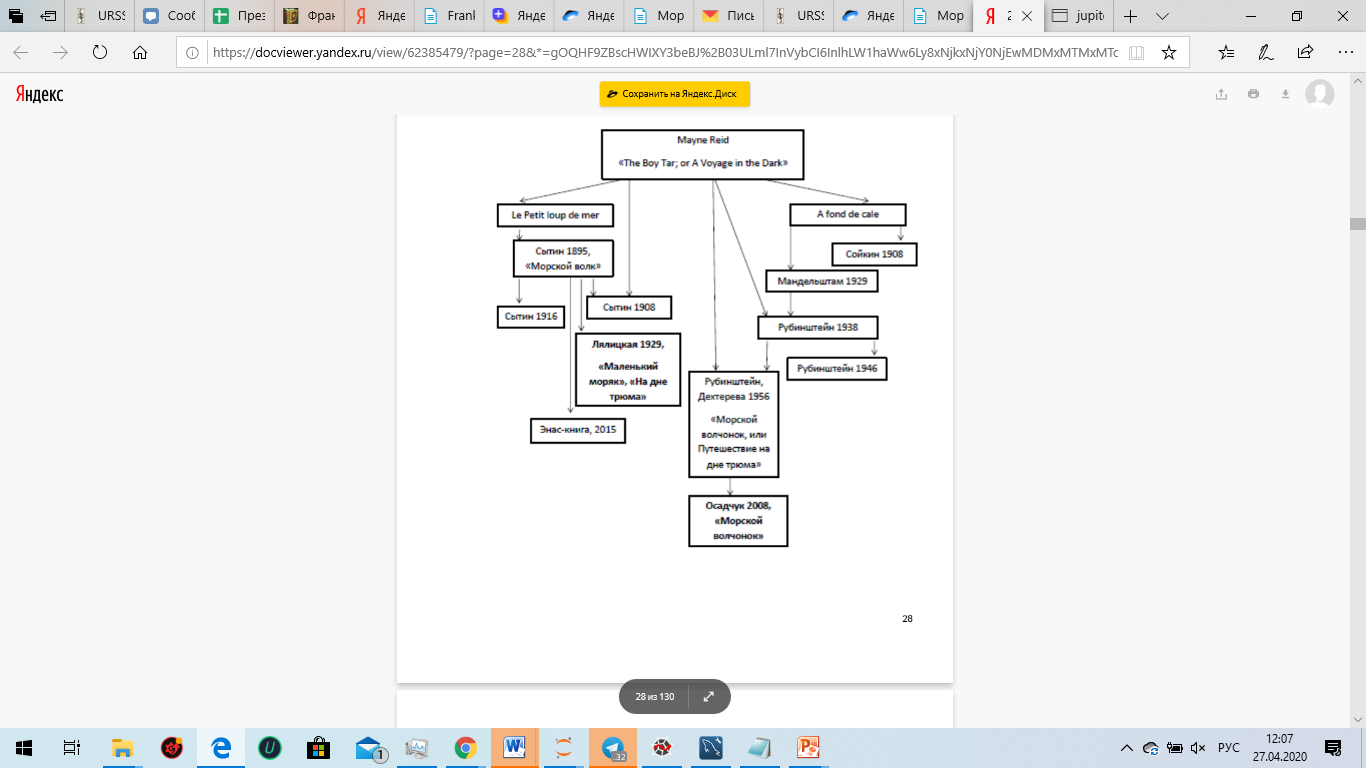
Тема проекта:

Анализ русских переводов «Морского волчонка» Майн Рида: readability и изменения при редактировании

Описание данных

Во время написания бакалаврской ВКР я выяснила, что история перевода этого текста весьма запутанная.

Она изображена на схеме: 

Доступен полный текст следующих переводов:

Рубинштейн 1938/1946 (различия крайне незначительны);

Рубинштейн, Дехтерева 1956;

Издательство Энас-книга 2015;

Аудиокнига Осадчук 2008 (расшифрована мной)

Для следующих текстов оцифрованы отдельные фрагменты:  
издательство Сытина 1895,

Издательство Сойкина 1908,

Мандельштам 1929,

Лялицкая 1929

Также есть таблица, в которой выписаны фрагменты разных переводов (предположительно все фрагменты), содержащие метатекст.

**Гипотезы**

1. Readability

Предполагаю, что по показателям readability (средняя длина слова и предложения, количество причастий и деепричастий, отглагольных существительных) дореволюционные полные переводы и перевод Мандельштама отличаются от перевода Рубинштейна, а адаптированные тексты отличаются от полных текстов, на основании которых они подготовлены.

1. Предполагаю, что при подготовке адаптированных для детей переводов редакторы были больше склонны исключать фрагменты определенных типов (метатекст, служебные слова)

Таким образом, предполагается сопоставить:

- метрики readability для доступных полных текстов и всех частично доступных текстов,

- количество случаев употребления метатекста для переводов, для которых выписаны данные фрагменты.

Также предполагается с помощью библиотеки python difflib получить данные о том, какие слова добавлялись/исключались при редактировании текста Дехтеревой и Лялицкой. В случае, если качество выявления вставок и исключений окажется приемлемым, планируется определить, есть ли отклонения от случайного распределения/распределения, пропорционального доле в исходном тексте, у определенных категорий слов (служебные слова, причастия и деепричастия, слова с семантикой непосредственного восприятия органами чувств, глаголы в форме второго лица).

Основные методы: t-test, ANOVA, хи-квадрат