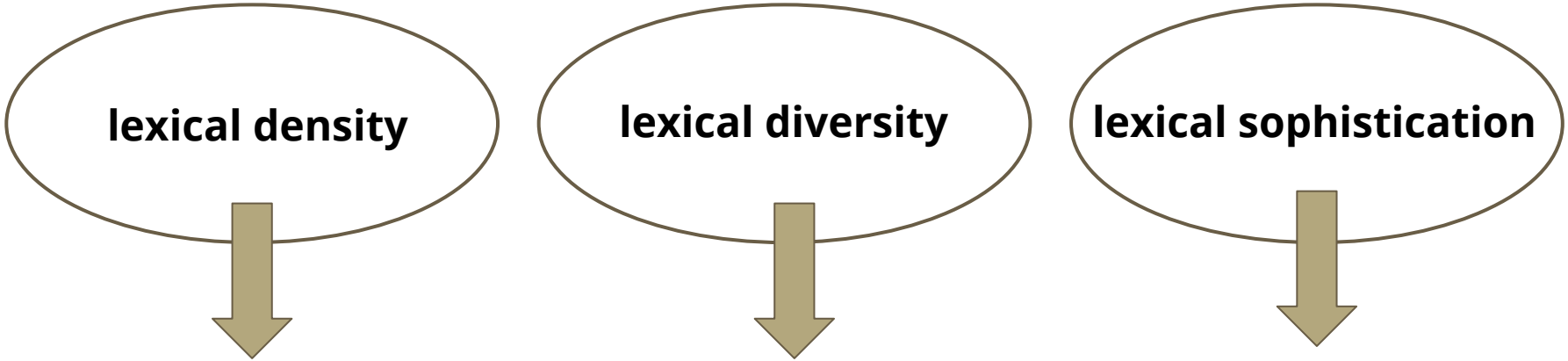
Способы измерения лексической сложности текста

Яркова В.В.,
стажер-исследователь научно-учебной лаборатории
учебных корпусов НИУ ВШЭ — Пермь

Определение лексической сложности



the ratio of the number of lexical (open-class) words to the total number of words in a text

the range of a learner's vocabulary

the type-token ratio

the ratio of the number of advanced words to the total number of words

Способы измерения лексической сложности

- **Обработка вручную**

- **Использование цифровых инструментов (!)**

  Coh-Metrix (Graesser et al, 2004)
  Lexical Complexity Analyzer (Ai & Lu, 2010)
  Text Inspector (Bax, 2012)
  TAALED and TAALES (Kyle & Crossley, 2015)
  Pitt Lexical Toolkit (Naismith, forthcoming)
  Lex Complexity Tool (Bottini, forthcoming)

Какие могут быть проблемы?

Отсутствие доступа

Платный контент

Отсутствие возможности измерения **всех параметров** ЛС

Lexical Complexity Analyzer

- Бесплатная и стабильная программа
- Измеряет **все 3 параметра** лексической сложности
- Можно как скачать, так и работать **онлайн** (!)

Не поддерживает ОС Windows

Нужны навыки программирования

Отсутствие автоматической обработки текстов (лемматизация и разметка)

Материал для анализа

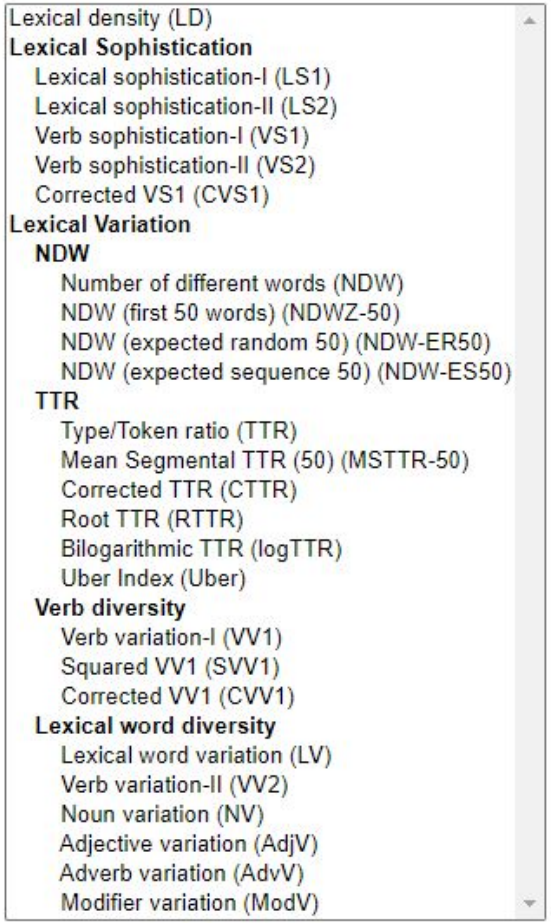
Корпус докладов русскоговорящих студентов
с научной конференции iTELL

