



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

ПОРТАЛ ДЛЯ АВТОМАТИЗИРОВАННОЙ ОЦЕНКИ АКАДЕМИЧЕСКОГО ТЕКСТА

НУЛ УЧЕБНЫХ КОРПУСОВ



О ПРОЕКТЕ

Научно-исследовательская группа «Разработка программного обеспечения для проведения корпусных исследований английского языка» г. Пермь работала с 2017 по 2019 г.

Научно-исследовательская группа «REALES для реально необходимых слов» г. Москва работала с 2017 по 2019 г.

Научно-учебная лаборатория учебных корпусов работает с 2019 г.



Типы анализируемых текстов: студенческие эссе, ВКР, научные статьи

Подход: сравнение текста пользователя с «идеальным» текстом соответствующего жанра –

эссе сравниваются с эссе, получившими высокий балл
статьи и ВКР - со статьями, опубликованными в ведущих журналах по той или иной дисциплине (6 экспертных корпусов)



РАЗДЕЛЫ ПОРТАЛА

- 1) оценка текста на предмет использования маркеров академического стиля
- 2) оценка правильности текста (Adwisor)
- 3) оценка сложности текста (на основе Inspector (Lyashevskaya et al., 2020))



GATE - General Architecture for Text Engineering

- Среда для решения лингвистических задач
- Открытый исходный код, java
- Набор готовых инструментов
- Модули расширения (Plugins)
- Графический интерфейс пользователя
- Программная библиотека
- Хорошая документация
- Облачная версия (платная)



МАРКЕРЫ АКАДЕМИЧЕСКОГО СТИЛЯ: ОТБОР

- Анализ методической литературы по академическому письму (e.g. Swales & Feak, 2004; Siepmann et al., 2011; Wallwork, 2016) – вопрос о специфичности предлагаемых маркеров
- Литература по корпусным исследованиям академического дискурса (Biber et al., 1999; Biber & Gray, 2016) – маркеры, характерные для научных текстов



МАРКЕРЫ АКАДЕМИЧЕСКОГО СТИЛЯ: ПРИМЕРЫ

Adjectives and adjective phrases		
Feature	Example	Comment
Attributive adjectives	<i>Robust data; a valid measure</i>	an adjective functioning as a premodifier before a noun
Specific predictive adjectives	<i>Seems unlikely; appears reasonable</i>	an adjective that does not premodify a noun; it occurs in the subject predicative position, following a copular verb
Derived adjectives	<i>Obvious, subjective</i>	Adjectives with specific derivational suffixes



ПРИМЕР РАЗМЕТКИ ТЕСТА

Register

Create a new account.

Full name

Email

Password

Confirm password

Register

© Portal Prototype

Documents

Create New

Name	Uri	CorpusEntity	
test_doc	demo.txt	Super Cool Texts	Edit Details Delete
new_article	new_article.txt	Super Cool Texts	Edit Details Delete

NewText

© Portal Prototype

Research

Corpuses

Documents

Annotations

Corpuses	Articles	Text	To
CorpusName Super Cool Texts	Article Title test_doc new_article NewText	<p>ORGANIZATION STRATEGIC ANALYSIS This paper is aimed at providing answers to the list of research questions on the subject of the strategic analysis of the company. Within the framework of these questions there was developed the following theoretical purpose of the current proposal: to structure the available data on the strategic analysis in context of the strategic management. The above issue is addressed in the 5 following parts of the article: 1) Introduction 2) Literature Review 3) Methodology 4) Results Anticipated 5) Conclusion The first part of the proposal provides explanation of the research questions and the main purpose of the study, states delimitations, indicates the professional significance and presents the definitions of key terms. The second part of the proposal introduces the theoretical and practical background of the study. The third part represents a sort of practical recommendations for strategic analysis of a firm. The fourth part of the proposal presents the possible results of the practical part of our study. And the last part is the findings of current paper. Introduction Background of the study. Market economy is very unstable phenomenon. It is based on consumer demand, rather than at a pre-approved plan. The former can vary depending on external and internal environment of the organization. Therefore, when it</p>	<input checked="" type="checkbox"/> Paragraph <input checked="" type="checkbox"/> Split <input checked="" type="checkbox"/> Sentence <input checked="" type="checkbox"/> Token <input checked="" type="checkbox"/> Postpositive <input checked="" type="checkbox"/> Noun <input checked="" type="checkbox"/> Abstract_s <input checked="" type="checkbox"/> Prepositive <input checked="" type="checkbox"/> Space <input checked="" type="checkbox"/> PassiveVoic <input checked="" type="checkbox"/> Present <input checked="" type="checkbox"/> Desemantis <input checked="" type="checkbox"/> Abstract_se <input checked="" type="checkbox"/> Logic_conn <input checked="" type="checkbox"/> IPronoun <input checked="" type="checkbox"/> PersonalPro <input checked="" type="checkbox"/> Past <input checked="" type="checkbox"/> WePronoun <input checked="" type="checkbox"/> Future <input checked="" type="checkbox"/> Complex_p <input checked="" type="checkbox"/> OrSuffix <input checked="" type="checkbox"/> Complex_c <input checked="" type="checkbox"/> YouHeShe

