BATS:
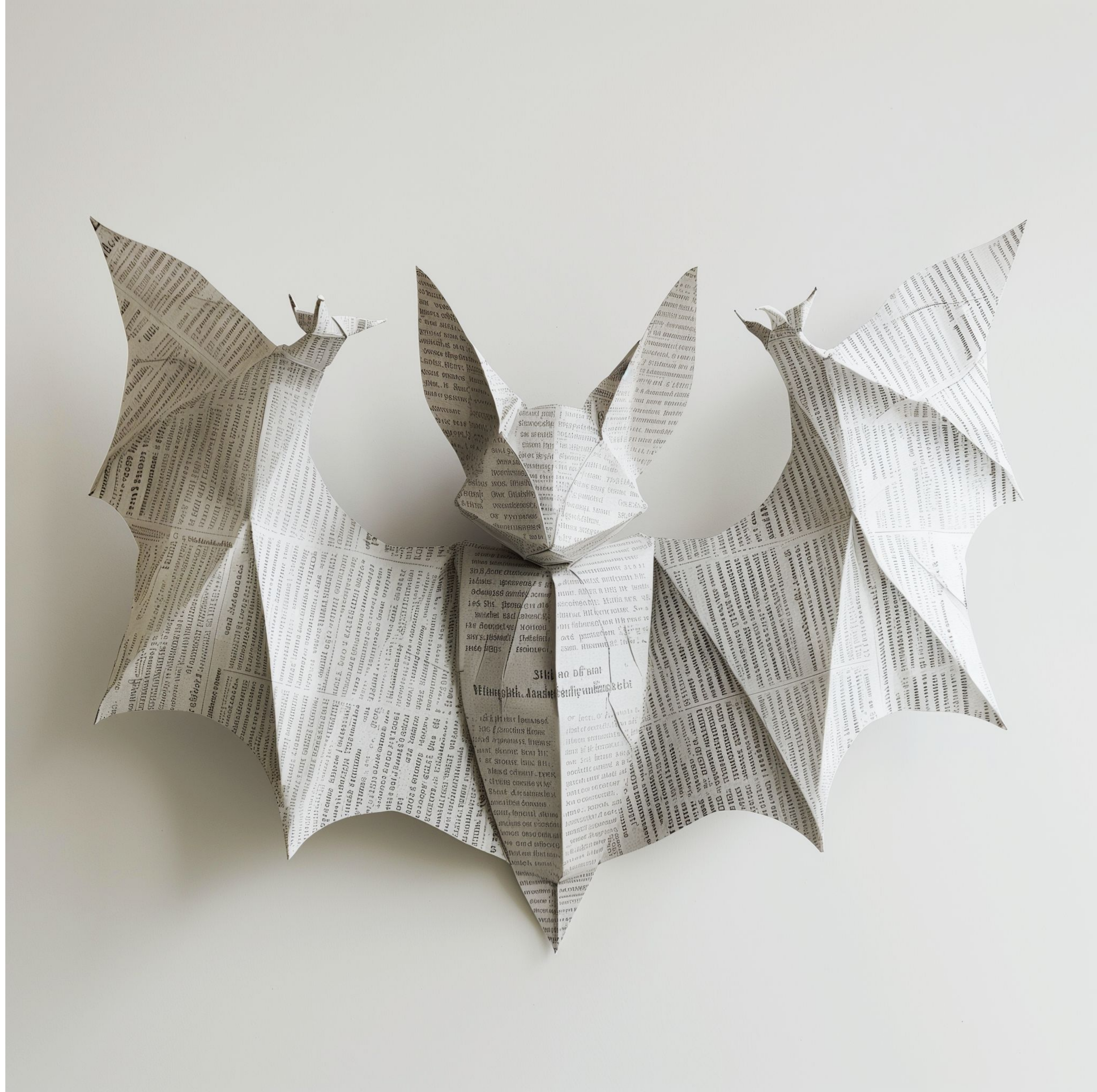# BenchmArking
Text Simplicity

Christin Katharina Kreutz,
Fabian Haak, Björn Engelmann,
Philipp Schaer

THM
TECHNISCHE HOCHSCHULE MITTELHESSEN

HERDER-INSTITUT
für historische Ostmitteleuropaforschung
INSTITUT DER LEIBNIZ-GEMEINSCHAFT

