



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

CORPUS-BASED ANALYSIS OF DISCIPLINARY VARIATION IN PROFESSIONAL ACADEMIC WRITING: IMPLICATIONS FOR EAP/ESP PEDAGOGY

Elizaveta A. Smirnova,
HSE University - Perm
easmirnova@hse.ru



Plan

- complexity features
- cause-and-effect strategies
- stance expressions





Corpus

Discipline	No. texts	No. tokens
Hard sciences		
Chemistry	34	197,806
Physics	44	200,206
Mathematics	28	199,380
Engineering	34	198,926
Totals	140	796,318
Soft sciences		
Business	20	197,956
Linguistics	22	200,997
History	21	199,394
Political science	25	202,040
Totals	88	800,387

Methods

- AntConc 3.4.4 (Anthony, 2014)
- Manual pruning of the data
- Statistical significance tests





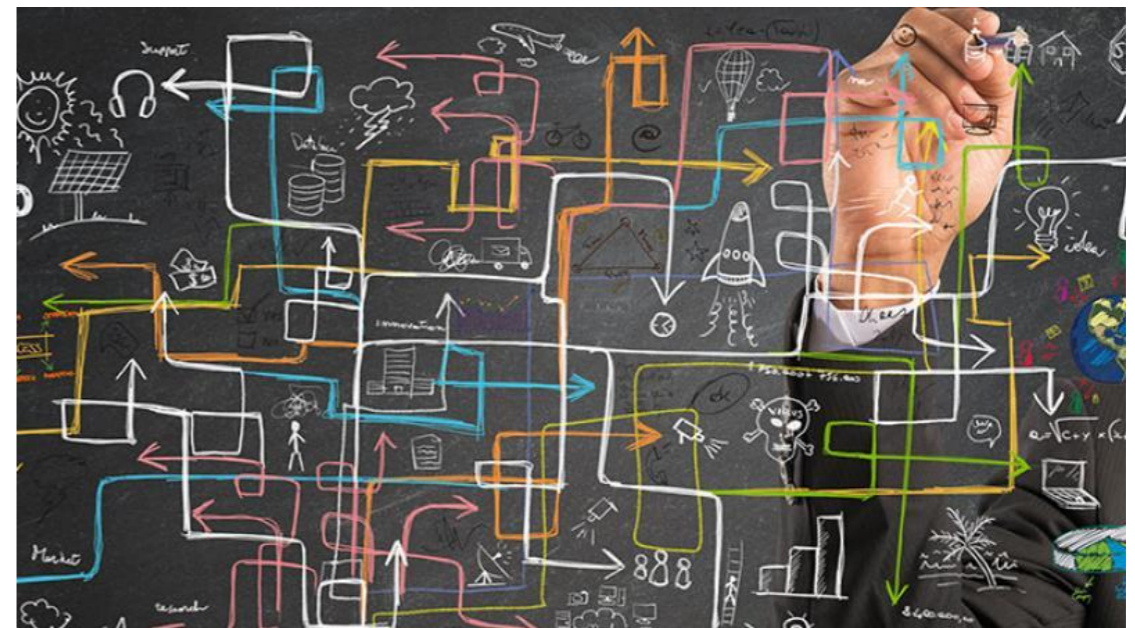
Complexity features: Linguistic items

Clausal complexity features:

- finite adverbial clauses of purpose, condition, concession, time, place, reason, result, manner, and contrast;
- *wh*-complement clauses;
- verb + *that*-clauses.

Phrasal complexity features:

- nouns;
- attributive adjectives;
- premodifying nouns;
- *of*-genetives.





Complexity features: Results

- Most of the differences were found to be **significant**.

- Clausal complexity:
 - overall, **adverbial clauses** were found to be more frequent in the corpus of the **hard-science** papers;
 - **wh-clauses** and **that-clauses** prevail in the **soft** research papers.

- Phrasal complexity:
 - the preference for **adjectival** and **prepositional** phrases in the **soft-science** texts;
 - the preference for **nominal** categories in the **hard** sciences.



Complexity features: Implications (1)

The high frequency of **condition** (1), **place** (2), and **reason** (3) **adverbial clauses** in **mathematics** where they are used in the comments for calculations and formulas:

(1) *If we apply the recurrences $\sigma_1 \sigma_3 \dots \sigma_{n-1}$ to an arbitrary ca_0, a_1, \dots, a_n , we obtain a linear combination of lower coefficients (CM-2016-2).*

(2) *For example, let $G = SL_4(\mathbb{C})$, and let $w = w_0 s_3$, where w_0 is the longest element in $W = S_4$, the symmetric group with four letters (CM-2016-1).*

(3) *Since the off-diagonal factors of R are all even in x_0, x_2, \dots, x_n and $P(x)$ is even, the diagonal factors of R must be even as well (CM-2016-2).*



Complexity features: Implications (2)

The high frequency of **time** (4) and **reason** (5) **adverbial clauses** in the **political-science** research papers, specifically in comments for mathematical and statistical models :

(4) Conversely, multiple imputation will offer small gains in bias reduction *when variables of theoretical interest have a low proportion of missing values* (PA-2016-1).

(5) Such bias inducers may not be troublesome in practice, however, either *because they can be identified for exclusion, as is sometimes the case for post treatment variables, or because the bias they induce tends to be small* (PA-2016-3).



