Сравнительный анализ использования маркеров научного стиля в учебных и экспертных корпусах по разным дисциплинам

Смирнова Е.А.
Туляков Д.С.
Авраменко И.А.

Introduction

2016 г. - начало работы по созданию списка маркеров

2017-2019 гг. - работа НУГ

2020-2022 - НУЛ учебных корпусов

Подходы к отбору маркеров

1. Изучение методической литературы по академическому письму

   Проблема: рекомендуемые конструкции зачастую просто характерны для высокого уровня владения языком, а не для академического дискурса

1. Маркеры академического стиля, отобранные в результате корпусного анализа большого объема корпусных данных, относящихся к разным жанрам (Biber et al., 1999; Biber & Gray, 2016)

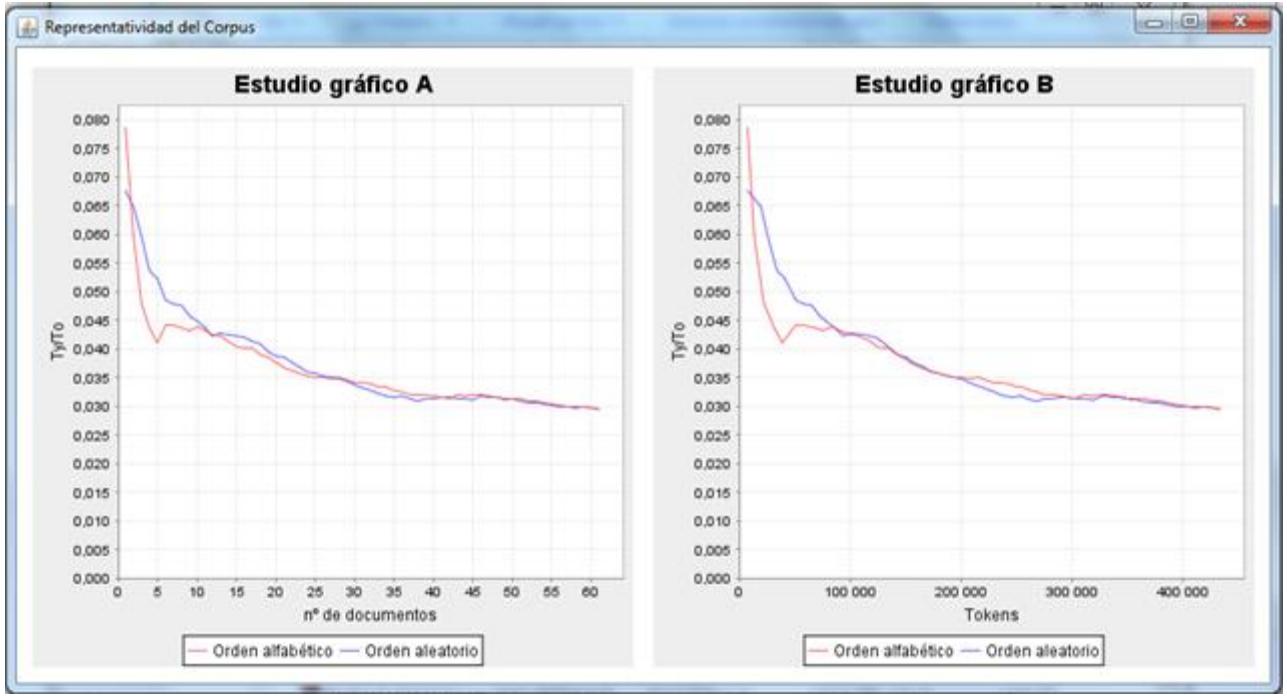
   Плюс: эмпирическое обоснование принадлежности маркеров к академическому дискурсу

Groups of Features	Examples
1 NOUNS AND NOUN PHRASES	
1.1 NOUN TOKENS	Nominalisations (e.g. education, happiness)
1.2 DETERMINER TOKENS	Demonstrative determiners (e.g. these costs)
1.3 NOUN PHRASES [token combinations]	noun + prepositional phrase (e.g. the hallmark of)
2 ADJECTIVES AND ADJECTIVE PHRASES	
2.1 ADJECTIVE TOKENS	predicative adjectives (e.g. are not independent)
3 VERBS AND VERB PHRASES	
3.1 VERB TOKENS	Specific prepositional verbs (e.g. occur in)
3.2 TOKEN COMBINATIONS WITH VERBS	Passive voice (e.g. is used to)
4 ADVERBS AND ADVERBIALS	Specific amplifiers (e.g. extremely, highly)
5 DEPENDENT CLAUSE FEATURES	Extraposed to-clauses (e.g. It is a challenge to determine)
6 OTHER FEATURES	Semi-determiners (e.g. certain, such)