CIR

Technology
Arts Sciences
TH Köln

Objective: Evaluate Text Simplification

→ *how easy it is for a certain group of readers to understand a given text?*

Approach: Evaluate simplicity by **interpretable qualitative aspects** and data programming

HERDER-INSTITUT
für historische Ostmitteleuropaforschung
INSTITUT DER LEIBNIZ-GEMEINSCHAFT

THM
TECHNISCHE HOCHSCHULE MITTELHESSEN

CIR

Technology
Arts Sciences
TH Köln

Current Evaluation Challenges

- **Limitations of current evaluation approaches**
  - **too simple:** Flesch-Kincaid and similar metrics evaluate basic readability
  - **mostly reference-based:** SARI, BLEU etc. assess similarity to an optimal simplification reference

- **Lack of Explainability**

- **Not taking domain & target audience into account**

HERDER-INSTITUT
für historische Ostmitteleuropaforschung
INSTITUT DER LEIBNIZ-GEMEINSCHAFT

THM
TECHNISCHE HOCHSCHULE MITTELHESSEN

CIR

Technology
Arts
Sciences
TH Köln

1 Which characteristics from literature reflect the simplicity or complexity of texts?
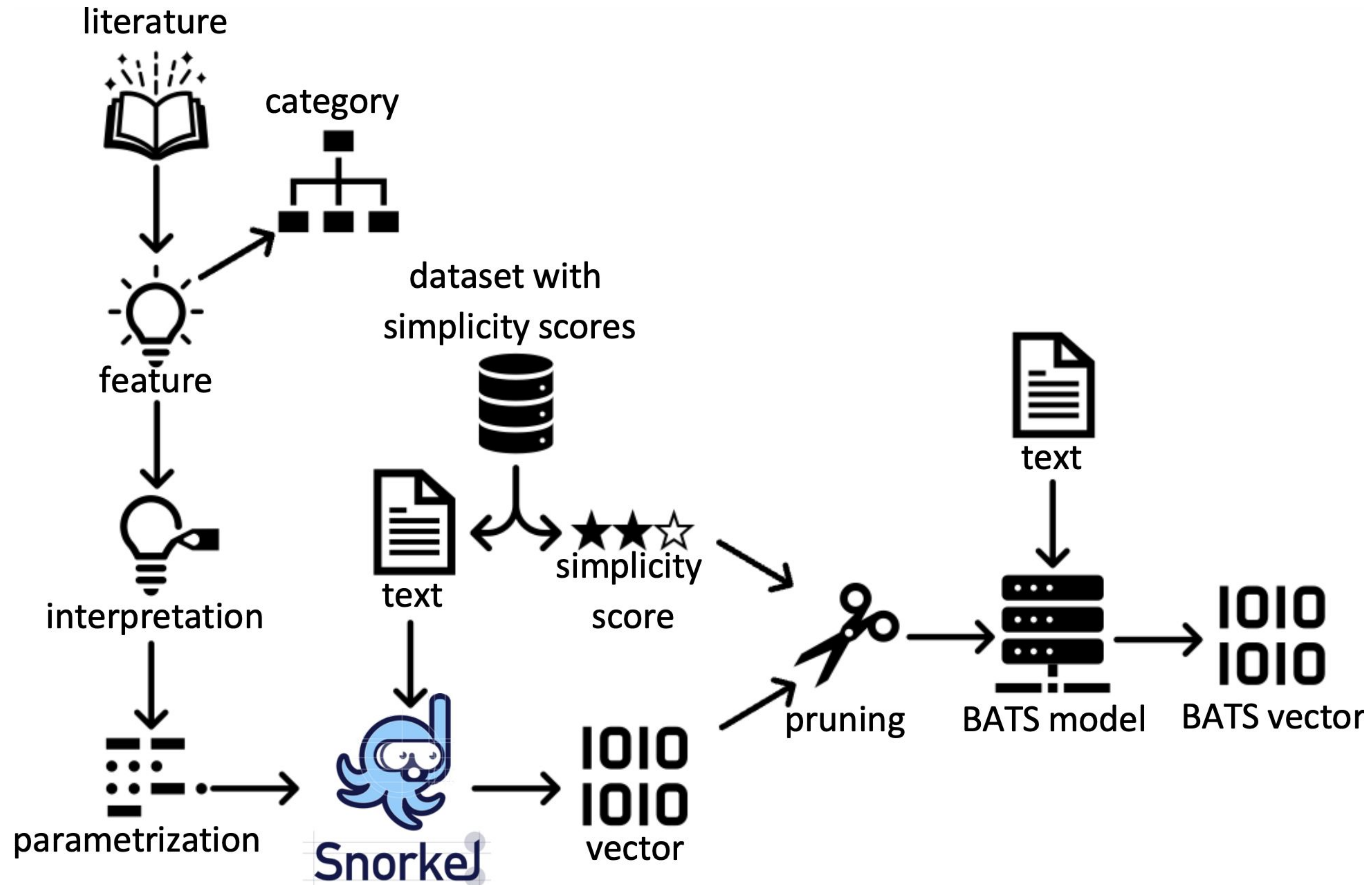
2 Which dataset-specific, target audience specific, and domain-specific characteristics can be found regarding simplicity?

