Maksims Konca
1868791
Bachelor Informatik
WS 2020
konca@stud.uni-frankfurt.de

Bachelor Thesis

Educational Texttechnology: Quantifying Task Descriptions in a Multilingual Setting

Maksims Konca

10.12.2020

Department of Computer Science and Mathematics
Supervisor: Prof. Dr. Alexander Mehler

Erklärung

Hiermit bestätige ich, dass ich die vorliegende Arbeit selbstständig verfasst habe und keine anderen Quellen oder Hilfsmittel als die in dieser Arbeit angegebenen verwendet habe.

Ort, Datum

Unterschrift

Abstract

This thesis explores a variety of methods of text quantification applicable in the field of educational text technology. Besides the cohort of existing linguistic, lexical, syntactic, and semantic text quantification methods, additional methods based on *Bidirectional Encoder Representations from Transformers* (BERT) are introduced and analysed. The model, developed in this thesis, is tested on a multilingual data composed of task descriptions used in *Test of Understanding in College Economics* (TUCE). Quantitative features extracted from raw textual data are analysed using an array of evaluation methods with the goal of finding the best predictors of the target variable - the rate of correct student responses in TUCE.

Contents

List of Figures

List of Tables

List of Abbreviations

BERT Bidirectional Encoder Representations from Transformers 1, 3

EBFEf Error Based Feature Elimination using forests 20

EMO Evolutionary Multiobjective Optimisation 1, 10, 11, 14

MCS Monte-Carlo Simulation 10

MDS Maximum Distance Correlation Search 10, 18

NLP Natural Language Processing 1

POS Part of Speech 3

RFC Random Forest Classifier 10

TUCE Test of Understanding in College Economics 1, 3

1. Introduction

The introduction of BERT (Devlin et al., 2018) made a great impact on the field of the *Natural Language Processing* (NLP). It improved the results on a variety of NLP tasks considerably[1]. BERT, as a Transformer (Vaswani et al., 2017) based deep neural network consisting of multiple layers and millions of parameters, allows to capture long(er) range intratextual dependencies (e.g. dependencies between parts of the text such as tokens, sentences, or paragraphs) by applying two kinds of attention mechanisms - the encoder-decoder attention (Bahdanau, Cho and Bengio, 2014) and a multi-head self-attention (Cheng, Dong and Lapata, 2016).

The main idea of this thesis is to explore the BERTs potential on the task of quantifying such dependencies, and test, whether it can provide an enhancement to existing approaches by extracting additional information, hidden from classical methods of quantitative linguistics. In particular, BERT is used to develop a set of features, that could be used in the analysis of multilingual learning data, as an extension to the methods, developed by Mehler, Zlatkin-Troitschanskaia et al., 2018; we test, whether these new features improve correlation between the quantitative features of task descriptions used in TUCE (Walstad, Watts and Rebeck, 2007) and perceived difficulty of these tasks manifested in the rate of correct responses.

The second objective is to identify the most salient features responsible for such correlation. To this end we test different feature importance ranking approaches.

The methodology of this thesis is largely based on Mehler, Zlatkin-Troitschanskaia et al., 2018 (Sections 3.1, 3.2, 3.3, and partly 3.4). This allows a direct comparison between the results of both studies. In addition we explore and apply the methods of the *Evolutionary Multiobjective Optimisation* (EMO) (Section 3.5).

The thesis is organised as follows: Chapter 2 describes the data analysed in this study, the preprocessing steps required prior to the analysis, the text characteristics used to quantify the input texts, and the classification and evaluation methods, that were applied. Chapter 3 presents the results, which are discussed in Chapter 4. Finally, Chapter 5 gives a conclusion and an outlook on future work.

[1] see `https://gluebenchmark.com/leaderboard`

2. Methods

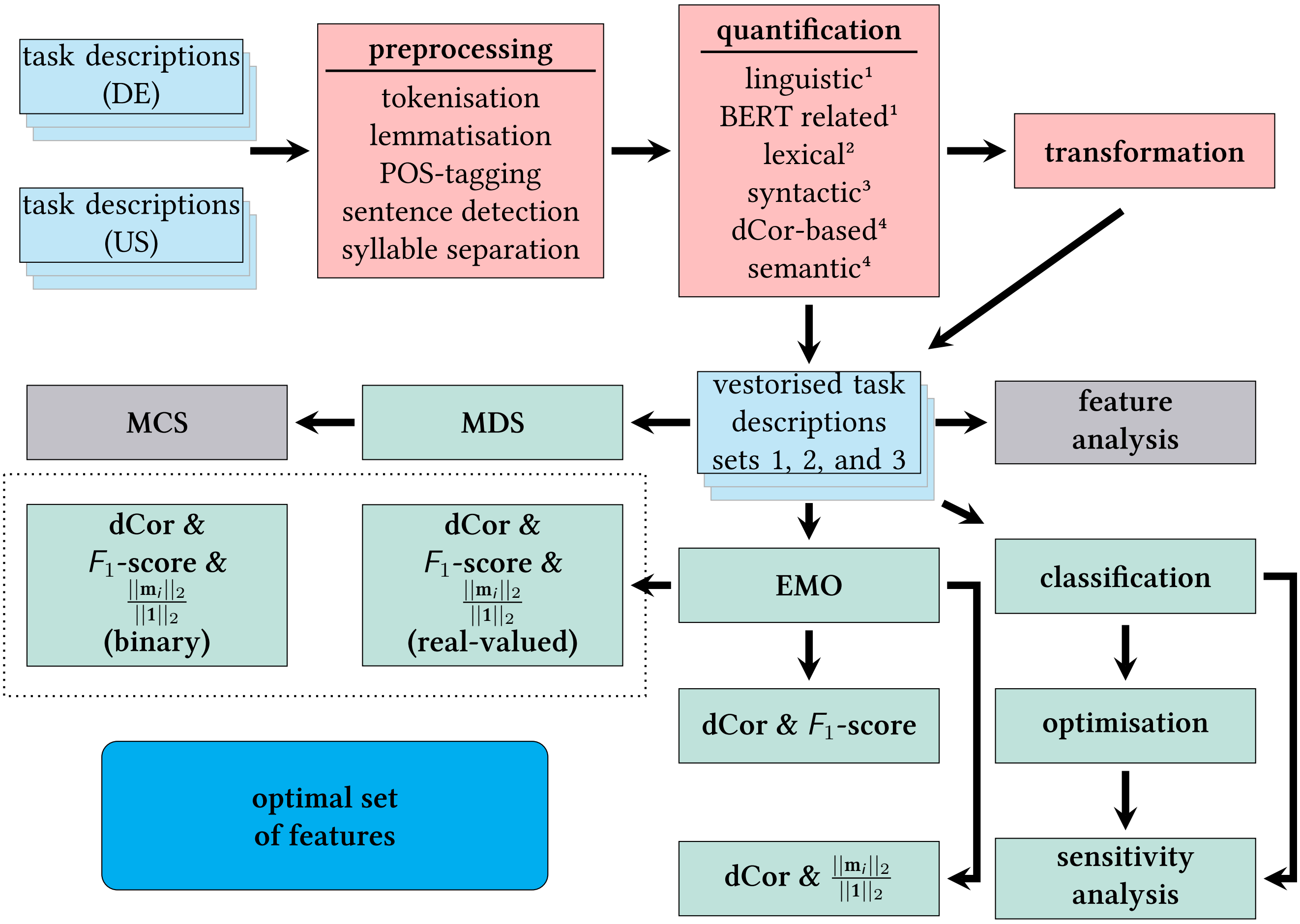


Figure 2.1.: The precessing stages: red - preliminary, green - optimisation stages, each producing own optimal set of features. MDS - Maximum Distance correlation (dCor) Search, MCS - Monte-Carlo Simulation, EMO - Evolutionary Multiobjective Optimisation (two and three objectives). $\frac{||\mathbf{m}_i||_2}{||\mathbf{1}||_2}$ - size/sparsity penalty.

Table 2.1.: Data overview. DE - German, US - English task descriptions.

	DE		US	
	Questions	Answers	Questions	Answers
∅ number of tokens	26	35	28	37
∅ sentence length	14	6	18	13
∅ token length	6	6	4	5

2.1. Data Overview

The data analysed in this thesis is a collection of 60 task descriptions used in the TUCE (Walstad, Watts and Rebeck, 2007). The tasks are multiple-choice type tasks, which can be divided into 60 question and 60 answer parts. Thus we can extend our data set by using 60 question and answer (fa), 60 question (f), and 60 answer (a) descriptions separately, which gives us 180 input texts for German (DE) and 180 for English (US) - 360 in total. Additionally, for each question/answer pair the rates of correct student responses were used as a target variable in our distance correlation maximisation experiments (for detailed description of the data see Mehler, Zlatkin-Troitschanskaia et al., 2018).

2.2. Preparation

To prepare the raw input data for the experiments, we performed NLP-based preprocessing including data cleansing, tokenisation, lemmatisation, POS-tagging, sentence detection, and syllable separation using spaCy's `de_core_news_lg`[5] and `en_core_web_lg`[6] models for German and English task descriptions respectively (see Figure 2.1). Next, each of the 236 text characteristics was calculated for all 360 texts.

Due to the short length of the input texts, especially questions part, see (see Table 2.1), not all features could be used for the downstream tasks. Features, that rely on longer token/sentence sequences (such as features based on BERT's next sentence prediction probabilities), did not produce meaningful results for shorter texts, and had to be eliminated. Total 114 features for each of the 360 question (f), answer (a), question and answer (fa) texts could be used for the further analysis.

To make the results comparable with Mehler, Zlatkin-Troitschanskaia et al., 2018 the feature set had to undergo a transformation (see Section 2.4).

[1]Konca et al., 2020
[2]Islam and Mehler, 2013
[3]Mehler, Hemati, Uslu et al., 2018
[4]Mehler, Zlatkin-Troitschanskaia et al., 2018
[5]`https://github.com/explosion/spacy-models/releases//tag/de_core_news_lg-2.3.0`
[6]`https://github.com/explosion/spacy-models/releases//tag/en_core_web_lg-2.3.1`

Table 2.2.: Parameters of quantitative text characteristics used in this study.

Parameter	Description
f_{r_i}	frequency of the rank r_i
h	h-point
N	total number of tokens
r	token's frequency rank
V	vocabulary size (number of types)

2.3. Quantification

In total, 256 features were implemented and applied in this thesis. The linguistic features were previously used in Konca et al., 2020, lexical in Islam and Mehler, 2013, syntactic in Mehler, Hemati, Uslu et al., 2018. The features marked with * are introduced in this thesis.

Linguistic features:

1. **A** – *adjusted modulus* (Kubát, Matlach and Čech, 2014):

$$A = \frac{h^{-1}(f_{r_1}^2 + V^2)^{1/2}}{\log_{10} N}$$

where h is the h-point (see below); for f_{r_1} and V see Table 2.2.

2. **alpha** – *writer's view* (Popescu and Altmann, 2007):

$$\cos\alpha = \frac{-((h-1)(f_{r_1}-h)+(h-1)(V-h))}{((h-1)^2+(f_{r_1}-h)^2)^{1/2}((h-1)^2+(V-h)^2)^{1/2}}$$

3. **ASL** – *average sentence length.*

4. **ATL** – *average token length.*

5. **G** – *Gini coefficient* (Popescu and Altmann, 2006):

$$G = \frac{1}{V}\left(V + 1 - \frac{2}{N}\sum_{i=1}^{V} r_i f_{r_i}\right)$$

6. **h** – *h-point* (Hirsch, 2005; Popescu, 2009): the point, where the token's frequency rank and the frequency itself are equal.

If no such point exists, two neighbouring frequencies, for which $f_{r_i} > r_i$ and $f_{r_j} < r_j$, are used to calculate h as follows:

$$h = \frac{f_{r_i} r_j - f_{r_j} r_i}{r_j - r_i + f_{r_i} - f_{r_j}}$$

7. **H** – *entropy* (Esteban and Morales, 1995):

$$H = \log_2 N - \frac{1}{N} \sum_{i=1}^{V} f_{r_i} \log_2 f_{r_i}$$

8. **hl** – *hapax legomena percentage*: the percentage of unique types.

9. **L** – *curve length* (Popescu, Čech and Altmann, 2011):

$$L = \sum_{i=1}^{V-1} \sqrt{(f_{r_i} - f_{r_{i+1}})^2 + 1}$$

10. **Λ** – *lambda* (Popescu, Čech and Altmann, 2011; Čech, 2015):

$$\Lambda = \frac{L \log_{10} N}{N}$$

where L is the curve length.

11. **Q** – *activity* (Altmann, 1988):

$$Q = \frac{v}{v + a}$$

where v is the number of verbs and a the number of adjectives.

12. **R1** – *vocabulary richness* (Kubát, Matlach and Čech, 2014):

$$R_1 = 1 - \left(\frac{\sum_{i=1}^{h} f_{r_i}}{N} - \frac{h^2}{2N} \right)$$

where h is the h-point.

13. **RR** – *repeat rate* (Kubát, Matlach and Čech, 2014):

$$RR = \frac{1}{N^2} \sum_{i=1}^{V} f_{r_i}^2$$

14. **RRR** – *relative repeat rate* (McIntosh, 1967):

$$RRR = \frac{1 - \sqrt{RR}}{1 - 1/\sqrt{V}}$$

15. **stc** – *secondary thematic concentration* (Čech, Popescu and Altmann, 2013):

$$stc = \sum_{i=1}^{2h} \frac{(2h - r_i) f_{r_i}}{h(2h - 1) f_{r_1}}$$

16. **tc** – *thematic concentration* (Popescu and Altmann, 2011):

$$tc = 2\sum_{i=1}^{T} \frac{(h - r_i) f_{r_i}}{h(h-1) f_{r_1}}$$

where T is the number of autosemantic words whose rank r is above the h-point.

17. **ttr** – *type-token-ratio* (Wimmer, 2005):

$$ttr = \frac{V}{N}$$

18. **UG** – *unique trigrams*: the ratio of the number of hapax legomena and the total number of character-based 3-grams that can be generated by them.

19. **VD** – *verb distances*: the average distance between verbs, measured by the number of tokens. Following Kubát, Matlach and Čech, 2014, all verbs were considered.

Lexical features: [7]

20. **adjpd** – number of adjectives
21. **advpd** – number of adverbs
22. **ipd** – number of interjections
23. **NDW** – number of "difficult" words, consisting of more than 8 letters
24. **ppd** – number of pronouns
25. **apd** – number of auxiliary words
26. **dpd** – number of determiners
27. **npd** – number of nouns
28. **preppd** – number of prepositions
29. **vpd** – number of verbs

Syntactic features:[8] Each sentence of the input text was parsed with spaCy dependency parser. Resulting dependency trees were analysed using 11 features (see features 30 to 40 below). To get one score for each text, the individual scores for each tree were aggregated using six aggregation functions: mean μ, Gini coefficient (Gini, 1912) (G*), distance autocorrelation (adc*) with lag 1, recursive autocorrelation (rac*[9]), recursive distance autocorrelation (radc*), dynamic time wrapping (adtw*[10]), and entropy (H) - 77 syntactic features in total.