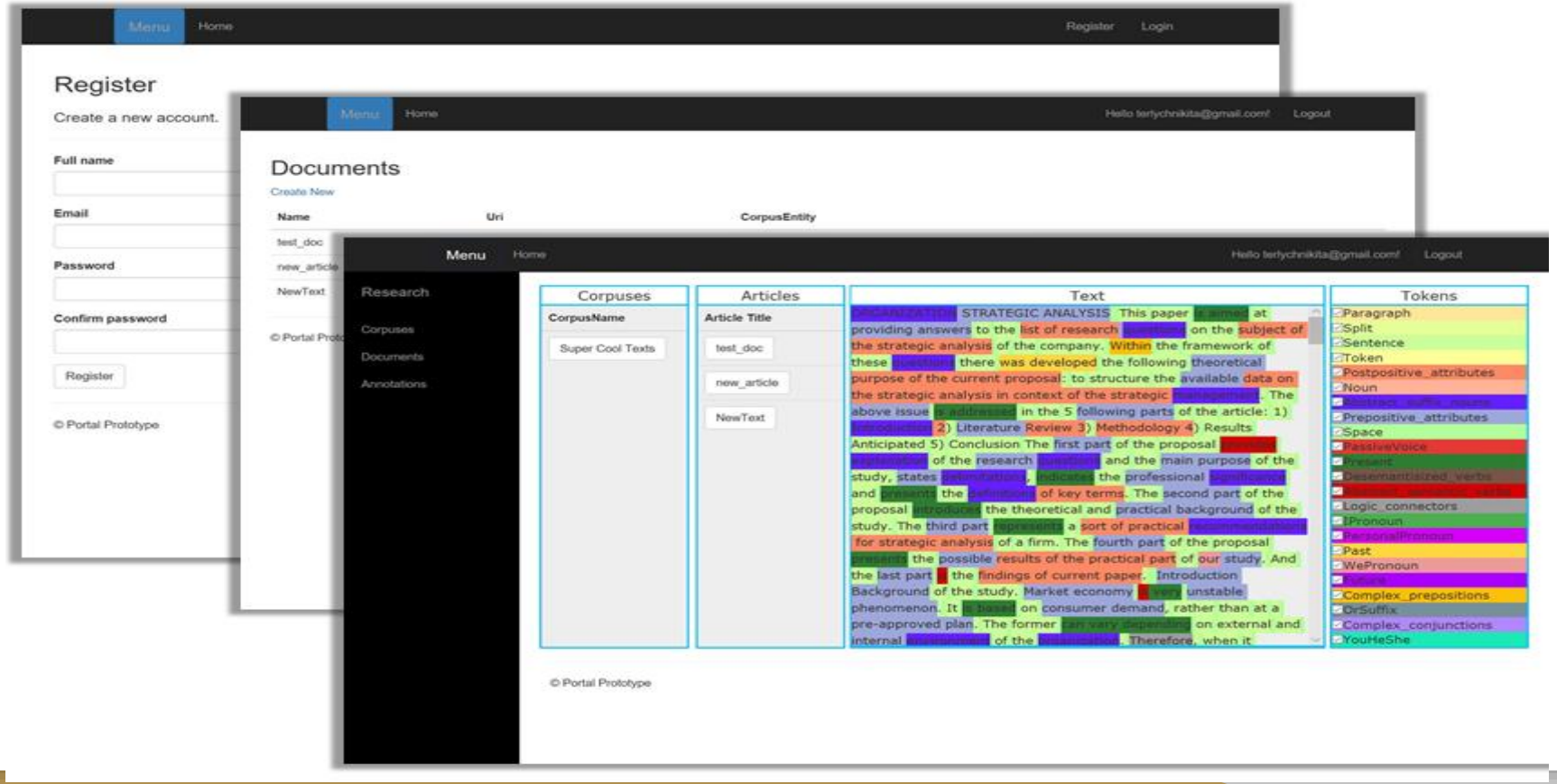
Data

Corpus	Number of tokens	Number of texts
Economics		
Expert	~ 654,000	57
Learner	~ 146,000	68
Management		
Expert	~ 683,000	61
Learner	~ 130,000	58
History		
Expert	~ 622,000	65
Learner	~ 91,000	43
Political Science		
Expert	~ 655,000	73
Learner	~ 57,000	29
Computer Science		
Expert	~455,000	52
Learner	~ 112,000	59
Law		
Expert	~738,000	91
Learner	~182,000	77