3 Can BATS be used to quantify simplicity effectively?

BATS:
BenchmArking Text Simplicity
ACL 2024

HERDER-INSTITUT
für historische Ostmitteleuropaforschung
INSTITUT DER LEIBNIZ-GEMEINSCHAFT

THM
TECHNISCHE HOCHSCHULE MITTELHESSEN

CIR

Technology
Arts Sciences
TH Köln

Data: English Parrallel Corpora

Dataset	# texts	Target Audiences (TA)	Domains (D)
ASSET	4718		
AutoMeTS	6994		medical
BenchLS	1856		
Britannica	926	children	encyclopedia
EW-SEW-Turk	1000		encyclopedia
HutSSF	652		news
METAeval	604		encyclopedia
MTurkSF	126	non-experts	medical
NNSeval	478	language learners	encyclopedia
OneStopEnglish	4144	language learners	news
QuestEval	282		encyclopedia
SemEval_2007	598		
SimPA	2204	language learners	administrative
SimpEval	324		encyclopedia
TurkCorpus	4718	children, language learners	encyclopedia

literature

category

feature

interpretation

parametrization

dataset with
simplicity scores

text

★★☆
simplicity
score

IOIO
IOIO
vector

text

pruning

BATS model

IOIO
IOIO
BATS vector

Snorkel

BATS:
BenchmArking Text Simplicity
ACL 2024

HERDER-INSTITUT
für historische Ostmitteleuropaforschung
INSTITUT DER LEIBNIZ-GEMEINSCHAFT