[7] Following Islam and Mehler, 2013
[8] For the definitions of the formulas see Mehler, Hemati, Uslu et al., 2018
[9] see definition of feature 42
[10] calculated using pyts package (Faouzi and Janati, 2020)

30. **c** – comp(T) – Formula (1)
31. **dep** – depend(T) – Formula (4)
32. **LDE** – LDE(T) – Formula (8)
33. **MDD** – MDD(T) – Formula (10)
34. **DDE** – DDE(T) – Formula (12)
35. **TCI** – TCI(T) – Formula (15)
36. **imb** – imbalance(T) – Formula (17)
37. **L** – $\hat{L}$(T) – Formula (18)
38. **w** – width(T) – Formula (23)
39. **l** – level(T) – Formula (24)
40. **W** – $\hat{W}$(T) – Formula (25)

BERT[11] related features:

41. **btac1, btac2, ..., btac10** – Pearson autocorrelation coefficients (Parzen, 1963) calculated from BERT's (Devlin et al., 2018) masked token prediction probabilities with lags 1 to 10.
42. **btrac*** – recursive (Pearson) autocorrelation coefficient

$$X^{(1)}_{k^{(1)}} = \frac{1}{n^{(0)} var^{(0)}} \sum_{t=1}^{n^{(0)}-k^{(1)}} (X^{(0)}_t - \mu^{(0)})(X^{(0)}_{t+k^{(1)}} - \mu^{(0)})$$

$$X^{(l)}_{k^{(l)}} = \frac{1}{n^{(l-1)} var^{(l-1)}} \sum_{t=1}^{n^{(l-1)}-k^{(l)}} (X^{(l-1)}_t - \mu^{(l-1)})(X^{(l-1)}_{t+k^{(l)}} - \mu^{(l-1)})$$

$$btrac = X_1^{(n^{(0)}-1)}$$

where l is level of recursion, $l \in \{x \in \mathbb{N} | 1 \leq x \leq n^{(0)} - 1\}$; k the lag, $k^{(l)} \in \{x \in \mathbb{N} | 1 \leq x \leq n^{(l-1)} - 1\}$; $\mu^{(l)}$ the mean; $var^{(l)}$ the variance. On the 0's level the X_i^0 is ith token prediction probability, $i \in \{x \in \mathbb{N}, 1 \leq x \leq n^{(0)}\}$, $\mu^{(0)}$ and $var^{(0)}$ are the mean and the variance of the token prediction probabilities, and $n^{(0)}$ is the number of predictions

43. **btadc1, btadc2, ..., btadc10*** – distance (`dCor`) (Székely, Rizzo, Bakirov et al., 2007) autocorrelation coefficients calculated from BERT's masked token prediction probabilities with lags 1 to 10

[11]calculated using huggingface transformers python library (Wolf et al., 2020)

44. **btradc*** – recursive distance autocorrelation coefficient
45. **btly*** – the largest Lyapunov exponent (Eckmann et al., 1986) of the series of masked token prediction probabilities[12] as a measure of the degree of chaos in the sequence of next token probabilities
46. **bth*** – Hurst (Harold Edwin Hurst, 1957 and Harold E Hurst, 1956) exponent as a measure of the long-term memory of the sequence of next token probabilities
47. **btdfa*** – detrended fluctuation analysis (Hardstone et al., 2012)
48. **btH*** – entropy of token prediction probabilities
49. **btsH*** – sample entropy token prediction probabilities
50. **bstsim*** – average cosine similarity of the BERT encoded sentence vectors
51. **bsesim*** – average cosine similarity of the BERT sentence embedding vectors
52. **bsacn*** – (per sentence) average number of BERT's sentence embedding distance (`dCor`) similarity clusters (calculated by means of HDBSCAN (McInnes, Healy and Astels, 2017))
53. **bsH*** – entropy of sentence prediction probabilities
54. **bsac1**, **bsac2**, ..., **bsac10** – Pearson autocorrelation coefficients (Parzen, 1963) calculated from BERT's (Devlin et al., 2018) masked token prediction probabilities with lags 1 to 10
55. **bsrac*** – recursive (Pearson) autocorrelation coefficient
56. **bsadc1**, **bsadc2**, ..., **bsadc10*** – distance (`dCor`) (Székely, Rizzo, Bakirov et al., 2007) autocorrelation coefficients calculated from BERT's masked token prediction probabilities with lags 1 to 10
57. **btradc*** – recursive distance autocorrelation coefficient

Additionally, the features from Mehler, Zlatkin-Troitschanskaia et al., 2018 were implemented (see Table A.2 for complete list).

2.4. Feature Transformation

To preserve consistency with the experiments of Mehler, Zlatkin-Troitschanskaia et al., 2018, the calculation of the distance correlation between the target variable (the rate of correct answers) and vectorised task descriptions required a transformation of the original feature set (see Table 2.4). The questions and answers (fa), questions (f), and answers (a) formed two

[12]calculated using nolds package (Schölzel, 2019)

additional sets containing vectorised task descriptions in each language - German (DE) and English (US). In total, both new (stacked) sets contained 324 features for each of the 60 task descriptions. The downstream tasks used the most appropriate representation of the data.

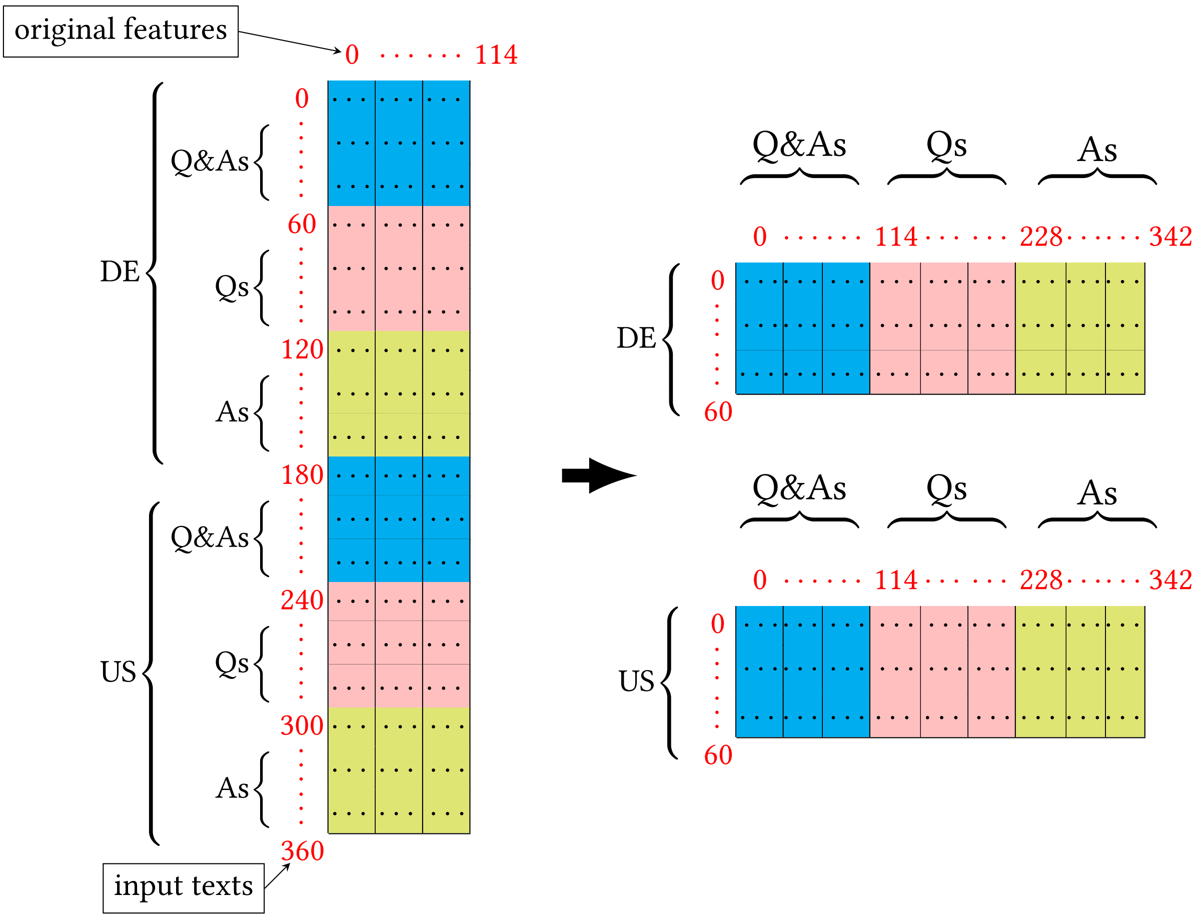


Figure 2.2.: Feature set transformation. DE - German, US - English; Q - questions, A - answers, Q&A - questions and answers.

Figure 2.3 shows the distance correlation of the features used in this experiment before the transformation. Apart from several small intragroup clusters the features show a low degree of similarity, which can be interpreted as a sign of the feature orthogonality, i.e. each feature (group) captures different aspects of a text. Figure 2.3 also shows the distance correlation of the quantified task descriptions. Six clusters are recognisable - from top left to bottom right: German (DE) questions and answers (fa), questions (f), and answers (a), and English (US) questions and answers (fa), questions (f), and answers (a). Figure 2.4 shows the distance correlations after the transformation was performed. Now the feature correlations form three similarity clusters (fa, f, and a), whereas the distance correlations of the vestorised task descriptions became noisier.

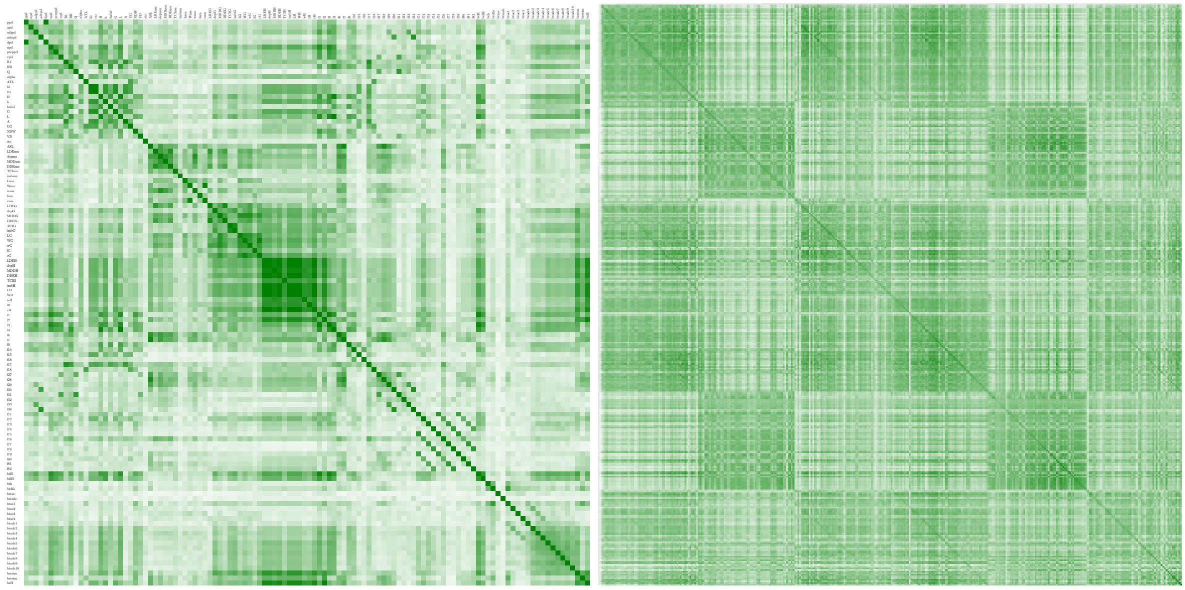


Figure 2.3.: Distance Correlation of text features (left) and vectorised TUCE task descriptions (right).

2.5. Analysis

The following analysis was performed on the resulting feature sets:

1. *Maximum Distance Correlation Search* (MDS) - an evolutionary[13] search with population size of 2000 individuals iterating for 2000 generations.

2. *Monte-Carlo Simulation* (MCS) test of the significance of MDS results.

3. Network-theoretic feature analysis.

4. Classification experiments by means of the *Random Forest Classifier* (RFC). The RFC was chosen for all classification experiments, inasmuch as it is easy to use, has few hyper-parameters (which could be left unaltered without hampering the results), and has a high degree of interpretability. Gini impurity[14] was used as a splitting criterion.

5. EMO for simultaneous optimisation of:
    - $F1$-score and sparsity penalty
    - distance correlation and sparsity penalty
    - distance correlation and $F1$-score
    - distance correlation, sparsity penalty, and $F1$-score (binary mask)
    - distance correlation, sparsity penalty, and $F1$-score (real-valued mask)

[13] using DEAP (Fortin et al., 2012) python package

[14] $gini(t) = 1 - \sum_{j=1}^{k} p^2(j|t)$, where $p(j|t)$ is a probability to pick a point with label i splitting the tree t

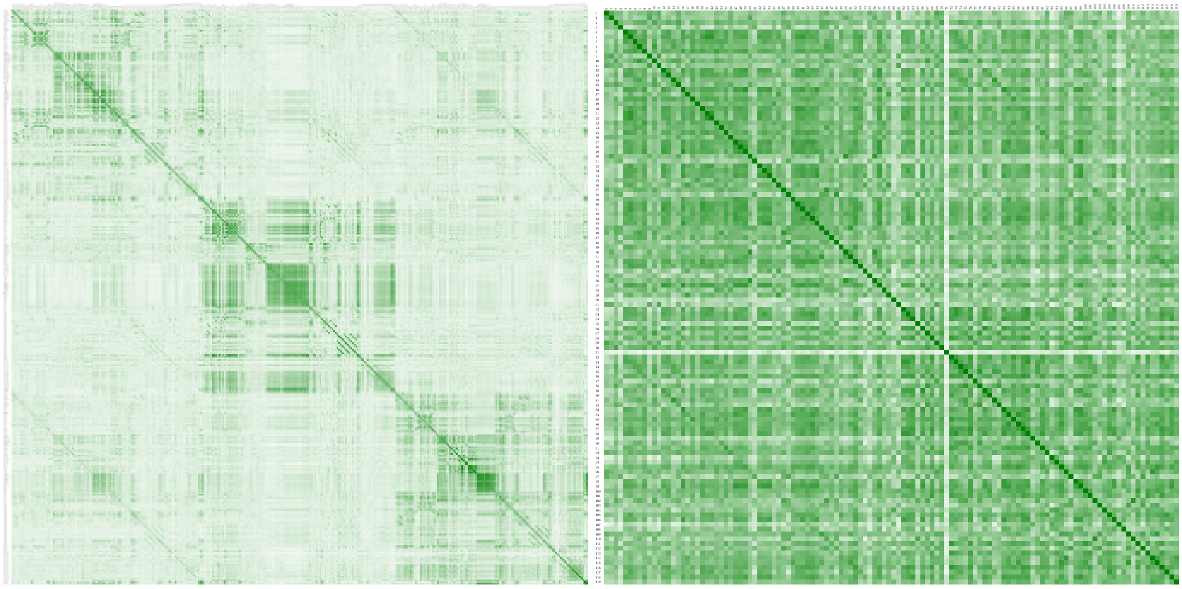


Figure 2.4.: Distance Correlation of text features (left) and vectorised TUCE task descriptions (right) after the transformation (see Section 2.4).

The use of EMO had several purposes:

- To test, if the evolutionary search could isolate features responsible for high F_1-score. Since the separation of German and English task descriptions using all 342 features was trivial, there was, presumably, a high degree of redundancy in the model.
- Provided, the evolutionary search works well on easier task, to test, whether a set of features could be identified, that optimises both the language separability (F_1-score of classification experiment) and distance correlation with the rate of correct responses.
- Narrow down the search for the most salient features by applying a sparsity penalty.
- Emulate the sensitivity analysis by using real-valued mask for feature weighting.

For two objective optimisation and binary three objective optimisation the NSGA-II (Deb, Pratap et al., 2002) algorithm was applied. For real-valued three objective optimisation we used the NSGA-III (Deb and Jain, 2013) algorithm.

6. Sensitivity analysis of the classification using methods of Sobol (Sobol, 2001; Saltelli et al., 2010; Saltelli, 2002) and Morris (Campolongo, Cariboni and Saltelli, 2007; Morris, 1991).

This way we arrive at several (task-specific) optimal feature sets.

3. Results

3.1. Correlation Analysis

The results of Mehler, Zlatkin-Troitschanskaia et al., 2018 could be improved in almost every scenario (see Tables 3.1, 3.2, 3.3, 3.4, 3.5, and Figure 3.2). However, the magnitude of the improvements was mostly small. It can be concluded, that our findings support the hypotheses, that certain characteristics of the task descriptions are related to the rates of correct student responses, and, that there are differences between languages (Hypothesis 1a and 1b in Mehler, Zlatkin-Troitschanskaia et al., 2018). However, our results strongly suggest, that this relation gets stronger with years of study. For German task descriptions the distance correlations increase from 0.47 (optimised for the 1st year) to 0.52 (3rd year); distance correlations for English task description show similar, but weaker trend - distance correlations increase from 0.46 (1st year) to 0.48 (3rd year of study), which is a contradiction to the Hypothesis 3.