Complexity features: Implications (3)

Physics was found to employ a large number of **attributive adjectives** which, as in **chemistry**, are often part of terms used in the discipline, e.g. *interatomic* bonds, *local* minimum, *magnetic* field, *massive* gravity, *black* hole.

Chemistry also tends to extensively use **complex nominals** (6) due to the pervasiveness of outstandingly long names of chemical entities and processes in the field.

(6) We have demonstrated a new pathway of *unsaturated fatty acid synthesis* that is catalyzed by an enzyme, FabX, that has dual dehydrogenase/isomerase activities (CCB-2016-1)



Cause-and-effect: Linguistic items

	Nouns	Conjunctions	Verbs	Prepositions
CAUSE	<i>cause</i>	<i>because</i>	<i>contribute</i>	<i>because of</i>
	<i>factor</i>	<i>given that</i>		<i>due to</i>
	<i>ground</i>	<i>since</i>		<i>following</i>
	<i>origin</i>			<i>given</i>
	<i>reason</i>			<i>in view of</i>
	<i>root</i>			<i>on account of</i>
	<i>source</i>			<i>owing to</i>
	Nouns	Conjunctions	Verbs	Adverbs
EFFECT	<i>consequence</i>	<i>so</i>	<i>result</i>	<i>accordingly</i>
	<i>effect</i>	<i>so that</i>	<i>lead</i>	<i>consequently</i>
	<i>implication</i>			<i>hence</i>
	<i>outcome</i>			<i>naturally</i>
	<i>result</i>			<i>thereby</i>
				<i>therefore</i>
			<i>thus</i>	



Cause-and-effect: Results

- Cause:
 - **prepositions** are characteristic of **hard** sciences, whereas the key **nouns** and the **verb** (*contribute*) are especially frequent in **soft-science** papers.

- Effect:
 - In **soft** sciences, effect tends to be expressed via **nouns**, whereas **conjunctions** and **adverbs** are characteristic of **hard** sciences. The key verbs have not proved to be clear indicators of the hard/soft distinction.



Cause-and-effect: Implications (1)

The objective and impersonal style of the hard-science discourse - the frequent use of linguistic strategies that provide arguments and justify the statements. Expressions introduced by linking items such as **conjunctions** (mostly in adverbial clauses) (7) and **prepositions** (commonly introducing modifiers of nominal categories) are optimal ways of contributing towards the expression of causality and effect in these **hard-science** texts.

(7) In fact df has maximal rank over N_{sm} so *that* f is a totally geodesic immersion with the induced flat metric there and in particular $N_0 = N_{00}$ (JDG 2016-3).



Cause-and-effect: Implications (2)

The epistemic character of the soft-science discourse reveals a more speaker-centric assertive style, where linking functional items introducing ‘probes’ such as adverbials or modifiers are not pervasive, which justifies the frequency of grammatical categories such as nouns. Also, **soft** sciences might be more responsive to the use of **nouns** (8) expressing cause/effect because soft-knowledge domains implicate less control of variations and thus more possibilities for varied outcomes.

(8) The key pedagogical *implication* of this study is that a simple act of copying novel words, while processing meaningful L2 input, may significantly boost quality of lexical knowledge (AL 2016-4)

Stance: Linguistic items

- modal and semi-modal verbs;
- stance adverbs;
- stance complement clauses controlled by verbs, adjectives, and nouns;
- hedges;
- boosters;
- attitude-markers;
- self-mention expressions.





Stance: Results

Overall, **stance** is explicitly expressed in the **soft-science writings** to a greater extent than in the hard disciplines, with the exception of **a few features** that are more common in the **hard sciences**, namely:

- complement *that*-clauses controlled by epistemic verbs;
- *to*-clauses controlled by adjectives of difficulty;
- expressions of self-mention.



Stance: Implications (1)

Epistemic **adverbs** expressing likelihood (8) are more prevalent in soft-sciences, for example, in **history** papers, where the adverbs are used to express authorial speculations on some historic events.

(8) Wealthy parents may have chosen to inoculate their children, and this *probably* accounted for much of the mortality advantage associated with wealth in the period 1752-94. However this advantage was *probably* very substantially diminished with the very widespread acceptance of and access to vaccination after 1800. (HF-2019-1).



Stance: Implications (2)

Complement clauses controlled by verbs are more strongly associated with the linguistic repertoire of hard sciences. For instance, ***that*-clauses governed by verbs** expressing certainty (9) are significantly more frequent in the hard-science texts, specifically in **mathematics**, where they are used to comment on calculations.

(9) *We shall prove that* $x(0)$ satisfies the condition (3.21). It is *easy to prove that* conditions (1.5) can be formulated in the following equivalent form where... (AMC-2019-3)



Stance: Implications (3)

Expressions of self-mention were also found to occur more frequently in hard-science papers. For instance, **mathematics** writings were found to employ personal pronouns *I* and *we* more frequently than any other science, which contradicts Burton and Morgan's (2000: 435) claim that in this discipline "the apparent absence of the author from the text fits with positivist epistemologies in which the mathematician's role is subordinate to that of the mathematics itself".

Conclusions

- The analysis of disciplinary variation in the use of the discourse markers can aid in the development of genre-specific language resources for ESP and EAP courses.
- Understanding how different disciplines employ complexity features, express cause/effect and stance can help learners to write more effectively in their respective fields.





НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

easmirnova@hse.ru