THM
TECHNISCHE HOCHSCHULE MITTELHESSEN

CIR

Technology
Arts Sciences
TH Köln

From Features to Snorkel Labels

37
Features

- few unique entities
- few long words
- few words per sentence
- few negations
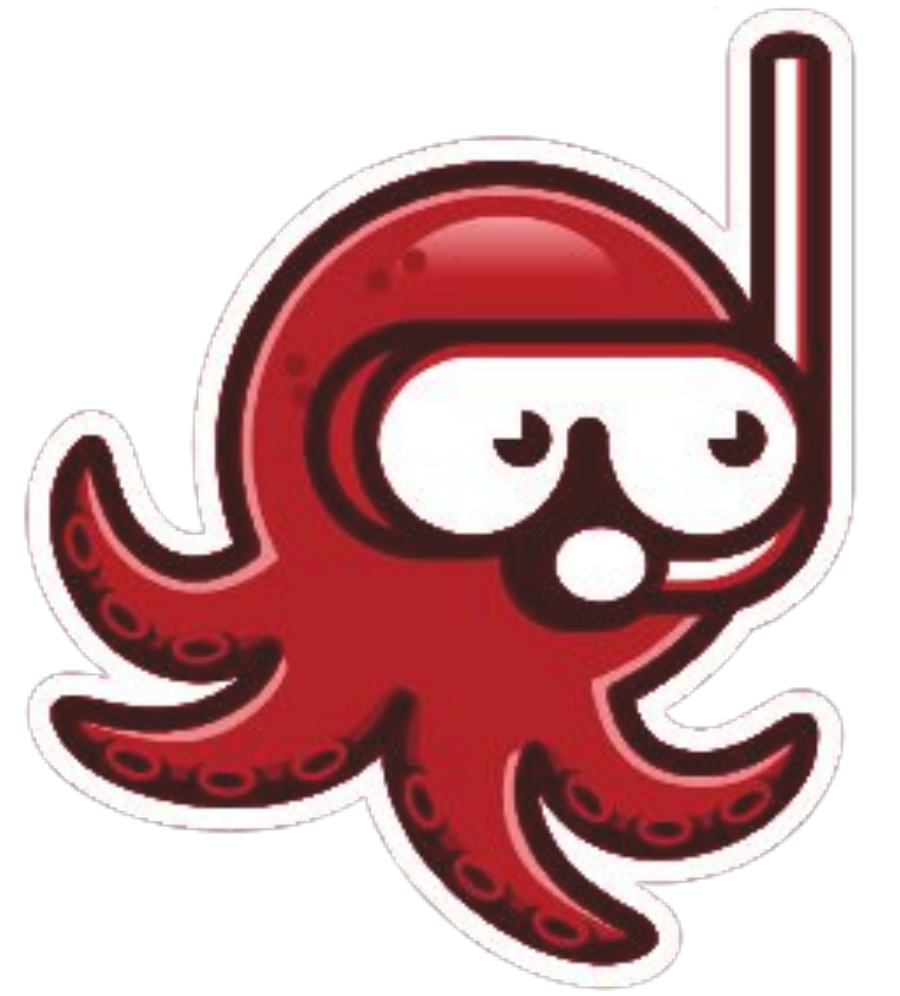- low depth of the syntactic tree
- ...

135
Interpretations

- few unique entities
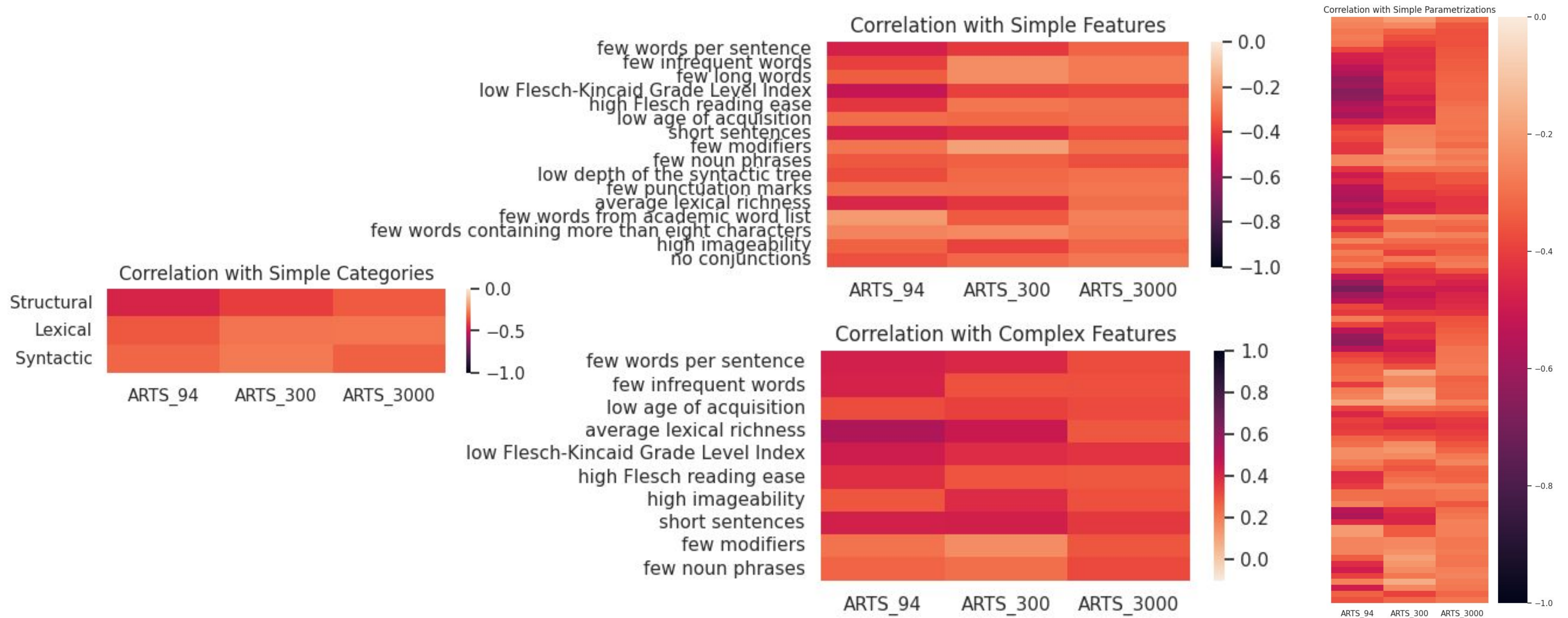 - entities/sentence
  - entities/text
  - entities/tokens