In contrast to Mehler, Zlatkin-Troitschanskaia et al., 2018, we could identify 20 features, that were selected in all 8 experiments of the evolutionary search (see Figure 3.1). They are: btrac(fa), bsH(f), TCIH(fa), btac1(f), btadc4(fa), btac4(fa), Lmu(f), hl(a), f31(a), btac3(fa), WG(fa), btac2(fa), LDEmu(f), MDDmu(f), imbG(a), ATL(fa), wG(f), ATL(f), f79(fa), and alpha(fa) (see Tables A.1 to A.12 MDS columns).

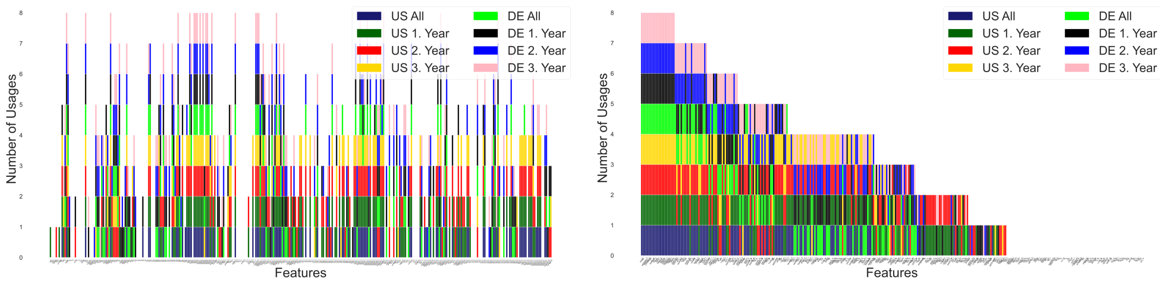


Figure 3.1.: Frequency distribution of features selected in each setting of evolutionary search. DE- German, US - English task descriptions,. Left - unsorted, right - sorted.

3.2. Monte-Carlo Simulation

We performed the same Monte-Carlo test, as Mehler, Zlatkin-Troitschanskaia et al., 2018: the features were permuted and average distance correlations of 10,000 permutations together with the probability, that a random permutation shows at least as high distance correlation, as the optimised, were calculated (see the results in Table 3.6). The probabilities differ in both directions, the average distance correlations, however, are considerably higher than in

Table 3.1.: Distance correlations of the full set of features before applying evolutionary search algorithm; DE - German, US - English task descriptions. Top 20, 10, and 5 of the features showing largest distance correlation with target variable (rate of correct responses) calculated individually.

	DE	US
All features	0.28	0.27
Top 20 features	0.33	0.30
Top 10 features	0.37	0.39
Top 5 features	0.41	0.41

(a) All

	DE	US
All features	0.25	0.28
Top 20 features	0.30	0.39
Top 10 features	0.34	0.38
Top 5 features	0.39	0.41

(b) 1st year

	DE	US
All features	0.29	0.27
Top 20 features	0.34	0.37
Top 10 features	0.37	0.39
Top 5 features	0.42	0.41

(c) 2nd year

	DE	US
All features	0.28	0.26
Top 20 features	0.35	0.33
Top 10 features	0.47	0.42
Top 5 features	0.46	0.38

(d) 3rd year

Table 3.2.: Distance correlations after applying evolutionary algorithm optimising for all years of study. DE - German, US - English task descriptions.

	DE	US
All	0.50	0.47
1. Year	0.47	0.47
2. Year	0.51	0.47
3. Year	0.52	0.47

Mehler, Zlatkin-Troitschanskaia et al., 2018 (e.g. 0.40 compared to 0.22 for all years of study). However, in most experiments the main mass lies significantly lower than the optimised distance correlation (see Figure 3.3).

We also examined the interdependence of the average distance correlation and the number of features selected (see Figure 3.4). Without any feature selection (all features, no mask applied) there is no correlation observable (left Figure). However, if the mask is applied the average distance correlation increases, which also supports the feature non-randomness assumption.

Table 3.3.: Distance correlations after applying evolutionary algorithm optimising for first year of study. DE - German, US - English task descriptions.

	DE	US
1. Year	0.47	0.46
All	0.50	0.45
2. Year	0.50	0.45
3. Year	0.51	0.45

Table 3.4.: Distance correlations after applying evolutionary algorithm optimising for second year of study. DE - German, US - English task descriptions.

	DE	US
2. Year	0.51	0.46
All	0.50	0.45
1. Year	0.47	0.45
3. Year	0.51	0.45

3.3. Feature Analysis

The network cohesion values (the ratio of the number of the edges of a graph and the number of edges in a completely connected graph of the same order, see Mehler, Zlatkin-Troitschanskaia et al., 2018) do not show any decline until the distance correlation bound is decreased to approximately 0.1. The all year experiments (both DE and US) show faster declines in cohesion values, whereas the cohesion values for DE 1st year exhibit relatively high values even at the 0.5 distance correlation bound. The fraction of largest component stays almost 1.0 till the 0.5 bound is reached. These facts suggest, that the model presented in this thesis has a higher degree of redundancy, which could be exploited in the EMO experiments.

3.4. Classification

To evaluate the consistency of the model, and test whether the additional features (e.g. BERT related or syntactic tree based) improve the classification results of Mehler, Zlatkin-Troitschanskaia et al., 2018, several different classification experiments were carried out.

- A four-class experiment, where the classifier had to find a separation between German questions, German answers, English questions, and English answers. This classification task was the most difficult, producing the lowest F_1-score. Nevertheless, the results, obtained using new features improved from 0.80 to 0.83 unoptimised and 0.88 optimised (see Figure 3.6)
- Less challenging (see Figure 3.7 Q vs. A) task was to separate the questions from the answers. The classifier could reach the F_1-score of 0.89 (unpotimised) or 0.935 (optim-

Table 3.5.: Distance correlations after applying evolutionary algorithm optimising for third year of study. DE - German, US - English task descriptions.

	DE	US
3. Year	0.52	0.48
All	0.50	0.46
1. Year	0.47	0.45
2. Year	0.51	0.46

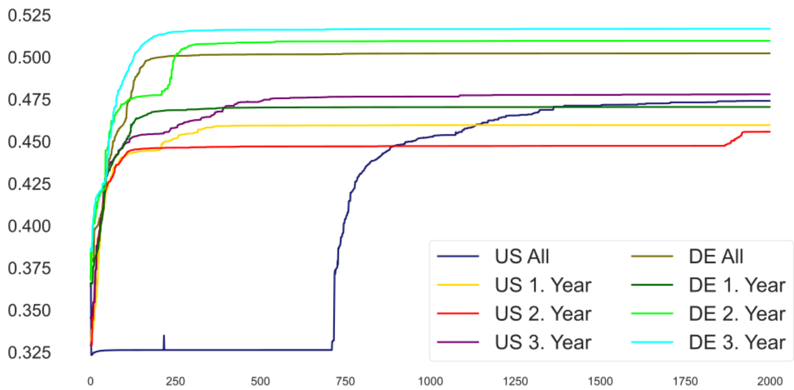


Figure 3.2.: Distance correlation as a function of the number of iterations of the evolutionary search. DE - German, US - English task descriptions.

ised). The result is not surprising - the most obvious difference between questions and answers, was the length of the texts.

- The most interesting question, considering the focus of this study, was how well would the classifier perform on language separation task. The experiment was split into two using *i*) original and *ii*) stacked (see Figure 2.1 and Figure 2.4) versions of the data set. The idea of Mehler, Zlatkin-Troitschanskaia et al., 2018 to generate new features by using the additional types of input texts, hence, making the classes more homogeneous without any information loss, had a positive impact on the results. On the original data set the F_1-score of 0.93 (unoptimised) or 0.97 (optimised) was reached, on the stacked version the scores 0.97 (unoptimised) and 1.0 (optimised) could be obtained (see Figure 3.7). However, optimisation of the F_1-score alone had mostly negative impact on the distance correlation with the rate of correct responses (compare Table 3.1 and Table 3.7). This suggests, that both tasks are not highly correlated implicitly. Possible reason is the vast number of available feature combinations, that produce high F_1-score; i.e. not every combination, that can separate German task descriptions from their English counterparts is a good predictor of the rate of correct responses. The triviality of the class separation is also supported by the *Error Based Feature Elimination with Forests* EBFEf experiment with 20 randomly selected features (see Figure 3.8). In each run the random number generator was initialised with different seed, i.e. the

Table 3.6.: Results of the Monte-Carlo simulation for German (DE) and English (US) task descriptions. Left: probability of reaching the maximum distance correlation; Right: average distance correlation reached during the simulation.

	DE	US
All	0.0031	0.0094
1. Year	0.0164	0.0014
2. Year	0.0000	0.0003
3. Year	0.0001	0.0000

	DE	US
All	0.40	0.40
1. Year	0.41	0.39
2. Year	0.36	0.37
3. Year	0.37	0.36

starting conditions, the train/test set split points varied in each execution.

Observed degree of task orthogonality demands an explicit objectives coupling, that enforces the optimisation along both directions.

Table 3.7.: Distance correlations between the set of features, that achieved highest F_1-score, and the rate of correct student responses (All, 1st year, 2nd year, and 3rd year); DE - German, US - English task descriptions.

	DE	US
All features	0.26	0.24
Top 20 features	0.32	0.30
Top 10 features	0.32	0.33
Top 5 features	0.43	0.37

(a) All

	DE	US
All features	0.24	0.26
Top 20 features	0.30	0.31
Top 10 features	0.30	0.38
Top 5 features	0.40	0.37

(b) 1st year

	DE	US
All features	0.28	0.25
Top 20 features	0.33	0.32
Top 10 features	0.33	0.35
Top 5 features	0.43	0.37

(c) 2nd year

	DE	US
All features	0.27	0.23
Top 20 features	0.35	0.33
Top 10 features	0.35	0.36
Top 5 features	0.35	0.39

(d) 3rd year

[1]Konca et al., 2020

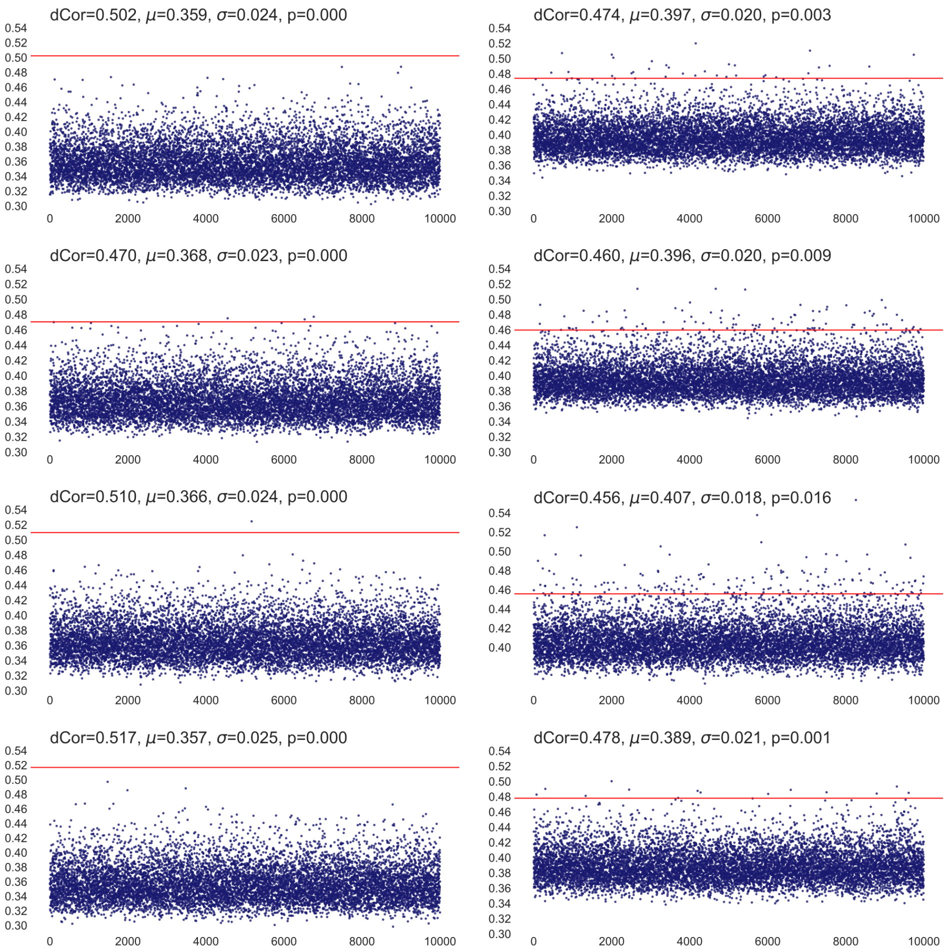


Figure 3.3.: Results of the Monte-Carlo simulations for German (left column) and English (right column) task descriptions; top to bottom: all years, 1st year, 2nd year, and 3rd year. Red horizontal line represents the maximum distance correlation, reached during the evolutionary search algorithm.

3.5. Evolutionary Multi Objective Optimisation

3.5.1. Two Objective Optimisation

We performed several two objective optimisation experiments. First, we optimised F_1-score with sparsity penalty. Initial unpotimised F_1-score of all 342 features was 0.97. After the first generation of the evolutionary search 50% of the features were eliminated yielding the F_1-score of 1.0, which was maintained until the number of features dropped to 9. The remaining features were ATL(fa), lG(fa), btadc6(fa), bsesim(fa), adjpd(f), Q(f), DDEG(f), MDDH(a), btadc2(a).

The simultaneous optimisation of distance correlation and F_1-score after 2000 generations of evolutionary search yielded dCor of 0.50 and F1-Score of 1.0, selecting 125 features (see Figure 3.9). Notice the F_1-score isolines, the multitude of different feature combinations allow

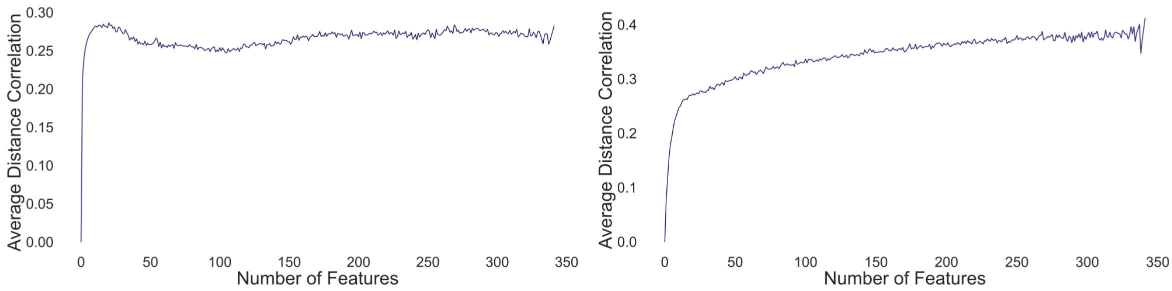


Figure 3.4.: Monte-Carlo simulation: average distance correlation dependence on the number of selected features; left: all features, right: after the mask application.