Representativeness



Работа с корпусами



Results

- Собраны статистические данные по 40 маркерам
- Делятся на 2 группы:
  - ‘Недопредставленные’ (underrepresented)
  - ‘Перепредставленные’ (overrepresented)
- Удовлетворяют критериям:
  - Различие в нормализованной частотности статистически значимо
  - Величина различия ощутима (top-3)
  - Различие характерно для всех дисциплин

Underrepresented features

Feature	mean effect size	coefficient of variation
concessive clauses	-69	-0.2
degree adverbs	-54	-0.3
linking adverbials	-44	-0.3

Overrepresented features

extraposed to cl-s	148	0.4
subj pred to-clause	24	0.3
this as a pronoun	19	0.4

Case: concessive clauses

mean effect size: **-69%**; coefficient of variation: **0.2**

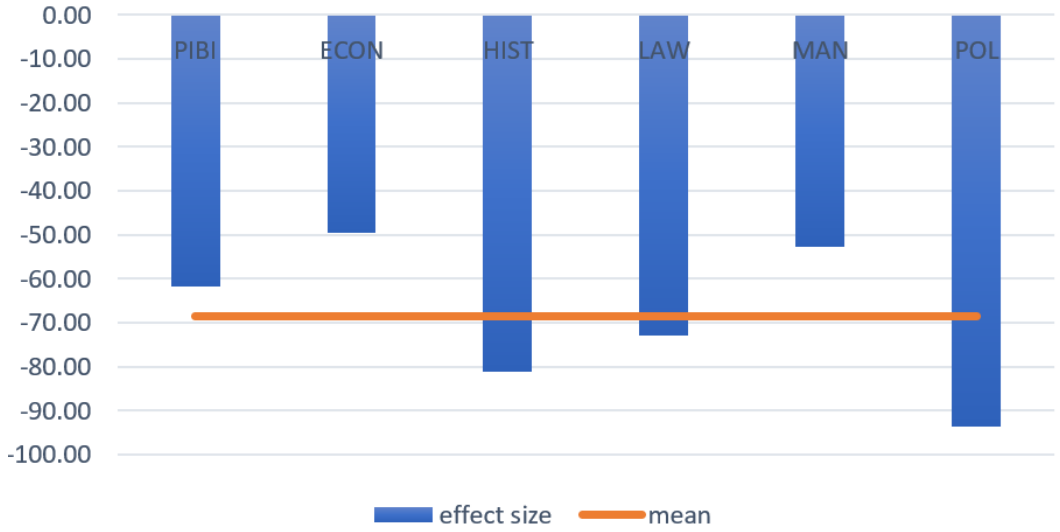
Although *there are many forecasts devoted to the description of the future economic situation, they are often inaccurate and mutually contradictory.* (Learner Corpus, Economics, 2015)

These same errors explain why ship prices are so high in booms, ***even though*** *earnings mean revert so rapidly in the data.* (Expert Corpus, Economics, 2015)

'Concessive clauses are often used to show the limitations of certain facts, events, or claims' (Biber et al. 1999, 825)

Case: concessive clauses

mean effect size: **-69%**; coefficient of variation: **0.2**



Underrepresented features

feature	example	effect size
concessive clauses	***Even though*** *in most testing cases the background is filled mostly by the vehicle color*, the use of complex scenery in the generation process avoids overfitting.	**-69%**
degree adverbs	At this time, the communication cost is equal to or ***slightly*** greater than the calculation cost.	**-54%**
linking adverbials	***Therefore***, the wavelet transform can be used to make reveal the hidden information of fault	**-44%**

Examples from Expert Corpus

Overrepresented features

feature	example	effect size
extraposed *to*-clauses	***It is*** *crucial not only* ***to*** *choose* a right model, but also to select features.	**+148%**
subject predicative *to*-clause	As we already stated above, *the purpose* of this research ***is to*** *develop* a smart office equipment control application	**+24%**
this as a pronoun	**This** means, that to reduce the number of accidents, the telecommunication company incorporate some changes to the network organization.	**+19%**

Examples from Learner Corpus

Further steps

- Проверить и улучшить механизмы выявления маркеров (точность и полнота)
- Выявить наиболее значимые маркеры академического регистра путем сопоставления экспертного корпуса и письменного корпуса общего характера
- Провести качественный анализ по наиболее значимым маркерам, чтобы определить причину различий:
  - learner v expert language user?
  - project proposal v journal article?