1249
Parametrizations

- few unique entities
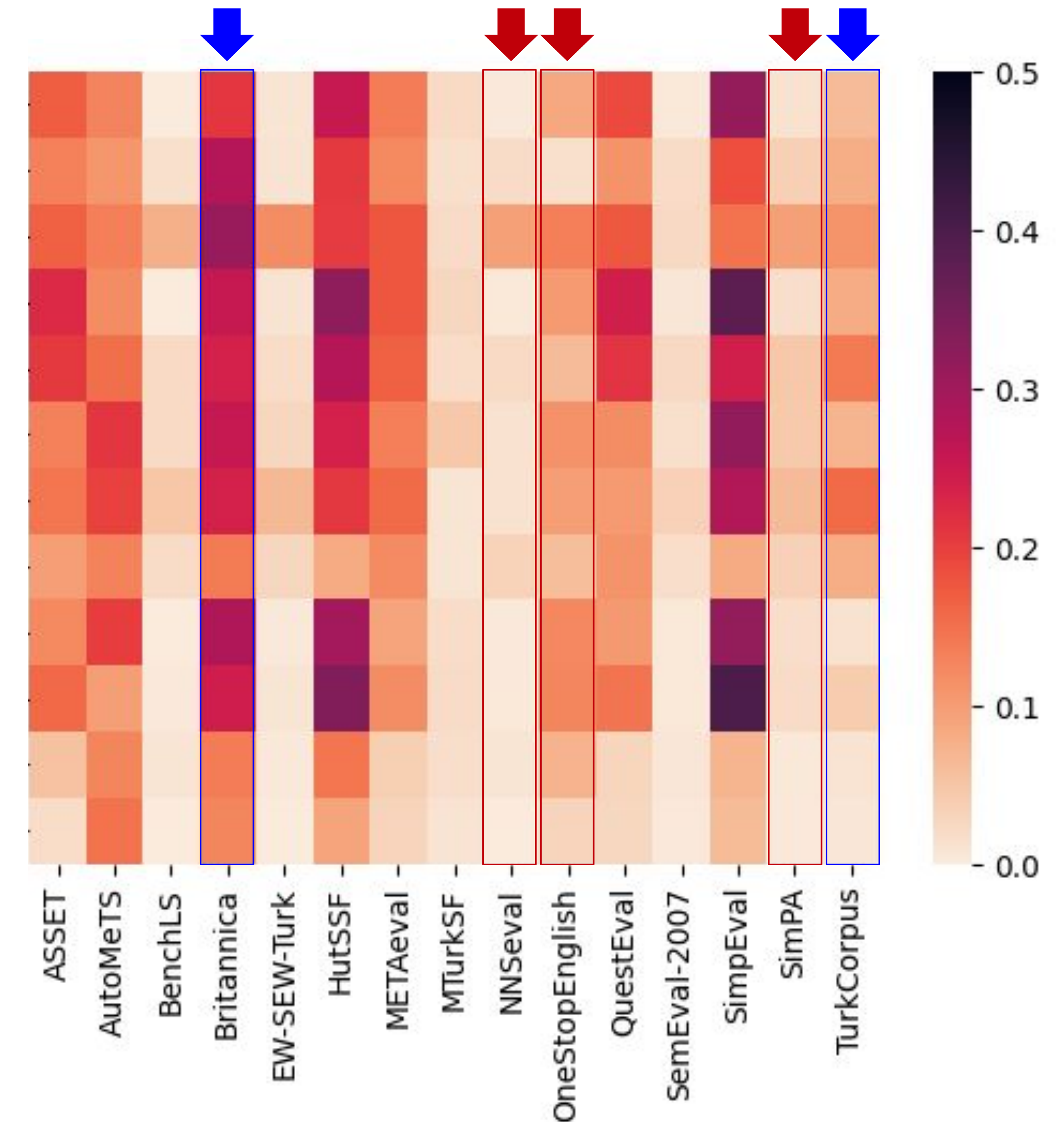