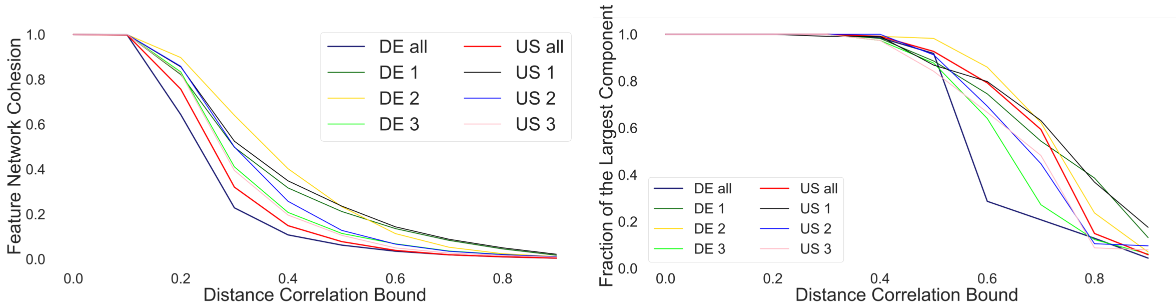


Figure 3.5.: Network cohesion (left). Fraction of vertices belonging to the largest component (right).

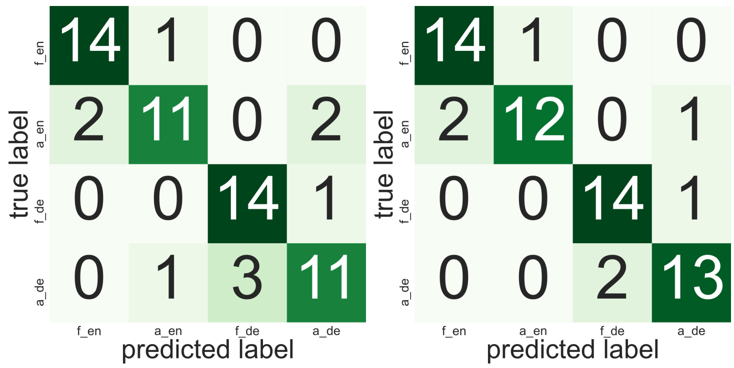
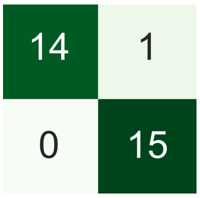


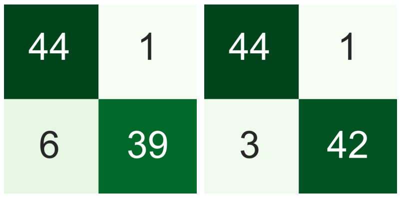
Figure 3.6.: Confusion matrices of the four-class experiment: German questions (f_de) versus German answers (a_de) versus English questions (f_en) versus English answers (a_en). Unoptimised(left), optimised (right)

the optimisation of the distance correlation along the constant F_1-score of 1.0.

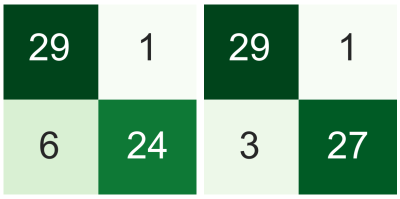
Sparsity constrain had almost no impact on distance correlation optimisation (see Figure 3.10. A combination of 39 features reached distance correlation of 0.49, falling just short of the results of evolutionary search with one objective (MDS). See Tables A.1 to A.12 (dCor+ columns) for features, that were selected by the two objective optimisation algorithms.



(a) DE vs. US (stacked)



(b) DE vs. US



(c) Q vs. A

Figure 3.7.: Confusion matrices of the binary classifications; German (DE) vs. English (US); Questions (Q) vs. Answers (A). The stacked version uses the same feature generation process as the evolutionary search for maximum distance correlation.

3.5.2. Three Objective Optimisation

To find a three-dimensional Pareto manifold and, hence, reduce the number of feature combinations, the simultaneous optimisation of distance correlation and F_1-score under the sparsity constrain was performed (see Figures 3.11 for German and 3.12 for English task descriptions). As expected, the F_1-score shows similar structure, as in the two objective experiments - there exist several isoplanes. For German task descriptions the Pareto manifold is mostly flat, until the number of selected features reaches approximately 40, after which there is a rapid drop of the dCor. English task descriptions show similar, but more perturbed manifold. Another major difference is the subset of selected features (see Tables A.1 to A.12 EMO columns). Only few features were selected in both experiments - bsesim(fa), f73(fa), ATL(fa/f), stc(fa), TCIG(fa), lH(fa), bsH(f), f2(f), f10(f), MDDmu(f), TCImu(f), lmmu(f), f80(a), advpd(a), preppd(a), ttr(a) - 17 out of 342.

3.5.3. Three Objective Optimisation revisited

Not surprisingly, additional constraint did not improve the results of the optimisation. For example, in DE All years experiment after 2000 generations dCor of 0.50 and F1-Score of 0.9666 (54 features selected) could be obtained. To circumvent the rigidity of the binary mask restrictions, an alternative method of the three way optimisation was implemented (see Figure 3.13). As before this method searched simultaneously for maximum distance correlation and maximum F_1-score under the sparsity constraint, expressed as the ratio of the euclidean norms of the mask vector and the unity vector. The modification was aimed to relax the binarity constrain, and allow individual gene to take values between 0 and 1 (and not just 0 or 1). Such genes could be interpreted as weights or measures of importance of each feature, hence, act as an additional sensitivity measure. That resulted in the significant drop of the euclidean norm ratio, especially in the experiment with English (US) task descriptions.

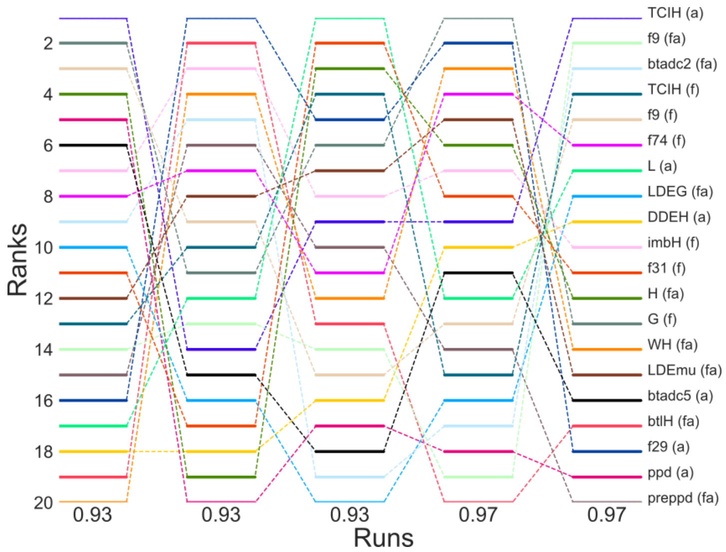


Figure 3.8.: Ranking by *Error Based Feature Elimination using forests* (EBFEf)[1]of 20 randomly selected features.

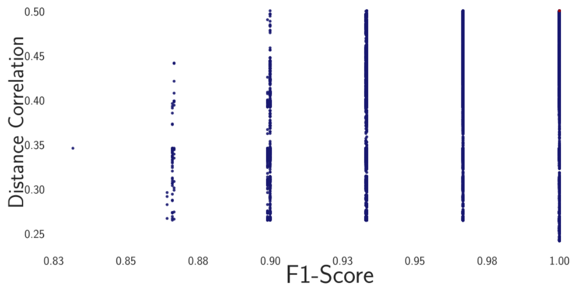


Figure 3.9.: Distance correlations dependence on the F_1-score (two objective optimisation).

In DE all years experiment dCor of 0.51, F_1-score of 0.9666 and norm ratio of 0.14 were obtained. The dominating features were lmu(f) with weight 1.0, preppd(a) - 0.72, and f28(fa) - 0.72 (see Tables A.13 to A.30).

In US all years experiment dCor of 0.54, F_1-score of 0.9666 and norm ratio of 0.01, which is order of magnitude lower than in DE experiment, were reached. The dominating features were depG(a) with weight 0.13, bth((fa) - 0.07, and stc(fa) - 0.07. The drawback of this approach was the increased difficulty of the interpretation of the absolute number of selected features, still allowing the relative comparison e.g. 0.14 to 0.01.

3.6. Sensitivity Analysis

We also performed two kinds of sensitivity analyses - *i)* classification based and *ii)* distance correlation based. Each kind by the means of two methods (see Tables A.13 to A.30). The classification based sensitivity analysis by the method of Sobol (see Figure 3.14[2] a) and b))

[2]Courtesy of `https://github.com/antonia-had/Radial_convergence_plot`

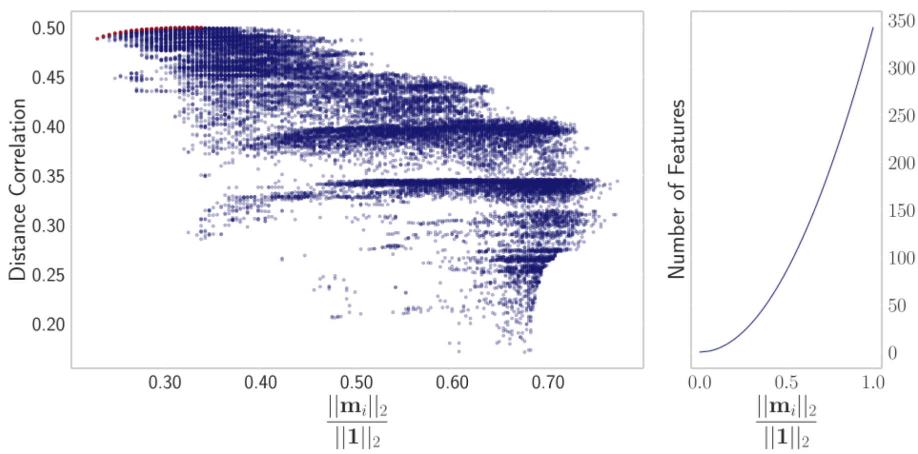


Figure 3.10.: Distance correlations dependence on the sparsity constrain (number of features selected, two objective optimisation).

showed differences between experiments with original and transformed (stacked) feature sets. Features, both experiments had in common, were the bsesim, bstsim, bsH, and to some extent ATL/ASL.

In addition to the classification based, we experimented with the distance correlation based sensitivity analysis, which is a continuation of the idea to use real-valued feature weighting (described in Section 3.5.3). It allows features to be either fully selected (weight 1.0), partly selected (weight $\in (0, 1)$) or not selected (weight 0.0). This way the sensitivity analysis can be applied to the dCor optimisation. Figure 3.14 c) and d) shows the results of the experiment. The f75(fa), which is the number of *General-Language Words* (see Mehler, Zlatkin-Troitschanskaia et al., 2018 Section 4.1 for detailed description) members in relation to text length, calculated on questions and answers (fa) subset, shows a multitude of second order interactions with syntactic, BERT related and other features (DE experiment). In the US experiment the imbmu(f) - the average imbalance of the syntactic trees, plays central role. It has the highest sensitivity and a variety of second order interactions with other features. The wH(f), feature measuring entropy of the widths of syntactic trees, also plays an important role.

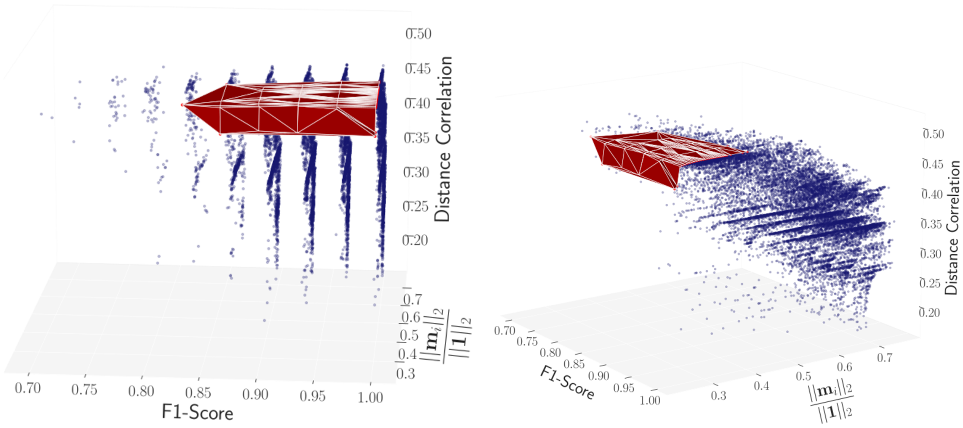


Figure 3.11.: Distance correlations dependence on F_1-score and sparsity constrain (number of features selected, binary mask) of the German task descriptions; Pareto frontier (red).

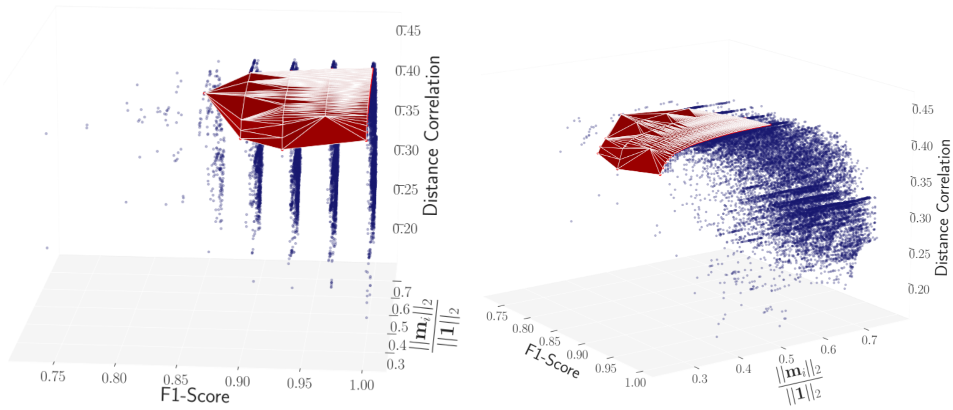


Figure 3.12.: Distance correlations dependence on F_1-score and sparsity constrain (number of features selected, binary mask) of the English task descriptions; Pareto frontier (red).