© Portal Prototype



ОЦЕНКА СЛОЖНОСТИ ТЕКСТА

2 понимания сложности:

- широкое использование в тексте специфических синтаксических конструкций, например, придаточных предложений (Biber & Gray, 2016)

- лексическое и грамматическое разнообразие и уровень сложности, которое можно оценить при помощи компьютерных инструментов (Bulté & Housen, 2012)



РАБОТА INSPECTOR'A

- студенческие эссе анализируются по 59 параметрам
- Парсер UD-Pipe
- сравнение показателей проверяемого эссе с показателями лучших эссе
- Наглядное представление результатов

The screenshot shows the 'Inspector' web application interface. The browser address bar displays 'linghub.ru/inspector_realec/'. The interface is divided into two main sections: 'Essay' and 'Result'.

Essay: A text box contains the following text: "Online customer reviews, being an essential factor that determines success or failure in business, namely, in the tourism industry, demand close attention since the investigation of this type of discourse might bear some implications both for language specialists and for hotel managers. This paper is a quantitative and qualitative study of evaluation devices occurring in the corpus of hotel reviews from Booking.com. The analysis is based on the methods associated with corpus linguistics and Systemic Functional Linguistics (Halliday 1994; Eggins 2004), specifically, the Appraisal framework (Martin & White, 2005) which is applied to a new type of discourse, the discourse of hotel ratings. The study focuses on the categories of Attitude, Graduation and Engagement suggested by Martin and White and investigates inscribed and invoked evaluation. New subcategories within the Attitude dimensions of the Appraisal framework were created to adapt them to the corpus under study, therefore, contributing to theorising evaluation. The practical result of the study is a list of evaluation devices employed in hotel reviews which might be useful for further corpus analyses and for designing systems for automated analysis of customer evaluation."

Result: A message states: "Your essay is not good enough. The probable grade in the exam is 7. To improve your writing skills look at the recommendations below."

Statistics: A table provides the following data:

Number of words	184
Number of lemmas	109
Number of sentences	8

Information about academic words and linking phrases: A table provides the following data:

Number of academic words	9
Number of linking phrases	11

At the bottom of the interface, there are two buttons: "Файл не выбран" (File not selected) and "Inspect".

Below the statistics, a preview of the analyzed text is shown with key words highlighted in color: "factor" (red), "or" (orange), "since" (green), "discourse" (red), "and" (green), "methods" (red), and "and" (green).



ПОКАЗАТЕЛИ СЛОЖНОСТИ

- показатели, используемые в L2 Syntactic Complexity Analyzer (Lu, 2010)
- 14 индексов, выбранных из работ Wolfe-Quintero et al. (1998) и Ortega (2003)
- специфичны для академических текстов и для текстов неносителей языка



ПОКАЗАТЕЛИ СЛОЖНОСТИ

- 5 групп метрик:

1) длина синтаксической единицы (например, средняя длина клаузы)

2) сложность предложения (соотношение количества клауз и количества предложений)

3) подчинительные конструкции (например, соотношение количества придаточных предложений и общего количества клауз)

4) сочинительные конструкции (например, соотношение сочинительных конструкций и общего количества клауз)

5) определенные структуры (например, соотношение существительных с определениями и общего количества клауз)

(Lu, 2010: 479)



Литература

- Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E., & Quirk, R. (1999). *Longman grammar of spoken and written English*. London: Longman.
- Biber, D., & Gray, B. (2010). Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *Journal of English for Academic Purposes*, 9(1), 2-20.
- Biber, D., & Gray, B. (2016). *Grammatical complexity in academic English: Linguistic change in writing*. Cambridge, England: Cambridge University Press.
- Bulté, B. & Housen, A. (2012). Defining and operationalising L2 complexity. In A. Housen, F. Kuiken & I. Vedder (Eds.), *Dimensions of L2 Performance and Proficiency: Complexity, Accuracy and Fluency in SLA* (pp. 21-46). Amsterdam: John Benjamins.
- Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, 15(4), 474-496.
- Lyashevskaya, O., Panteleeva, I., Vinogradova, O. (2020). Automated Assessment of Learner Text Complexity (upcoming).
- Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics*, 24, 492–518.
- Siepmann, D., Gallagher, J. D., Hannay, M., & Mackenzie, J. L. (2011). *Writing in English: a Guide for Advanced Learners*. Tuebingen: UTB.
- Swales, J. M., & Feak, C. B. (2004). *Academic Writing for graduate students: Essential tasks and skills* (Vol. 1). Ann Arbor, MI: University of Michigan Press.
- Wallwork, A. (2016). *English for writing research papers*. New York: Springer.
- Wolfe-Quintero, K., Inagaki, S., & Kim, H. Y. (1998). *Second language development in writing: Measures of fluency, accuracy, and complexity*. Honolulu, HI: University of Hawaii Press.



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