

Goals

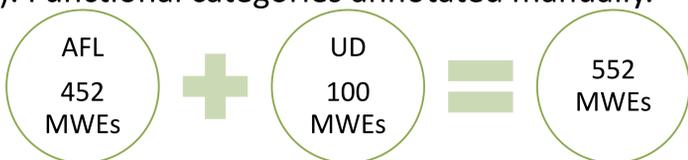
- ▶ What are the functions MWEs have fulfilled in scientific English writing across 300 years?
- ▶ How to characterize the diachronic development of their usage over time?

Data

- ▶ Royal Society Corpus (RSC) 6.0 (Fischer et al. 2020)
 - ▶ Diachronic corpus of scientific English.
 - ▶ From 1665 until 1996.
 - ▶ 47 837 texts (295 895 749 tokens).

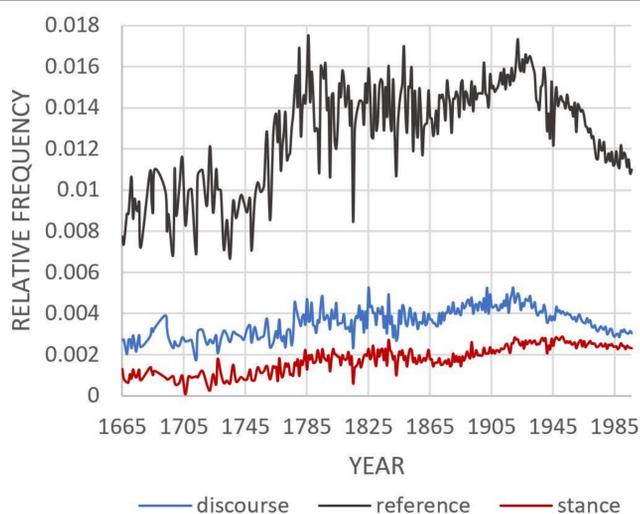
Multi-word Expressions

- ▶ Two complementary corpus-based approaches:
 1. **Academic Formulas List (AFL)**: most common formulaic sequences in academic English: List of 603 MWEs extracted using the formula teaching worth (FTW) and classified according to Biber et al. (2004).
 2. **Universal Dependencies (UD) method**: Extraction of the 100 most frequent MWEs (*fixed* dependency relation) from the RSC parsed with Stanza (combined English model). Functional categories annotated manually.



Function	Type	MWEs	Examples
Stance	epistemic	84	<i>it is important, according to</i>
	attitudinal/modality	24	<i>we have to, needs to be</i>
	intention/prediction	11	<i>if you want to, to do so</i>
	ability	34	<i>can be found, it is possible to</i>
Discourse	topic introduction/focus	31	<i>in this article, for example in</i>
	topic elaboration/clarification	70	<i>due to the fact, the reason for</i>
Reference	identification/focus	61	<i>such as the, as can be seen in</i>
	imprecision	3	<i>and so on, and so forth</i>
	specification of attributes	177	<i>a form of, on the basis of</i>
	time/place/text reference	57	<i>at the end of, in between</i>