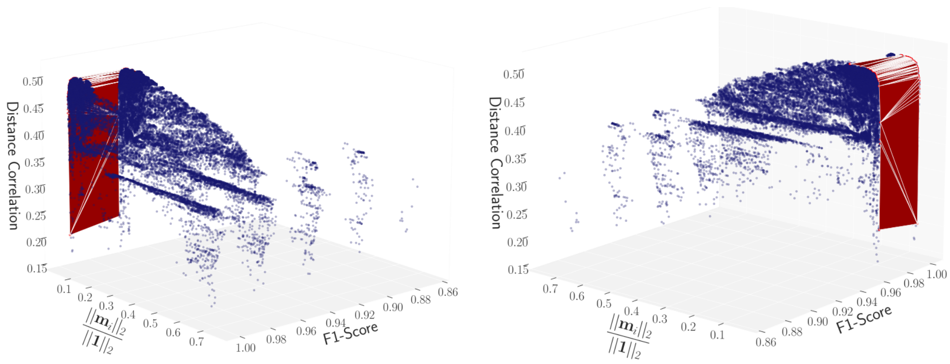
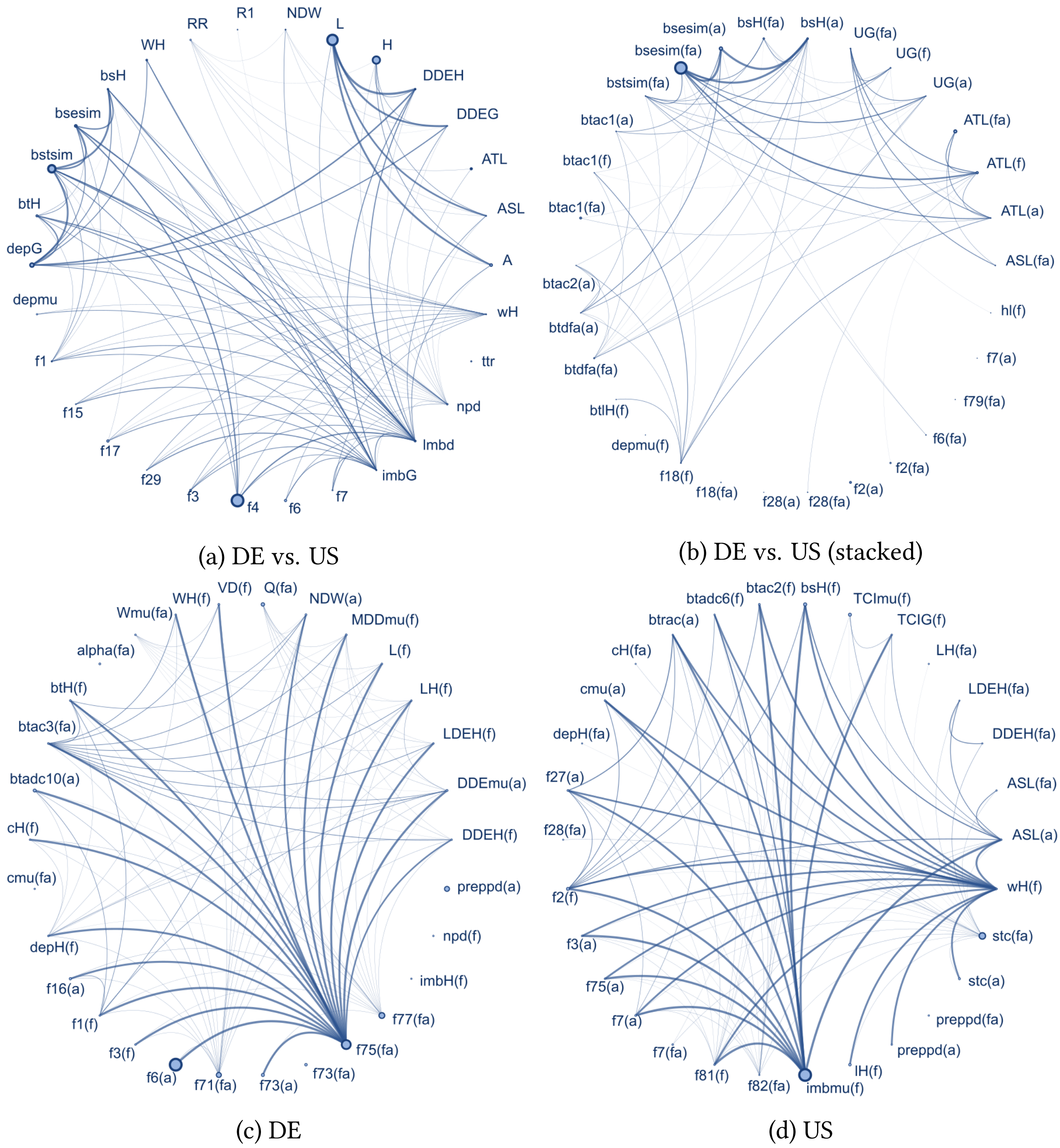


Figure 3.13.: Distance correlations dependence on F_1-score and sparsity constrain (real-valued mask) and the Pareto frontier (red).



(a) DE vs. US

(b) DE vs. US (stacked)

(c) DE

(d) US

Figure 3.14.: Sobol sensitivity. a) and b) DE vs. US classification; c) and d) - distance correlation based sensitivity (only top 30 features are shown). The node size indicates the first order index (S1) per parameter, the node border thickness indicates the total order index (ST) per parameter, and the thickness of the line between two nodes indicates the second order index (S2).

4. Discussion

From results presented so far, it is apparent, that first - there exist a clear correlation between a variety of linguistic, lexical, syntactic, and semantic features of task descriptions and the rate of correct student responses, and second - different input classes, e.g. different forms (questions/answers) or different languages (German/English) exhibit divergent and distinguishable structure. The model described in this thesis is able to illuminate this hidden structure from different angles. It performs well in classification experiments and in distance correlation maximisation by its size alone. Thus, it becomes important to be able to dissect and highlight the most important features, that can be universally applied across a variety of different experimental settings. Figure 4.1 shows the summary of the importance

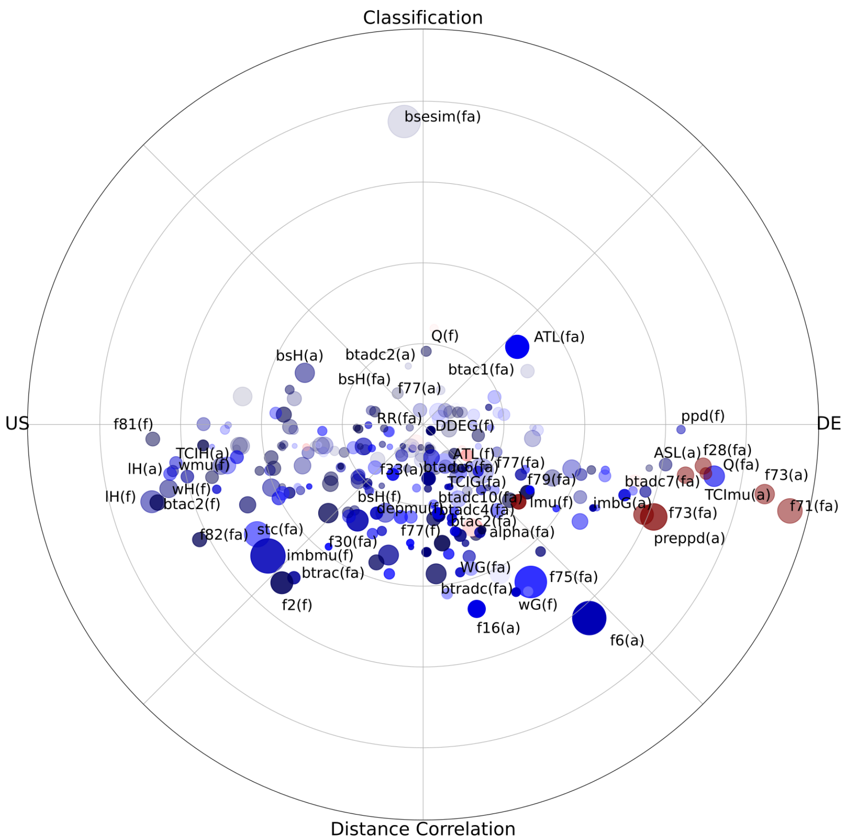


Figure 4.1.: Feature importance. Size of the points reflects sensitivity (Sobol and Morris); colour - real-valued NSGA3 feature weighting; The central points represent features, that were equally important for language separation and distance correlation maximisation simultaneously for US and DE experiments.

scores of analysed features in different experiments. Ideally, all features would lie inside a small radius near the centre, but the experiments show, that not all features are equally important in all experiments. One clear outlier is bsesim (average cosine similarity of the BERT embedding vectors), which plays a major role in the classification experiments, but is mostly ignored in the distance correlation maximisation, making it a bad predictor of the rate of correct student responses in both US and DE experiments. The rest of the features are situated much closer to the centre, meaning, they tend to have at least some degree of importance in both classification and distance correlation experiments.

Another dimension, along which the features could be ranked is the language (DE-US) axis. Here we can observe a greater variance of the ranks. Some features, such as f71(fa), f73(a), Q(fa), and TCImu(a) are highly biased towards DE side, others, like f81(f), lH(f), btac2(fa) play major role only in US experiments. Nevertheless, there are a multitude of features, that are important in all settings.

As we have seen, additional features improve the distance correlation by relatively small amount. This fact naturally rises the question of saturation of the distance correlation. Moreover, not every clear functional dependence produces high distance correlation score (see Figure 4.2). It is a task for future research to find a way to improve the explainability and comparability of different levels of text similarity based on distance correlation methods.

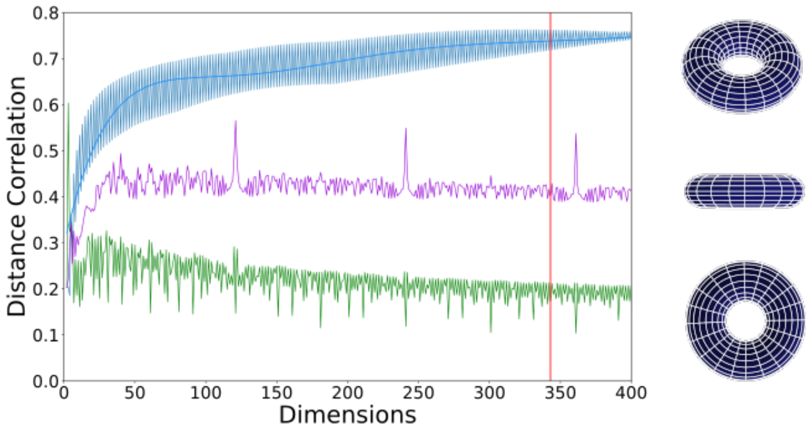


Figure 4.2.: Distance correlation between the torus dimensions depending and their number. The distance correlation between the last two dimensions (blue); the last and all previous dimensions (violet); the first and all subsequent dimensions (green); the level line of the 342nd dimension (red). Example of a torus (three figures on the right).

The counter intuitive increase in distance correlation of task descriptions with the rate of correct student responses over the academic years, which means, that the dependency on a variety of selected textual features is becoming stronger, remains an interesting open question for future research.

5. Conclusion

In this thesis we explored a variety of existing and new text quantification methods in a multilingual setting. We could confirm, that there there exist features, that act as good predictors for the rates of correct student responses. Moreover, extending the set of features used in Mehler, Zlatkin-Troitschanskaia et al., 2018, we could improve distance correlation between a sets of vectorised task descriptions and the rate of correct student responses. We also found, that these results are not random, but are based on a certain underlying structure.

Classification and distance correlation maximisation show no implicit correlation, but are rather orthogonal to each other. This fact demands an explicit constraining, which would force the model in the desired direction. We have also seen, that our model has a high degree of redundancy, when applied to the classification tasks only, possibly overfitting on the amount of data, that was at our disposal.

There are several possible directions future research could follow, that may lead to further improvements - find a better evolutionary search strategy, explore the possibilities of BERT fine tuning (an unaltered pre-trained model was used in this thesis), explore different zoom factors i.e. character based versus paragraph based quantification methods, implying the size of the input texts allows it.

Another auspicious direction would be an automation of the feature generation. In particular, VieNNA (Mehler, Hemati, Gleim et al., 2018) can offer a basis for the generative exploration of the feature space, by providing a framework for both rule based text (feature) representation genesis and genesis of the evaluation algorithms.

Bibliography

Altmann, Gabriel (1988). *Wiederholungen in Texten.* Bochum: Brockmeyer.

Bahdanau, Dzmitry, Kyunghyun Cho and Yoshua Bengio (2014). 'Neural machine translation by jointly learning to align and translate'. In: *arXiv preprint arXiv:1409.0473.*

Campolongo, Francesca, Jessica Cariboni and Andrea Saltelli (2007). 'An effective screening design for sensitivity analysis of large models'. In: *Environmental modelling & software* 22.10, pp. 1509–1518.

Čech, Radek (2015). 'Text length and the lambda frequency structure of a text'. In: *Sequences in Language and Text.* Berlin, Boston: De Gruyter Mouton, pp. 71–88.

Čech, Radek, Ioan-Iovitz Popescu and Gabriel Altmann (2013). 'Methods of analysis of a thematic concentration of the text'. In: *Czech and Slovak Linguistic Review* 3, pp. 4–21.

Cheng, Jianpeng, Li Dong and Mirella Lapata (Nov. 2016). 'Long Short-Term Memory-Networks for Machine Reading'. In: *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing.* Austin, Texas: Association for Computational Linguistics, pp. 551–561. DOI: `10.18653/v1/D16-1053`.

Deb, Kalyanmoy and Himanshu Jain (2013). 'An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, part I: solving problems with box constraints'. In: *IEEE transactions on evolutionary computation* 18.4, pp. 577–601.

Deb, Kalyanmoy, Amrit Pratap, Sameer Agarwal and TAMT Meyarivan (2002). 'A fast and elitist multiobjective genetic algorithm: NSGA-II'. In: *IEEE transactions on evolutionary computation* 6.2, pp. 182–197.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee and Kristina Toutanova (2018). 'BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding'. In: *arXiv preprint arXiv:1810.04805.*

Eckmann, J. -P., S. Oliffson Kamphorst, D. Ruelle and S. Ciliberto (Dec. 1986). 'Liapunov exponents from time series'. In: *Phys. Rev. A* 34 (6), pp. 4971–4979. DOI: `10.1103/PhysRevA.34.4971`. URL: `https://link.aps.org/doi/10.1103/PhysRevA.34.4971`.

Esteban, Maria Dolores and Domingo Morales (1995). 'A summary on entropy statistics'. In: *Kybernetika* 31.4, pp. 337–346.

Faouzi, Johann and Hicham Janati (2020). 'pyts: A Python Package for Time Series Classification'. In: *Journal of Machine Learning Research* 21.46, pp. 1–6. URL: `http://jmlr.org/papers/v21/19-763.html`.

Fortin, Félix-Antoine, François-Michel De Rainville, Marc-André Gardner, Marc Parizeau and Christian Gagné (July 2012). 'DEAP: Evolutionary Algorithms Made Easy'. In: *Journal of Machine Learning Research* 13, pp. 2171–2175.

Gini, Corrado (1912). 'Variabilità e mutabilità (Variability and Mutability)'. In: *Reprinted in Memorie di metodologica statistica (Ed. Pizetti E, Salvemini, T). Rome: Libreria Eredi Virgilio Veschi (1955) ed. Bologna.*

Hardstone, Richard et al. (2012). 'Detrended fluctuation analysis: a scale-free view on neuronal oscillations'. In: *Frontiers in physiology* 3, p. 450.
Hirsch, Jorge E (2005). 'An index to quantify an individual's scientific research output'. In: *Proceedings of the National academy of Sciences* 102.46, pp. 16569–16572.
Hurst, Harold E (1956). 'The problem of long-term storage in reservoirs'. In: *Hydrological Sciences Journal* 1.3, pp. 13–27.
Hurst, Harold Edwin (1957). 'A suggested statistical model of some time series which occur in nature'. In: *Nature* 180.4584, pp. 494–494.
Islam, Zahurul and Alexander Mehler (2013). 'Automatic readability classification of crowd-sourced data based on linguistic and information-theoretic features'. In: *Computación y Sistemas* 17.2, pp. 113–123.
Konca, Maxim, Alexander Mehler, Daniel Baumartz and Wahed Hemati (2020). 'From Distinguishability to Informativity: A Quantitative Text Model for Detecting Random Texts'. In: *Manuscript accepted for publication.*
Kubát, Miroslav, Vladimír Matlach and Radek Čech (2014). 'QUITA'. In: *Quantitative Index Text Analyzer. Lüdenscheid: RAM-Verlag.*
McInnes, Leland, John Healy and Steve Astels (2017). 'hdbscan: Hierarchical density based clustering'. In: *The Journal of Open Source Software* 2.11, p. 205.
McIntosh, Robert P (1967). 'An index of diversity and the relation of certain concepts to diversity'. In: *Ecology* 48.3, pp. 392–404.
Mehler, Alexander, Wahed Hemati, Rüdiger Gleim and Daniel Baumartz (2018). 'VienNA: Auf dem Weg zu einer Infrastruktur für die verteilte interaktive evolutionäre Verarbeitung natürlicher Sprache'. In: *Forschungsinfrastrukturen und digitale Informationssysteme in der germanistischen Sprachwissenschaft* 6, pp. 149–176.
Mehler, Alexander, Wahed Hemati, Tolga Uslu and Andy Lücking (2018). 'A multidimensional model of syntactic dependency trees for authorship attribution'. In: *Quantitative analysis of dependency structures* 72, p. 315.
Mehler, Alexander, Olga Zlatkin-Troitschanskaia et al. (2018). 'Integrating computational linguistic analysis of multilingual learning data and educational measurement approaches to explore learning in higher education'. In: *Positive learning in the age of information.* Springer, pp. 145–193.
Morris, Max D (1991). 'Factorial sampling plans for preliminary computational experiments'. In: *Technometrics* 33.2, pp. 161–174.
Parzen, Emanuel (1963). 'On spectral analysis with missing observations and amplitude modulation'. In: *Sankhyā: The Indian Journal of Statistics, Series A*, pp. 383–392.
Popescu, Ioan-Iovitz (2009). *Word frequency studies.* Vol. 64. Walter de Gruyter.
Popescu, Ioan-Iovitz and Gabriel Altmann (2006). 'Some aspects of word frequencies'. In: *Glottometrics* 13, pp. 23–46.
– (2007). 'Writer's view of text generation'. In: *Glottometrics* 15, pp. 71–81.
– (2011). 'Thematic concentration in texts'. In: *Issues in quantitative linguistics* 2, pp. 110–116.
Popescu, Ioan-Iovitz, Radek Čech and Gabriel Altmann (2011). *The lambda-structure of texts.* Ram-Verlag Lüdenscheid.
Saltelli, Andrea (2002). 'Making best use of model evaluations to compute sensitivity indices'. In: *Computer physics communications* 145.2, pp. 280–297.