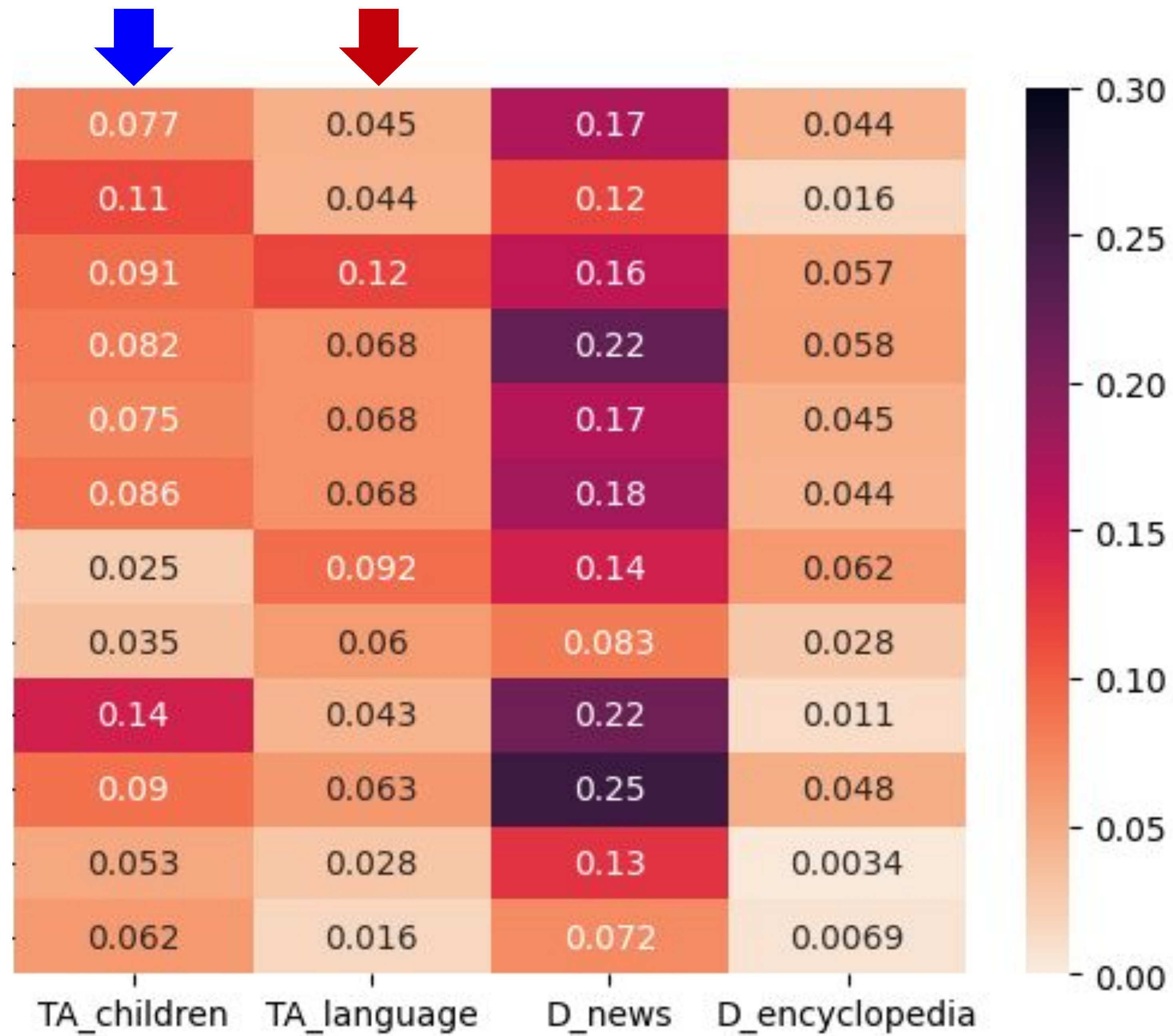 - entities/sentence
    - < 5
    - < 4
    - < 3
    - < 2
    - < 1