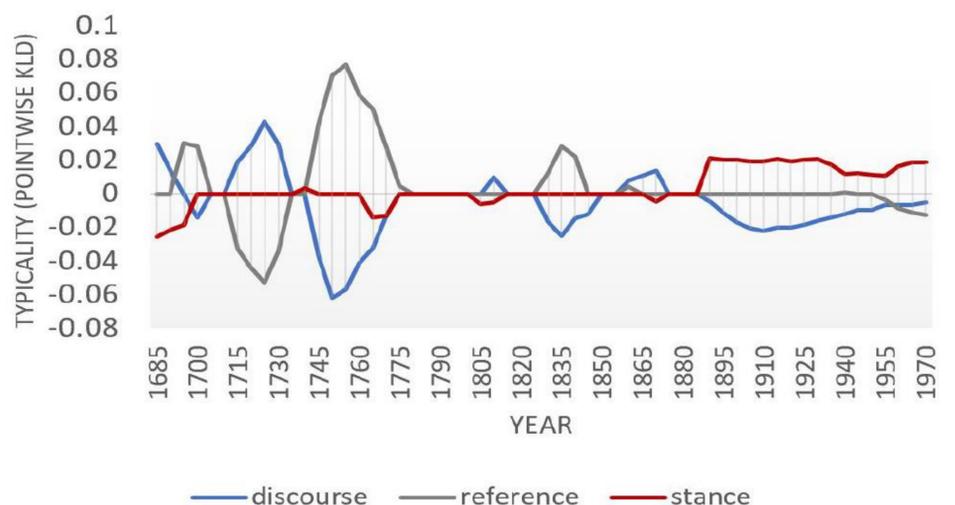
Frequency-based Trends



- ▶ Upward trend until the early 20th century.
- ▶ Referential expressions notably increase in the latter half of the 18th century.
- ▶ Stance expressions demonstrate a consistent rise from 1925 onwards.

Diachronic Changes by Divergence - Functions

- ▶ Kullback-Leibler Divergence (KLD): $D_{KL}(A||B) = \sum_{x \in X} A(x) \log \left(\frac{A(x)}{B(x)} \right)$



- ▶ From the 17th to the beginning of the 20th century, reference and discourse MWEs tend to behave in opposite directions.
- ▶ In the 20th century, stance expressions become more typical.

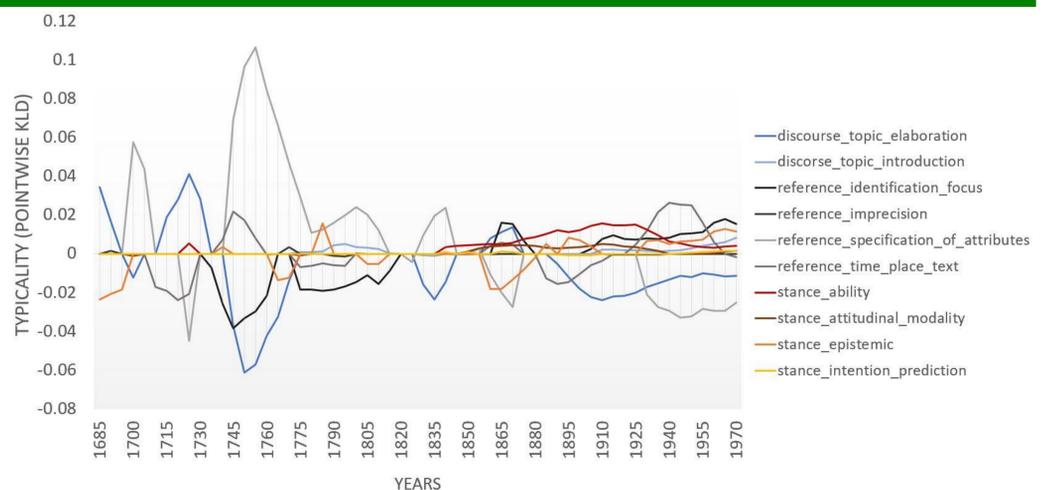
Diachronic Changes by Divergence – Function Types

Trends in Discourse and Referential Expressions:

- ▶ Topic Elaboration/Clarification: Historical and cultural contexts emphasizing explicit reasoning favor elaborated discourse.
- ▶ Specification of Attributes: Evolving academic standards demand precise attribute specification.

Divergence in Stance Expressions:

- ▶ By 1825, ability expressions become prominent, followed by attitudinal expressions, declining in the 1930s. Epistemic expressions then gain prevalence.
- ▶ MWEs increasingly articulate evidence-based reasoning.



Conclusion and Future Work

- ▶ Dynamic landscape of MWE usage, marked by significant shifts in function that reflect changing priorities and practices within the scientific community over time.
- ▶ Future work: (1) increase the number of MWEs; (2) model MWEs at the paradigmatic level; and (3) apply probabilistic measures of processing (e.g. surprisal).