Saltelli, Andrea et al. (2010). 'Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index'. In: *Computer Physics Communications* 181.2, pp. 259–270.

Schölzel, Christopher (June 2019). *Nonlinear measures for dynamical systems.* Version 0.5.2. DOI: 10.5281/zenodo.3814723. URL: https://doi.org/10.5281/zenodo.3814723.

Sobol, Ilya M (2001). 'Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates'. In: *Mathematics and computers in simulation* 55.1-3, pp. 271–280.

Székely, Gábor J, Maria L Rizzo, Nail K Bakirov et al. (2007). 'Measuring and testing dependence by correlation of distances'. In: *The annals of statistics* 35.6, pp. 2769–2794.

Vaswani, Ashish et al. (2017). 'Attention is all you need'. In: *Advances in neural information processing systems*, pp. 5998–6008.

Walstad, William B, Michael W Watts and Ken Rebeck (2007). *Test of understanding in college economics: Examiner's manual.* Council for Economic Educat.

Wimmer, Gejza (2005). 'The Type-Token-Relation'. In: *Quantitative Linguistik. Ein internationales Handbuch / Quantitative Linguistics. An International Handbook.* Ed. by Reinhard Köhler, Gabriel Altmann and Raijmund G. Piotrowski. Berlin/New York: De Gruyter, pp. 361–368.

Wolf, Thomas et al. (Oct. 2020). 'Transformers: State-of-the-Art Natural Language Processing'. In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations.* Online: Association for Computational Linguistics, pp. 38–45. URL: https://www.aclweb.org/anthology/2020.emnlp-demos.6.

A. Appendix

Table A.1.: BERT related feature selection in binary experiments. Rows: features, columns: experiments; green: selected. Questions and answers (fa) subset.

	F_1 + sparsity	MDS DE 1. year	MDS DE 2. year	MDS DE 1. year	MDS DE all years	MDS US 1. year	MDS US 2. year	MDS US 3. year	MDS US all years	dCor + sparsity DE	dCor + sparsity US	dCor + F_1 DE	dCor + F_1 US	EMO DE	EMO US
btH(fa)															
btlH(fa)							✓								
bth(fa)						✓	✓	✓	✓				✓		
btdfa(fa)			✓			✓	✓	✓	✓			✓	✓		✓
btrac(fa)			✓	✓	✓	✓	✓	✓	✓						✓
btradc(fa)		✓	✓	✓	✓	✓	✓	✓	✓						
btac1(fa)															
btac2(fa)		✓	✓	✓	✓		✓	✓	✓				✓		
btac3(fa)		✓	✓	✓	✓	✓	✓	✓	✓				✓	✓	
btac4(fa)		✓	✓	✓		✓		✓				✓	✓		✓
btadc1(fa)						✓	✓	✓	✓			✓			
btadc2(fa)		✓	✓	✓	✓	✓	✓		✓				✓		
btadc3(fa)						✓	✓		✓				✓		
btadc4(fa)		✓	✓	✓	✓	✓	✓	✓	✓			✓	✓		
btadc5(fa)		✓	✓		✓										
btadc6(fa)	✓	✓	✓	✓	✓	✓	✓	✓	✓				✓		
btadc7(fa)		✓	✓	✓	✓	✓		✓	✓	✓		✓	✓	✓	
btadc8(fa)		✓	✓		✓	✓		✓	✓			✓	✓		
btadc9(fa)			✓			✓	✓		✓			✓	✓		
btadc10(fa)		✓	✓	✓	✓	✓	✓	✓	✓			✓	✓		
bstsim(fa)						✓	✓	✓	✓				✓		
bsesim(fa)	✓						✓					✓	✓	✓	✓
bsH(fa)											✓				

Table A.2.: F feature (Mehler, Zlatkin-Troitschanskaia et al., 2018 Table 2) selection in binary experiments. Rows: features, columns: experiments; green: selected. Questions and answers (fa) subset.

	F_1 + sparsity	MDS DE 1. year	MDS DE 2. year	MDS DE 1. year	MDS DE all years	MDS US 1. year	MDS US 2. year	MDS US 3. year	MDS US all years	dCor + sparsity DE	dCor + sparsity US	dCor + F_1 DE	dCor + F_1 US	EMO DE	EMO US
f1(fa)															
f2(fa)															
f3(fa)															
f4(fa)															
f6(fa)					■					■		■	■		
f7(fa)															
f9(fa)									■						■
f10(fa)															
f15(fa)		■		■	■	■		■	■				■		
f16(fa)		■	■		■	■	■	■	■				■		
f17(fa)															
f18(fa)															
f27(fa)															
f28(fa)		■	■	■	■					■		■		■	
f29(fa)		■	■	■	■					■		■		■	
f30(fa)		■	■	■		■	■	■	■						■
f31(fa)		■	■	■	■		■	■	■						
f32(fa)							■	■		■			■		
f33(fa)			■	■	■		■	■	■				■		
f34(fa)		■	■	■	■	■		■				■	■		
f71(fa)		■	■	■	■					■		■		■	
f72(fa)															
f73(fa)		■	■	■	■				■	■		■		■	■
f74(fa)															

Table A.3.: Lexical and linguistic feature selection in binary experiments. Rows: features, columns: experiments; green: selected. Questions and answers (fa) subset.

	F_1 + sparsity	MDS DE 1. year	MDS DE 2. year	MDS DE 1. year	MDS DE all years	MDS US 1. year	MDS US 2. year	MDS US 3. year	MDS US all years	dCor + sparsity DE	dCor + sparsity US	dCor + F_1 DE	dCor + F_1 US	EMO DE	EMO US
ppd(fa)													■		
apd(fa)															
adjpd(fa)															
advpd(fa)															
dpd(fa)							■								
npd(fa)															
preppd(fa)						■									
vpd(fa)															
R1(fa)						■	■	■	■			■		■	
RR(fa)		■					■	■	■			■	■		
Q(fa)		■	■	■	■				■	■		■	■	■	
alpha(fa)		■	■	■	■	■	■	■	■	■		■	■		■
ATL(fa)	■	■	■	■	■	■	■	■	■	■		■	■	■	■
hl(fa)		■											■		
ttr(fa)		■		■				■							
H(fa)															
h(fa)															
lmbd(fa)		■	■	■			■						■		
G(fa)				■		■		■	■			■	■		
L(fa)															
A(fa)															
UG(fa)															
NDW(fa)															
VD(fa)															
stc(fa)			■		■	■	■	■	■			■	■	■	■
ASL(fa)															

Table A.4.: Syntactic feature selection in binary experiments. Rows: features, columns: experiments; green: selected. Questions and answers (fa) subset.

	F_1 + sparsity	MDS DE 1. year	MDS DE 2. year	MDS DE 1. year	MDS DE all years	MDS US 1. year	MDS US 2. year	MDS US 3. year	MDS US all years	dCor + sparsity DE	dCor + sparsity US	dCor + F_1 DE	dCor + F_1 US	EMO DE	EMO US
LDEmu(fa)			x		x	x						x			
depmu(fa)															
MDDmu(fa)															
DDEmu(fa)												x			
TCImu(fa)			x												
imbmu(fa)		x	x	x	x				x						
Lmu(fa)		x	x	x			x		x			x			
Wmu(fa)		x	x	x	x					x		x			
wmu(fa)				x						x					
lmu(fa)			x		x	x	x	x	x		x	x	x		x
cmu(fa)		x	x	x	x			x				x			
LDEG(fa)			x	x	x									x	
depG(fa)		x	x			x	x	x	x			x	x	x	
MDDG(fa)		x	x		x			x	x						
DDEG(fa)		x		x	x				x			x	x		
TCIG(fa)		x	x	x	x	x	x	x				x	x	x	x
imbG(fa)						x		x	x						x
LG(fa)			x	x	x		x	x	x						
WG(fa)		x	x	x	x		x	x	x	x	x				
wG(fa)		x		x	x		x	x	x			x	x	x	
lG(fa)	x	x	x	x	x				x			x			
cG(fa)		x	x	x			x	x	x	x		x	x		x
LDEH(fa)		x										x			
depH(fa)		x			x							x			
MDDH(fa)		x	x					x							
DDEH(fa)		x	x									x		x	
TCIH(fa)			x	x		x	x	x	x		x	x	x		x
imbH(fa)		x	x	x	x									x	
LH(fa)		x	x	x						x					
WH(fa)		x			x										
wH(fa)						x	x	x	x				x		x
lH(fa)		x	x	x	x		x		x	x	x			x	x
cH(fa)		x										x			

Table A.5.: BERT related feature selection in binary experiments. Rows: features, columns: experiments; green: selected. Questions (f) subset.

	F_1 + sparsity	MDS DE 1. year	MDS DE 2. year	MDS DE 1. year	MDS DE all years	MDS US 1. year	MDS US 2. year	MDS US 3. year	MDS US all years	dCor + sparsity DE	dCor + sparsity US	dCor + F_1 DE	dCor + F_1 US	EMO DE	EMO US
btH(f)															
btlH(f)		●							●						
bth(f)			●	●		●	●	●	●				●		●
btdfa(f)		●			●	●	●	●	●	●		●	●		●
btrac(f)		●		●	●							●			
btradc(f)			●			●	●	●	●				●		
btac1(f)		●	●	●	●	●		●	●				●		
btac2(f)			●		●	●	●	●	●		●	●	●		●
btac3(f)		●	●	●	●	●	●		●			●	●	●	
btac4(f)				●		●	●	●	●			●	●		
btadc1(f)						●	●	●	●						
btadc2(f)		●		●	●	●	●	●	●				●	●	
btadc3(f)				●	●										
btadc4(f)			●	●	●	●	●		●				●		
btadc5(f)									●				●	●	
btadc6(f)						●	●	●	●		●		●	●	
btadc7(f)		●				●	●	●	●				●		●
btadc8(f)			●	●	●	●	●	●	●				●	●	
btadc9(f)		●	●	●	●		●		●	●					
btadc10(f)		●	●	●	●	●	●		●			●			
bstsim(f)		●	●	●	●	●	●		●				●		
bsesim(f)				●	●		●	●	●			●	●		
bsH(f)		●		●	●	●	●	●	●	●	●	●	●	●	●

Table A.6.: F feature selection in binary experiments. Rows: features, columns: experiments; green: selected. Questions (f) subset.

	F_1 + sparsity	MDS DE 1. year	MDS DE 2. year	MDS DE 1. year	MDS DE all years	MDS US 1. year	MDS US 2. year	MDS US 3. year	MDS US all years	dCor + sparsity DE	dCor + sparsity US	dCor + F_1 DE	dCor + F_1 US	EMO DE	EMO US
f1(f)															
f2(f)		■	■		■	■	■	■	■		■		■	■	■
f3(f)															
f4(f)															
f6(f)					■	■	■	■				■	■		
f7(f)															
f9(f)															
f10(f)			■	■	■			■	■	■	■	■	■	■	■
f15(f)					■	■	■		■			■	■		
f16(f)		■	■	■	■	■	■		■				■		
f17(f)															
f18(f)															
f27(f)		■				■			■				■		■
f28(f)			■			■	■		■		■		■		■
f29(f)															
f30(f)						■	■		■				■		■
f31(f)		■	■	■	■	■			■			■	■		
f32(f)						■		■	■				■		
f33(f)				■		■									
f34(f)		■	■		■			■	■			■			
f71(f)															
f72(f)						■									
f73(f)				■		■	■	■	■		■		■		■
f74(f)															
f75(f)		■				■	■	■	■				■		
f76(f)				■				■	■						
f77(f)		■	■		■	■	■	■	■			■	■		
f78(f)				■		■			■				■		
f79(f)								■	■		■		■		■
f80(f)		■	■		■					■		■		■	
f81(f)						■	■	■	■		■	■	■		■
f82(f)						■	■		■		■		■		■

Table A.7.: Lexical and linguistic feature selection in binary experiments. Rows: features, columns: experiments; green: selected. Questions (f) subset.

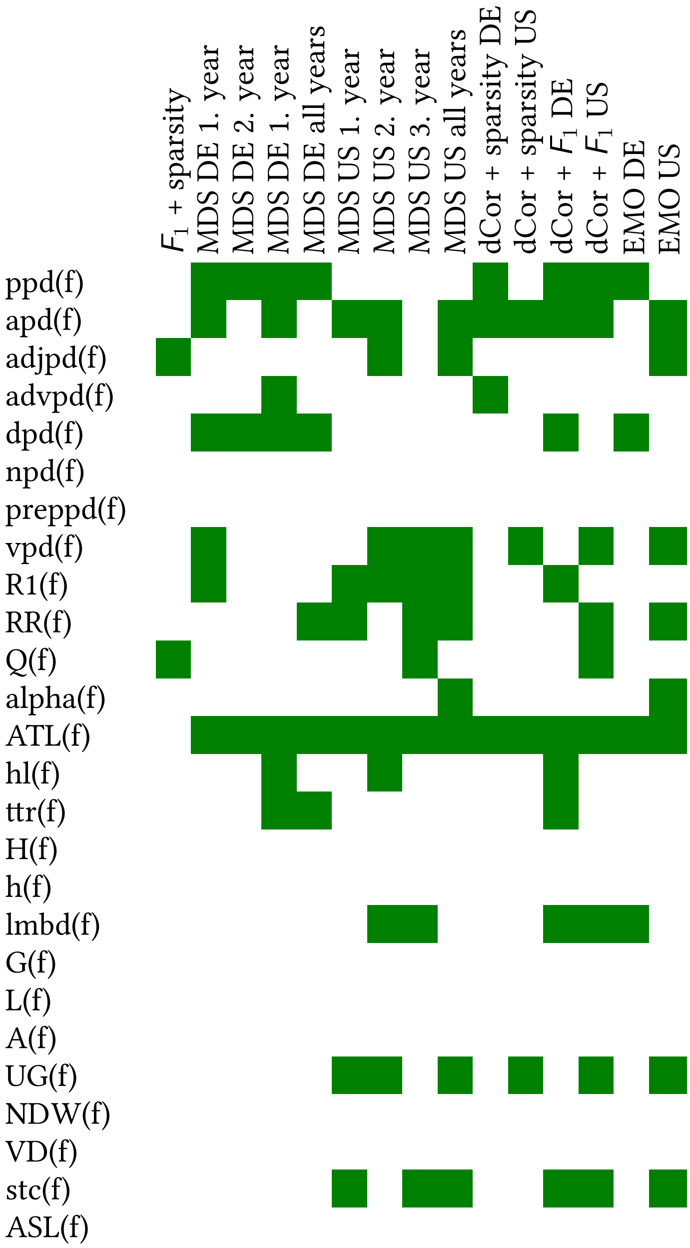


Table A.8.: Syntactic feature selection in binary experiments. Rows: features, columns: experiments; green: selected. Questions (f) subset.