Implications

ESL students’ challenges in academic writing:

a) language

b) genre

c) research

Data-driven learning (DDL)

for a student

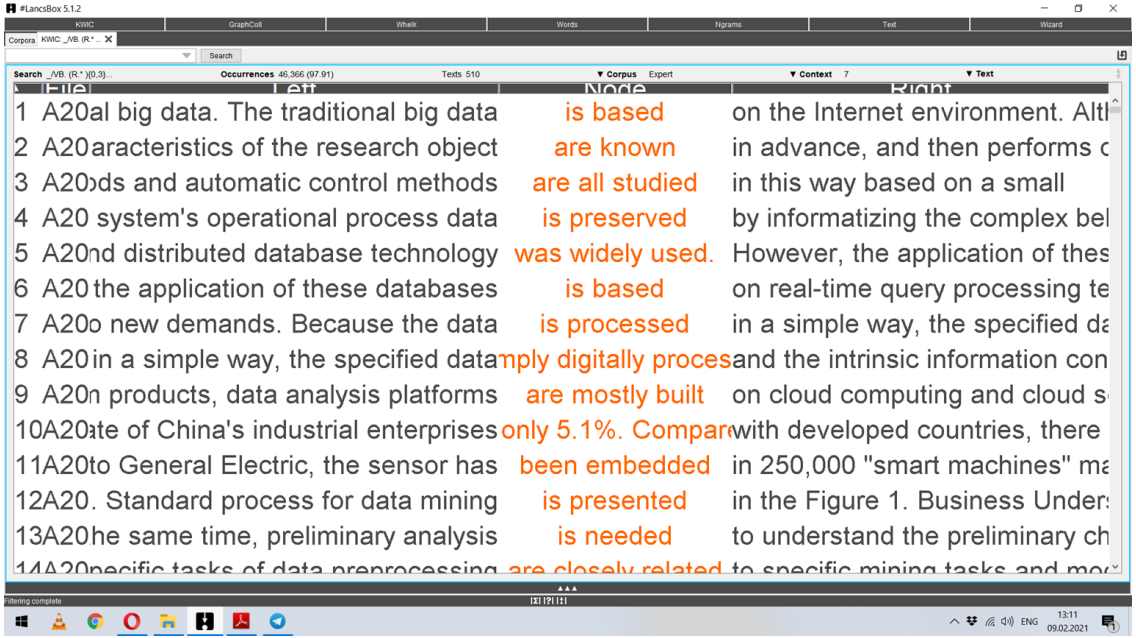
for a teacher

- reference materials
  - priority markers
- mistake prevention
  - materials for exercises
- *automated system of feedback*
  and tests

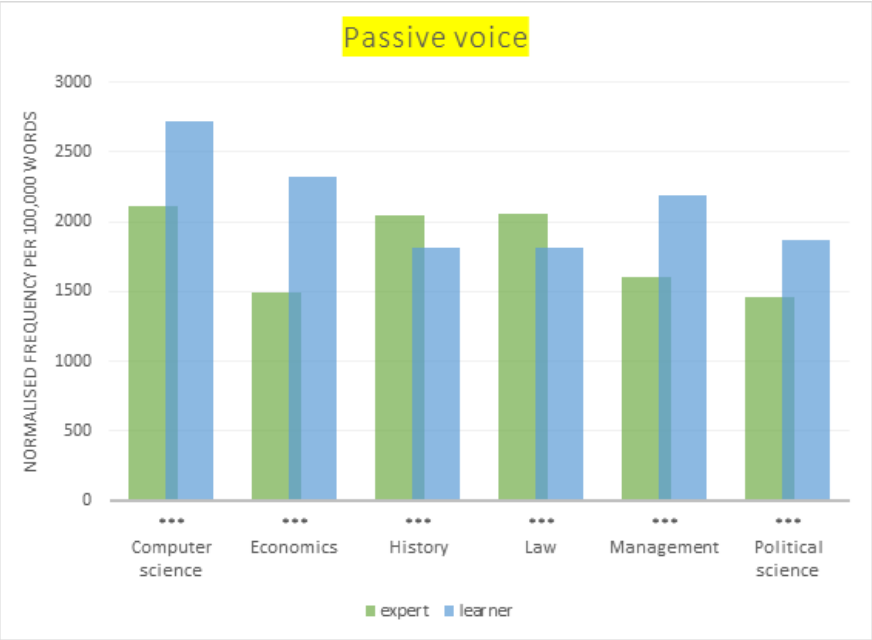
for both

- sample texts
- Ss' statistics

Concordance lines as reference material



Stats for mistake prevention



Exercises (1)

Find examples of the active and passive voice in the extract below:

