НУЛ Учебных корпусов

Cause and effect expressions in hard and soft sciences: a corpus-based analysis

Elizaveta Smirnova
Elena Kostareva

- Introduction

- Literature review

- Data & Method

- Discussion & conclusion

Introduction

Cause and effect, a relation established between two events, where the first event is considered to be a reason for the second one and the second event appears to be the result of the first event (Gopalan & Devi, 2017)

cause-effect relationship can be expressed by

- nouns (cause, reason, consequence, result)

- verbs (to lead to, to result in)

- adverbial clauses of cause and reason marked by conjunctions (because, since, so)

Studies are focused on :

- exampling functional differences between cause and effect expressions (Halliday & Hasan, 1976; van Dijk, 1977; Winter, 1977, 1982; Sanders & Spooren, 2015)

Studies are focused on :

• 	cognitive categorization of cause and effect relations in speech and writing

(Altenberg, 1984; Pander Maat & Sanders, 2001; Sanders & Spooren, 2015).

Introduction

The possible gap in studies:

the way cause and effect expressions are used in different disciplines appears to be under-research

Introduction

- RQ 1 Are cause and effect expressions used in different quantities in research articles in hard and soft sciences? Are there any particular trends in their use?

- RQ 2 Do hard and soft sciences employ cause and effect expressions in different grammatical patterns?

Hypothesis:

 - linguistic expression of cause and effect deviates in hard and soft sciences

 - differences are qualitative and quantitative

 - hard and soft sciences employ the words expressing cause and effect in different grammatical patterns

Literature review

- The focus of earlier studies:

the functional differences of cause-effect markers and their stylistic peculiarities expressions (Halliday & Hasan, 1976; van Dijk, 1977; Winter, 1977, 1982)

Literature review

- The focus of modern studies:

the role of causal-effect expressions in different types of discourses and considering the ways of their automated extraction and analysis(Marshman, 2004; Marshman & L'Homme, 2006; Sanders & Spooren, 2015 ; Chukharev-Hudilainen & Saricaoglu, 2016; Gopalan & Devi, 2017; Cao et al., 2018)

Data and method

 - Quantitative and qualitative approaches comprise frequency counts, statistical tests and text analysis of a corpus of published papers

 - corpora of randomly retrieved research articles from peer-reviewed journals in four hard (chemistry, biology, physics, mathematics)and four disciplines (business studies, history, linguistics, political sciences)

Data and method

Discipline	Number of texts	Number of words
Chemistry	16	97947
Physics	18	95852
Mathematics	13	98430
Engineering	17	99003
Hard sciences total	64	391232
Business Studies	10	95350
Linguistics	10	95603
History	10	99303
Political Science	11	93366
Soft sciences total	41	383622

Data and method

The precision rate - 93.5%

The recall rate – 98%

Log-likelihood tests to check the significance of the differences in frequencies in the two groups of sciences (Dunning, 1993)

Results

RQ1 Linguistic units expressing cause are used in similar quantities in hard and soft sciences

Linguistic units expressing effect are more frequent in hard sciences except for nouns

RQ2 The use of different parts of speech demonstrates different patterns

Results

Four patterns that occur in all the disciplines:
- *for some reason + clause*

   - *verb+ some effect of sth on sth,*

 - contribute + to + noun and

 - lead + to + noun

All the studied patterns with *reason* and *effect* are more common in soft sciences

The pattern *origin of sth + be* was not found in the corpus of soft sciences

The pattern *have + some consequence for sth* was not found in the corpus of hard sciences

Examples of deviations in particular sciences:

as a consequence *of sth* was not found in the papers in Engineering and Linguistics;

lead sb to do sth does not occur in Physics and is very rare in Engineering, Mathematics and Linguistics;

to contribute to + infinitive is quite rare in the corpora with one instance in the papers in Chemistry and three instances in History.

Conclusions
possible limitations

- The size of our corpora is not large;

- Studies in other disciplines and genres might yield a fuller picture of disciplinary variation in the use of cause and effect expressions.

Conclusions
possible applications

-to produce the discipline-specific language learning materials that effectively address the needs of learners of different sciences

- to ensure a more principled approach to ESP/EAP course design

Thank you for attention!