	F_1 + sparsity	MDS DE 1. year	MDS DE 2. year	MDS DE 1. year	MDS DE all years	MDS US 1. year	MDS US 2. year	MDS US 3. year	MDS US all years	dCor + sparsity DE	dCor + sparsity US	dCor + F_1 DE	dCor + F_1 US	EMO DE	EMO US
LDEmu(f)		■	■	■	■	■	■	■	■	■					
depmu(f)		■	■		■	■	■	■	■			■	■		
MDDmu(f)		■	■	■	■	■	■	■	■	■	■	■	■	■	■
DDEmu(f)		■	■	■	■	■			■			■		■	
TCImu(f)		■			■	■	■	■	■	■	■	■	■	■	■
imbmu(f)		■	■	■		■	■	■	■			■	■		
Lmu(f)		■	■	■	■	■	■		■			■	■		
Wmu(f)		■			■										
wmu(f)						■	■	■	■		■		■		■
lmu(f)		■	■	■	■	■	■		■	■	■	■	■	■	■
cmu(f)		■		■	■										
LDEG(f)		■		■	■	■	■		■			■	■	■	
depG(f)		■	■	■											
MDDG(f)		■	■		■							■			
DDEG(f)	■	■	■	■	■	■	■	■	■			■	■		
TCIG(f)					■	■	■	■					■		
imbG(f)				■		■	■	■	■			■	■		
LG(f)		■	■	■	■		■					■	■		■
WG(f)		■	■	■	■									■	
wG(f)		■	■	■	■	■	■	■	■					■	
lG(f)		■	■	■	■					■		■			
cG(f)		■	■	■	■				■			■			
LDEH(f)							■		■		■		■		■
depH(f)							■		■				■		■
MDDH(f)						■	■	■	■				■		■
DDEH(f)							■		■				■		■
TCIH(f)		■	■	■	■							■		■	
imbH(f)						■	■		■				■		■
LH(f)						■	■		■		■		■		■
WH(f)							■		■		■		■		■
wH(f)						■	■	■	■		■		■		■
lH(f)						■	■	■	■		■		■		■
cH(f)						■	■		■		■		■		■

Table A.9.: BERT related feature selection in binary experiments. Rows: features, columns: experiments; green: selected. Answers (a) subset.

	F_1 + sparsity	MDS DE 1. year	MDS DE 2. year	MDS DE 1. year	MDS DE all years	MDS US 1. year	MDS US 2. year	MDS US 3. year	MDS US all years	dCor + sparsity DE	dCor + sparsity US	dCor + F_1 DE	dCor + F_1 US	EMO DE	EMO US
btH(a)															
btlH(a)						■	■	■	■				■		■
bth(a)						■	■	■	■						
btdfa(a)						■	■	■	■				■		■
btrac(a)			■			■	■	■	■			■	■		■
btradc(a)				■		■	■	■	■			■	■		
btac1(a)															
btac2(a)						■	■	■	■			■			
btac3(a)						■	■	■	■				■		
btac4(a)				■		■	■	■	■			■	■		
btadc1(a)		■	■			■	■		■			■	■		
btadc2(a)	■					■	■	■	■			■	■	■	
btadc3(a)			■			■	■	■	■			■	■		
btadc4(a)						■	■	■	■						
btadc5(a)						■	■	■	■				■		
btadc6(a)						■	■	■	■			■			
btadc7(a)								■				■	■		
btadc8(a)															
btadc9(a)				■	■		■					■	■		
btadc10(a)		■	■	■	■	■		■							■
bstsim(a)															
bsesim(a)															
bsH(a)		■					■	■	■			■	■		■

Table A.10.: F feature selection in binary experiments. Rows: features, columns: experiments; green: selected. Answers (a) subset.

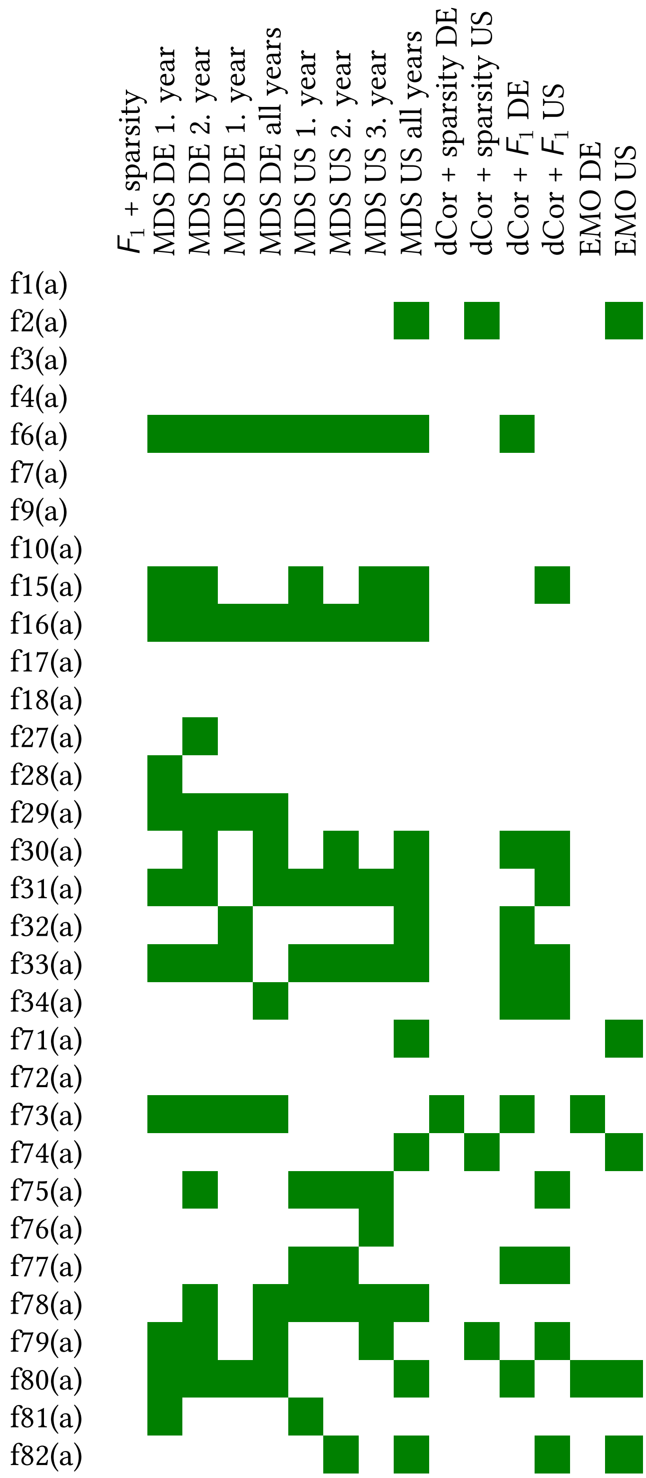


Table A.11.: Lexical and linguistic feature selection in binary experiments. Rows: features, columns: experiments; green: selected. Answers (a) subset.

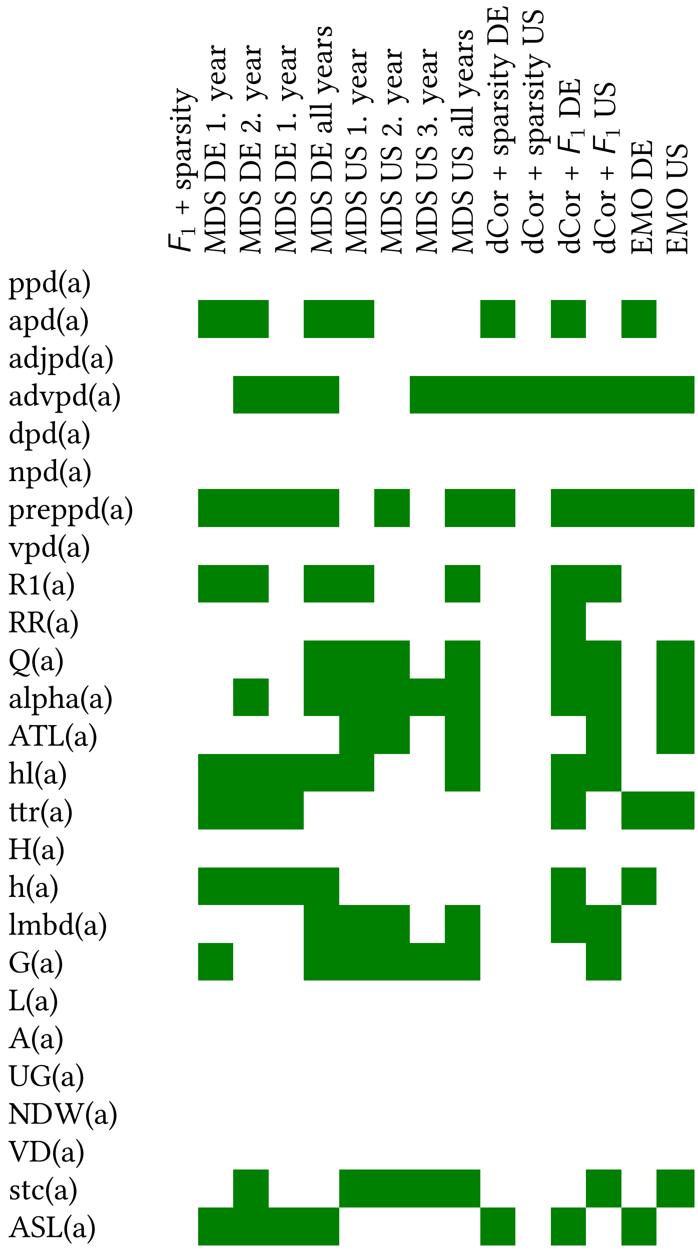


Table A.12.: Syntactic feature selection in binary experiments. Rows: features, columns: experiments; green: selected. Answers (a) subset.

	F_1 + sparsity	MDS DE 1. year	MDS DE 2. year	MDS DE 1. year	MDS DE all years	MDS US 1. year	MDS US 2. year	MDS US 3. year	MDS US all years	dCor + sparsity DE	dCor + sparsity US	dCor + F_1 DE	dCor + F_1 US	EMO DE	EMO US
LDEmu(a)												■			
depmu(a)					■										
MDDmu(a)							■								
DDEmu(a)		■	■		■										
TCImu(a)		■	■	■	■					■		■		■	
imbmu(a)			■	■	■							■	■		
Lmu(a)												■	■	■	
Wmu(a)		■				■	■		■			■	■		
wmu(a)															
lmu(a)						■	■		■				■		
cmu(a)		■			■	■	■	■	■			■			
LDEG(a)			■	■								■			
depG(a)		■	■	■	■	■		■							
MDDG(a)						■	■		■			■	■		
DDEG(a)			■	■								■			
TCIG(a)		■		■	■	■	■		■			■	■		■
imbG(a)		■	■	■	■	■	■	■	■	■		■	■	■	
LG(a)		■				■	■		■						
WG(a)						■	■		■						
wG(a)															
lG(a)			■										■		
cG(a)			■				■	■	■				■		
LDEH(a)						■			■		■		■		■
depH(a)						■	■		■				■		■
MDDH(a)	■	■				■	■		■				■		■
DDEH(a)						■	■		■				■		■
TCIH(a)						■	■	■	■		■		■		■
imbH(a)						■	■		■						■
LH(a)						■	■		■						■
WH(a)						■	■		■				■		■
wH(a)			■			■	■	■	■	■	■		■		■
lH(a)						■	■	■	■		■		■		■
cH(a)						■	■		■				■		■

Table A.13.: BERT related feature selection in real-valued experiments. Rows: features, columns: experiments. Questions and answers (fa) subset.

	Sobol RFC	Sobol dCor DE	Sobol dCor US	Morris RFC	Morris dCor DE	Morris dCor US	EMO DE	EMO US
btH(fa)	0.0078	0.0013	0.0027	0.0095	0.0005	0.0006	0.0962	0.0002
btlH(fa)	0.0040	0.0002	0.0000	0.0073	0.0001	0.0001	0.1151	0.0001
bth(fa)	0.0283	0.0000	0.0066	0.0339	0.0001	0.0010	0.0010	0.0748
btdfa(fa)	0.1055	0.0002	0.0000	0.1365	0.0002	0.0000	0.0038	0.0010
btrac(fa)	0.0000	0.0007	0.0039	0.0018	0.0004	0.0008	0.0533	0.0361
btradc(fa)	0.0047	0.0018	0.0004	0.0033	0.0006	0.0002	0.2033	0.0006
btac1(fa)	0.1722	0.0000	0.0002	0.2225	0.0000	0.0002	0.1775	0.0006
btac2(fa)	0.0120	0.0000	0.0000	0.0144	0.0001	0.0001	0.0878	0.0003
btac3(fa)	0.0051	0.0110	0.0074	0.0081	0.0015	0.0011	0.2392	0.0553
btac4(fa)	0.0029	0.0026	0.0000	0.0026	0.0007	0.0001	0.1627	0.0010
btadc1(fa)	0.0087	0.0001	0.0011	0.0121	0.0001	0.0004	0.0677	0.0021
btadc2(fa)	0.0011	0.0018	0.0001	0.0014	0.0006	0.0001	0.0352	0.0007
btadc3(fa)	0.0056	0.0009	0.0007	0.0095	0.0004	0.0003	0.0164	0.0002
btadc4(fa)	0.0000	0.0034	0.0003	0.0012	0.0008	0.0002	0.0261	0.0003
btadc5(fa)	0.0002	0.0004	0.0012	0.0018	0.0003	0.0004	0.1812	0.0006
btadc6(fa)	0.0018	0.0027	0.0026	0.0021	0.0007	0.0006	0.0412	0.0000
btadc7(fa)	0.0009	0.0030	0.0005	0.0018	0.0008	0.0002	0.0989	0.0002
btadc8(fa)	0.0016	0.0004	0.0002	0.0012	0.0003	0.0001	0.0292	0.0002
btadc9(fa)	0.0011	0.0003	0.0003	0.0012	0.0003	0.0002	0.0080	0.0005
btadc10(fa)	0.0045	0.0020	0.0006	0.0069	0.0006	0.0003	0.0951	0.0001
bstsim(fa)	0.0779	0.0000	0.0003	0.0646	0.0000	0.0002	0.1690	0.0000
bsesim(fa)	0.5285	0.0000	0.0000	0.6981	0.0000	0.0000	0.0160	0.0002
bsH(fa)	0.1499	0.0000	0.0033	0.2165	0.0001	0.0007	0.1137	0.0002

Table A.14.: F feature selection in real-valued experiments. Rows: features, columns: experiments. Questions and answers (fa) subset. Part 1.

	Sobol RFC	Sobol dCor DE	Sobol dCor US	Morris RFC	Morris dCor DE	Morris dCor US	EMO DE	EMO US
f1(fa)	0.0089	0.0015	0.0020	0.0092	0.0005	0.0005	0.0000	0.0000
f2(fa)	0.1060	0.0000	0.0019	0.0816	0.0001	0.0005	0.0540	0.0001
f3(fa)	0.0029	0.0020	0.0027	0.0048	0.0006	0.0006	0.0003	0.0000
f4(fa)	0.0025	0.0017	0.0001	0.0021	0.0005	0.0001	0.0000	0.0000
f6(fa)	0.0700	0.0006	0.0004	0.0633	0.0003	0.0002	0.0265	0.0003
f7(fa)	0.0172	0.0002	0.0138	0.0135	0.0002	0.0014	0.0372	0.0000
f9(fa)	0.0134	0.0001	0.0014	0.0163	0.0001	0.0005	0.0008	0.0000
f10(fa)	0.0000	0.0002	0.0008	0.0004	0.0002	0.0003	0.0480	0.0001
f15(fa)	0.0192	0.0002	0.0006	0.0193	0.0002	0.0003	0.1729	0.0011
f16(fa)	0.0018	0.0001	0.0001	0.0032	0.0001	0.0001	0.0909	0.0004
f17(fa)	0.0033	0.0012	0.0002	0.0037	0.0004	0.0002	0.0003	0.0000
f18(fa)	0.0482	0.0002	0.0039	0.0459	0.0002	0.0008	0.2316	0.0000
f27(fa)	0.0549	0.0000	0.0075	0.0531	0.0000	0.0010	0.0580	0.0001
f28(fa)	0.0462	0.0002	0.0078	0.0489	0.0002	0.0011	0.7155	0.0002
f29(fa)	0.0303	0.0007	0.0016	0.0330	0.0004	0.0005	0.0251	0.0003
f30(fa)	0.0040	0.0000	0.0003	0.0109	0.0001	0.0002	0.1218	0.0019
f31(fa)	0.0277	0.0032	0.0006	0.0381	0.0008	0.0003	0.2023	0.0005
f32(fa)	0.0089	0.0003	0.0001	0.0087	0.0002	0.0001	0.1122	0.0015
f33(fa)	0.0047	0.0041	0.0016	0.0036	0.0009	0.0005	0.0130	0.0007
f34(fa)	0.0016	0.0001	0.0005	0.0042	0.0001	0.0002	0.0578	0.0023
f71(fa)	0.0016	0.0408	0.0014	0.0010	0.0029	0.0005	0.6142	0.0000
f72(fa)	0.0011	0.0002	0.0015	0.0018	0.0001	0.0005	0.1455	0.0001
f73(fa)	0.0002	0.0176	0.0022	0.0022	0.0019	0.0006	0.5322	0.0001
f74(fa)	0.0011	0.0023	0.0005	0.0004	0.0007	0.0003	0.2313	0.0000

Table A.15.: F feature selection in real-valued experiments. Rows: features, columns: experiments. Questions and answers (fa) subset. Part 2.