Which characteristics from literature reflect the simplicity or complexity of texts?

1

BATS:
BenchmArking Text Simplicity
ACL 2024

Which dataset-specific, target audience specific, and domain-specific characteristics can be found regarding simplicity?

2

BATS:
BenchmArking Text Simplicity
ACL 2024

HERDER-INSTITUT
für historische Ostmitteleuropaforschung
INSTITUT DER LEIBNIZ-GEMEINSCHAFT

THM
TECHNISCHE HOCHSCHULE MITTELHESSEN

CIR

Technology
Arts Sciences
TH Köln

		MSE (lower values are better)					R^2 (higher values are better)				
		RF		GB			RF		GB		
Train	Predict	BATS	OAI	BATS	OAI	FRE	BATS	OAI	BATS	OAI	FRE
ARTS$_{94}$	ARTS$_{300}$	**.068**	.073	.08	.078	.078	**.188**	.133	.045	.074	.075
ARTS$_{94}$	ARTS$_{3000}$	**.059**	.076	.067	.081	.086	**.296**	.089	.192	.029	-.026
ARTS$_{300}$	ARTS$_{94}$	**.052**	.068	.059	.061	.065	**.393**	.203	.313	.282	.236
ARTS$_{300}$	ARTS$_{3000}$	**.055**	.07	.06	.07	.086	**.336**	.164	.286	.155	-.026
ARTS$_{3000}$	ARTS$_{94}$	.044	.06	**.039**	.055	.065	.485	.297	**.541**	.354	.236
ARTS$_{3000}$	ARTS$_{300}$	.048	.057	**.047**	.052	.078	.426	.325	**.435**	.378	.075

BATS:
BenchmArking Text Simplicity
ACL 2024

HERDER-INSTITUT
für historische Ostmitteleuropaforschung
INSTITUT DER LEIBNIZ-GEMEINSCHAFT

THM
TECHNISCHE HOCHSCHULE MITTELHESSEN

CIR

Technology
Arts Sciences
TH Köln

Conclusion

- The BATS model estimates a simplicity score and can be **explained by characteristics** obtained from the literature

- BATS can be **adapted** for specific **domains** and **target groups** by analyzing certain corpora

- BATS embeddings outperform state-of-the-art embeddings

BATS:
BenchmArking Text Simplicity
ACL 2024

HERDER-INSTITUT
für historische Ostmitteleuropaforschung
INSTITUT DER LEIBNIZ-GEMEINSCHAFT

THM
TECHNISCHE HOCHSCHULE MITTELHESSEN

CIR

Technology
Arts Sciences
TH Köln