Likert was interested in correlational relationships among certain internal influences (e.g., leadership), aspects of employees' future service potential (e.g., attitudes), and emergence-enabling systems (e.g., group process and climate), and in how those might then relate to changes in outcomes such as productivity (Likert & Bowers, 1973). He speculated that with enough units where people do very similar work and turnover is low or enough periods of data within one unit, an organization might be able to use these relationships to predict "savings or loss due to changes in productivity capability of the human organization" (Likert & Bowers, 1973: 21), and capitalize this as an asset. He also suggested that future changes in income streams due, for example, to changes in climate could be projected and used to estimate the present value of the human capital resource (Likert, 1967). While he proposed these relationships might exist, his work focused more on measuring predictors like job satisfaction and on identifying correlational relationships, and not on developing a formal human capital valuation model.

Exercises (2)

Correct the following sentences:

1) To handle this limitation my research will be base on data from “IR magazine Russia” …

2) Lorenzo (2011) studied the use of the term operational model and concluded that it has been first used to describe the organizational structure of business divisions.

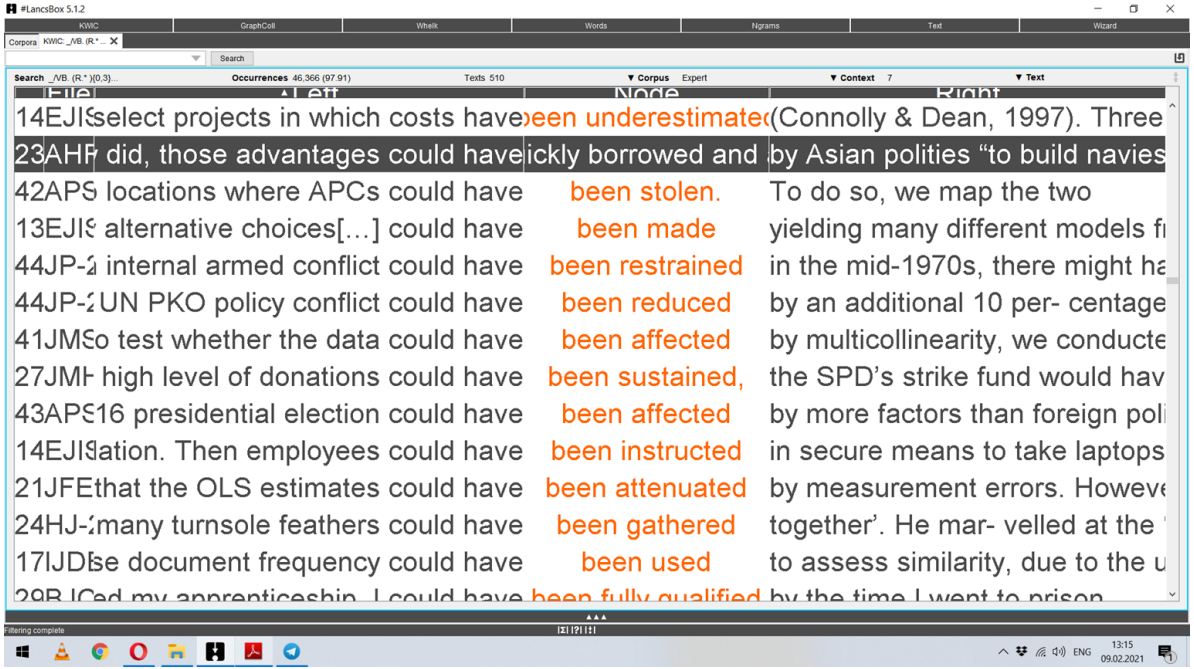
3) Moreover, the literature seems to be wanting of contributions focused on the design of logistics operations for roadshows from event management point of view.

4) The presentation of the all results will use an appropriate pattern in order to implications will be described accurately and objectively.

5) His model is also included 32 strategic tasks grouped according to these phases.

Exercises (3)

Which modal verbs precede Passive Voice and what do they signify?



Practical Implications & Further Research

1. Academic Writing на 4-м курсе
2. Семинар для AWC
3. Conference paper(s)
4. Research paper(s)

References

1. Biber, D., & Gray, B. (2016). *Grammatical complexity in academic English: Linguistic change in writing*. Cambridge University Press.

2. Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E., & Quirk, R. (1999). *Longman grammar of spoken and written English*. London: Longman.

3. Karpenko-Seccombe, T. (2021). *Academic Writing With Corpora: A Resource Book for Data-Driven Learning*. Oxon and New York: Routledge.