Материал за 2019-2020 гг.
52 доклада (11/41)
83508 слов

Материал для сравнения

MICASE vs BASE

Презентации студентов (11)
Междисциплинарность
143369 слов



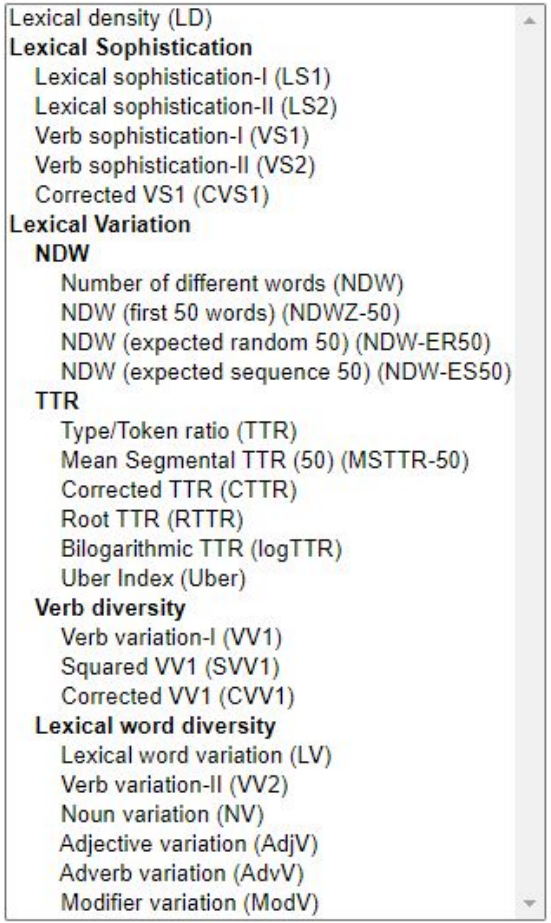
Полученные результаты

- **Lexical Density Values (Mean)**

  iTELL 0.487
  MICASE 0.483

- **Lexical Diversity Values, NDW (Mean)**

	NDW	NDW	NDWERZ	NDWESZ
iTELL	422.673	36.942	37.471	35.13
MICASE	1027.565	38.130	40.352	36.591



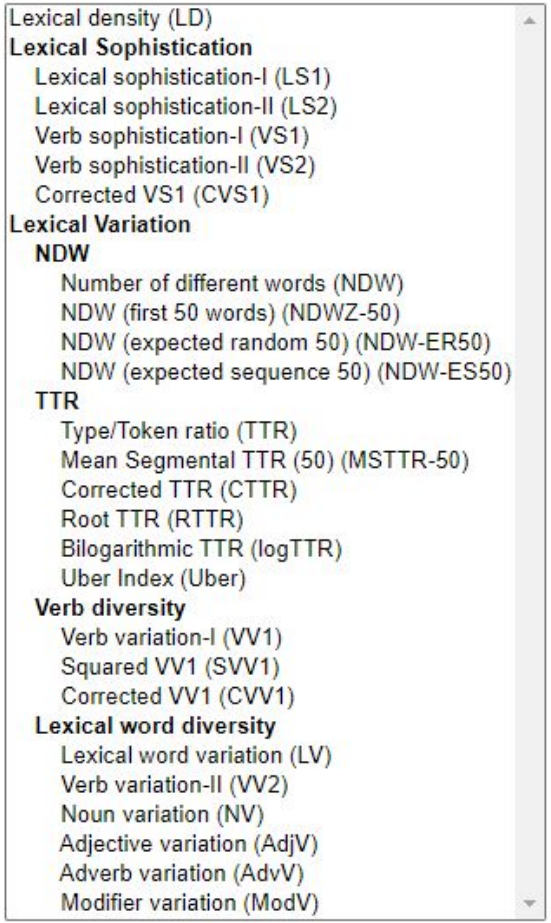
Полученные результаты

- **Lexical Diversity Values, TTR (Mean)**

	TTR	MSTTR	CTTR	RTTR	LOGTTR	UBER
iTELL	0.245	0.706	7.109	10.05	0.81	17.13
MICASE	0.146	0.726	8.54	12.078	0.781	17.588

- **Lexical Diversity, VD (Mean)**

	VV1	SVV1	CVV1
iTELL	41.408	4.516	0.408
MICASE	50.529	4.966	0.263



Полученные результаты

- **Lexical Diversity Values, LWD (Mean)**

	LV	VV2	NV	ADJV	ADVV	MODV
iTELL	0.525	0.095	0.368	0.082	0.039	0.12
MICASE	0.252	0.057	0.296	0.052	0.028	0.08

- **Lexical Sophistication Values (Mean)**

	LS1	LS2	VS1	VS2	CVS1
iTELL	0.312	0.304	0.109	2.004	0.952
MICASE	0.286	0.39	0.06	3.11	1.2

Перспективы исследования

- Изучение лексической сложности в динамике
- Совершенствование педагогических методик (как для неязыковых, так и языковых направлений подготовки)
- Изучение особенностей устной академической речи, создание инструмента для анализа

Спасибо за внимание!