	Sobol RFC	Sobol dCor DE	Sobol dCor US	Morris RFC	Morris dCor DE	Morris dCor US	EMO DE	EMO US
f75(fa)	0.0007	0.0802	0.0049	0.0024	0.0040	0.0009	0.1351	0.0002
f76(fa)	0.0011	0.0007	0.0008	0.0016	0.0004	0.0003	0.0309	0.0002
f77(fa)	0.0040	0.0476	0.0001	0.0036	0.0031	0.0001	0.0843	0.0008
f78(fa)	0.0011	0.0011	0.0010	0.0015	0.0005	0.0004	0.1161	0.0019
f79(fa)	0.0192	0.0029	0.0013	0.0187	0.0008	0.0004	0.0028	0.0066
f80(fa)	0.0076	0.0007	0.0018	0.0080	0.0004	0.0005	0.0499	0.0024
f81(fa)	0.0011	0.0064	0.0080	0.0033	0.0011	0.0011	0.0755	0.0002
f82(fa)	0.0022	0.0003	0.0086	0.0059	0.0002	0.0012	0.0059	0.0001

Table A.16.: Lexical and linguistic feature selection in real-valued experiments. Rows: features, columns: experiments. Questions and answers (fa) subset.

	Sobol RFC	Sobol dCor DE	Sobol dCor US	Morris RFC	Morris dCor DE	Morris dCor US	EMO DE	EMO US
ppd(fa)	0.0007	0.0006	0.0001	0.0004	0.0003	0.0001	0.1235	0.0000
apd(fa)	0.0013	0.0004	0.0010	0.0015	0.0003	0.0003	0.1110	0.0002
adjpd(fa)	0.0013	0.0000	0.0024	0.0016	0.0001	0.0006	0.0002	0.0001
advpd(fa)	0.0000	0.0000	0.0005	0.0000	0.0001	0.0002	0.0214	0.0000
dpd(fa)	0.0029	0.0006	0.0001	0.0024	0.0003	0.0001	0.0014	0.0003
npd(fa)	0.0060	0.0003	0.0007	0.0076	0.0002	0.0003	0.0003	0.0000
preppd(fa)	0.0020	0.0002	0.0100	0.0053	0.0002	0.0012	0.0153	0.0013
vpd(fa)	0.0060	0.0004	0.0046	0.0106	0.0003	0.0008	0.0185	0.0001
R1(fa)	0.0045	0.0006	0.0001	0.0054	0.0003	0.0001	0.0051	0.0006
RR(fa)	0.0004	0.0004	0.0017	0.0029	0.0003	0.0005	0.1343	0.0010
Q(fa)	0.0031	0.0265	0.0001	0.0037	0.0023	0.0001	0.1117	0.0002
alpha(fa)	0.0022	0.0125	0.0014	0.0037	0.0016	0.0005	0.2879	0.0021
ATL(fa)	0.2568	0.0002	0.0000	0.3507	0.0002	0.0000	0.1180	0.0003
hl(fa)	0.0060	0.0001	0.0012	0.0076	0.0001	0.0004	0.2299	0.0003
ttr(fa)	0.0031	0.0002	0.0017	0.0076	0.0001	0.0005	0.0068	0.0016
H(fa)	0.0025	0.0014	0.0026	0.0016	0.0005	0.0006	0.0007	0.0002
h(fa)	0.0011	0.0001	0.0037	0.0016	0.0001	0.0007	0.0576	0.0003
lmbd(fa)	0.0027	0.0001	0.0002	0.0034	0.0001	0.0002	0.1991	0.0005
G(fa)	0.0022	0.0003	0.0025	0.0039	0.0002	0.0006	0.0003	0.0001
L(fa)	0.0018	0.0017	0.0003	0.0018	0.0006	0.0002	0.0017	0.0000
A(fa)	0.0038	0.0005	0.0005	0.0036	0.0003	0.0003	0.1309	0.0001
UG(fa)	0.1022	0.0001	0.0001	0.1137	0.0001	0.0001	0.1090	0.0000
NDW(fa)	0.0738	0.0002	0.0000	0.0810	0.0002	0.0000	0.0004	0.0001
VD(fa)	0.0067	0.0009	0.0001	0.0076	0.0004	0.0001	0.0020	0.0001
stc(fa)	0.0009	0.0000	0.0662	0.0008	0.0001	0.0031	0.1593	0.0668
ASL(fa)	0.0638	0.0000	0.0102	0.0666	0.0000	0.0012	0.0093	0.0000

Table A.17.: Syntactic feature selection in real-valued experiments. Rows: features, columns: experiments. Questions and answers (fa) subset. Part 1.

	Sobol RFC	Sobol dCor DE	Sobol dCor US	Morris RFC	Morris dCor DE	Morris dCor US	EMO DE	EMO US
LDEmu(fa)	0.0031	0.0035	0.0012	0.0053	0.0008	0.0004	0.0555	0.0009
depmu(fa)	0.0040	0.0022	0.0016	0.0048	0.0006	0.0005	0.0797	0.0013
MDDmu(fa)	0.0058	0.0039	0.0007	0.0050	0.0009	0.0003	0.0502	0.0001
DDEmu(fa)	0.0002	0.0024	0.0040	0.0003	0.0007	0.0008	0.0389	0.0002
TCImu(fa)	0.0007	0.0001	0.0014	0.0010	0.0001	0.0004	0.3047	0.0000
imbmu(fa)	0.0025	0.0006	0.0028	0.0034	0.0003	0.0006	0.0006	0.0025
Lmu(fa)	0.0074	0.0002	0.0001	0.0135	0.0002	0.0001	0.1060	0.0005
Wmu(fa)	0.0016	0.0101	0.0004	0.0040	0.0014	0.0002	0.2617	0.0008
wmu(fa)	0.0087	0.0014	0.0007	0.0087	0.0005	0.0003	0.1172	0.0000
lmu(fa)	0.0038	0.0000	0.0014	0.0039	0.0000	0.0005	0.0909	0.0048
cmu(fa)	0.0040	0.0082	0.0002	0.0040	0.0013	0.0002	0.0013	0.0012
LDEG(fa)	0.0020	0.0041	0.0008	0.0032	0.0009	0.0003	0.0458	0.0004
depG(fa)	0.0042	0.0000	0.0045	0.0060	0.0001	0.0008	0.1833	0.0005
MDDG(fa)	0.0027	0.0054	0.0000	0.0026	0.0010	0.0001	0.0725	0.0009
DDEG(fa)	0.0018	0.0041	0.0004	0.0037	0.0009	0.0002	0.0078	0.0002
TCIG(fa)	0.0013	0.0000	0.0002	0.0012	0.0001	0.0002	0.0683	0.0000
imbG(fa)	0.0025	0.0001	0.0002	0.0039	0.0001	0.0002	0.2168	0.0007
LG(fa)	0.0007	0.0034	0.0000	0.0014	0.0008	0.0000	0.1191	0.0001
WG(fa)	0.0004	0.0019	0.0001	0.0010	0.0006	0.0001	0.0860	0.0001
wG(fa)	0.0018	0.0014	0.0000	0.0033	0.0005	0.0001	0.1660	0.0010
lG(fa)	0.0000	0.0003	0.0001	0.0008	0.0002	0.0001	0.0324	0.0010
cG(fa)	0.0040	0.0012	0.0001	0.0040	0.0004	0.0001	0.0000	0.0001

Table A.18.: Syntactic feature selection in real-valued experiments. Rows: features, columns: experiments. Questions and answers (fa) subset. Part 2.

	Sobol RFC	Sobol dCor DE	Sobol dCor US	Morris RFC	Morris dCor DE	Morris dCor US	EMO DE	EMO US
LDEH(fa)	0.0303	0.0014	0.0088	0.0496	0.0005	0.0011	0.1310	0.0002
depH(fa)	0.0132	0.0015	0.0088	0.0195	0.0005	0.0011	0.0000	0.0006
MDDH(fa)	0.0033	0.0000	0.0017	0.0048	0.0001	0.0005	0.2216	0.0000
DDEH(fa)	0.0132	0.0015	0.0088	0.0186	0.0005	0.0011	0.0428	0.0000
TCIH(fa)	0.0183	0.0000	0.0024	0.0262	0.0000	0.0006	0.0009	0.0099
imbH(fa)	0.0060	0.0016	0.0085	0.0102	0.0006	0.0011	0.1171	0.0002
LH(fa)	0.0154	0.0015	0.0087	0.0237	0.0005	0.0011	0.0319	0.0001
WH(fa)	0.0127	0.0016	0.0084	0.0201	0.0005	0.0011	0.1137	0.0008
wH(fa)	0.0076	0.0010	0.0007	0.0099	0.0005	0.0003	0.0760	0.0006
lH(fa)	0.0127	0.0000	0.0001	0.0187	0.0000	0.0001	0.0429	0.0000
cH(fa)	0.0125	0.0016	0.0084	0.0177	0.0005	0.0011	0.1393	0.0003

Table A.19.: BERT related feature selection in real-valued experiments. Rows: features, columns: experiments. Questions (f) subset.

	Sobol RFC	Sobol dCor DE	Sobol dCor US	Morris RFC	Morris dCor DE	Morris dCor US	EMO DE	EMO US
btH(f)	0.0025	0.0114	0.0009	0.0039	0.0015	0.0003	0.0011	0.0003
btlH(f)	0.0259	0.0040	0.0001	0.0370	0.0009	0.0001	0.0597	0.0001
bth(f)	0.0038	0.0000	0.0040	0.0066	0.0001	0.0008	0.0390	0.0001
btdfa(f)	0.0176	0.0001	0.0005	0.0255	0.0001	0.0003	0.0333	0.0227
btrac(f)	0.0009	0.0017	0.0002	0.0022	0.0006	0.0002	0.0556	0.0002
btradc(f)	0.0000	0.0002	0.0012	0.0006	0.0002	0.0004	0.1588	0.0002
btac1(f)	0.0306	0.0002	0.0000	0.0417	0.0002	0.0001	0.0929	0.0008
btac2(f)	0.0031	0.0001	0.0134	0.0032	0.0001	0.0014	0.0068	0.0501
btac3(f)	0.0007	0.0002	0.0008	0.0016	0.0002	0.0004	0.1226	0.0014
btac4(f)	0.0022	0.0001	0.0008	0.0016	0.0002	0.0004	0.0003	0.0001
btadc1(f)	0.0045	0.0018	0.0000	0.0054	0.0006	0.0001	0.0191	0.0000
btadc2(f)	0.0040	0.0002	0.0024	0.0045	0.0002	0.0006	0.0586	0.0001
btadc3(f)	0.0038	0.0000	0.0001	0.0050	0.0001	0.0001	0.0975	0.0008
btadc4(f)	0.0025	0.0000	0.0000	0.0033	0.0001	0.0001	0.1306	0.0010
btadc5(f)	0.0016	0.0017	0.0008	0.0018	0.0006	0.0003	0.1195	0.0003
btadc6(f)	0.0009	0.0022	0.0079	0.0008	0.0006	0.0011	0.0018	0.0009
btadc7(f)	0.0018	0.0009	0.0050	0.0009	0.0004	0.0009	0.1039	0.0009
btadc8(f)	0.0016	0.0002	0.0006	0.0018	0.0002	0.0003	0.0006	0.0006
btadc9(f)	0.0031	0.0003	0.0017	0.0056	0.0002	0.0005	0.0304	0.0013
btadc10(f)	0.0022	0.0008	0.0004	0.0018	0.0004	0.0002	0.1004	0.0005
bstsim(f)	0.0025	0.0038	0.0052	0.0042	0.0008	0.0009	0.0021	0.0003
bsesim(f)	0.0192	0.0080	0.0008	0.0137	0.0012	0.0003	0.0916	0.0003
bsH(f)	0.0013	0.0023	0.0306	0.0016	0.0007	0.0021	0.0674	0.0108

Table A.20.: F feature selection in real-valued experiments. Rows: features, columns: experiments. Questions (f) subset. Part 1.

	Sobol RFC	Sobol dCor DE	Sobol dCor US	Morris RFC	Morris dCor DE	Morris dCor US	EMO DE	EMO US
f1(f)	0.0058	0.0098	0.0003	0.0090	0.0014	0.0002	0.0004	0.0000
f2(f)	0.0011	0.0042	0.0287	0.0008	0.0009	0.0021	0.0121	0.0082
f3(f)	0.0036	0.0086	0.0012	0.0033	0.0013	0.0004	0.0001	0.0000
f4(f)	0.0009	0.0093	0.0001	0.0018	0.0013	0.0001	0.0008	0.0000
f6(f)	0.0004	0.0000	0.0004	0.0018	0.0000	0.0002	0.1510	0.0012
f7(f)	0.0022	0.0007	0.0001	0.0030	0.0003	0.0001	0.0018	0.0000
f9(f)	0.0016	0.0032	0.0003	0.0001	0.0008	0.0002	0.0181	0.0000
f10(f)	0.0004	0.0014	0.0021	0.0022	0.0005	0.0006	0.3087	0.0001
f15(f)	0.0071	0.0000	0.0002	0.0103	0.0000	0.0002	0.0287	0.0007
f16(f)	0.0033	0.0035	0.0003	0.0026	0.0008	0.0002	0.0374	0.0003
f17(f)	0.0118	0.0008	0.0021	0.0159	0.0004	0.0005	0.0157	0.0000
f18(f)	0.1167	0.0024	0.0000	0.1366	0.0007	0.0001	0.1820	0.0000
f27(f)	0.0074	0.0001	0.0042	0.0086	0.0002	0.0008	0.0516	0.0060
f28(f)	0.0009	0.0000	0.0003	0.0027	0.0001	0.0002	0.0143	0.0007
f29(f)	0.0000	0.0003	0.0005	0.0001	0.0002	0.0003	0.0827	0.0002
f30(f)	0.0020	0.0000	0.0002	0.0018	0.0000	0.0002	0.1860	0.0000
f31(f)	0.0020	0.0002	0.0005	0.0039	0.0002	0.0003	0.0057	0.0003
f32(f)	0.0018	0.0000	0.0003	0.0034	0.0000	0.0002	0.0144	0.0004
f33(f)	0.0054	0.0001	0.0001	0.0060	0.0001	0.0001	0.0561	0.0009
f34(f)	0.0011	0.0007	0.0006	0.0009	0.0004	0.0003	0.1159	0.0007
f71(f)	0.0000	0.0004	0.0015	0.0000	0.0003	0.0004	0.0244	0.0000
f72(f)	0.0013	0.0053	0.0014	0.0006	0.0010	0.0005	0.0663	0.0001
f73(f)	0.0020	0.0002	0.0062	0.0024	0.0002	0.0010	0.0544	0.0001
f74(f)	0.0002	0.0000	0.0011	0.0001	0.0001	0.0004	0.1559	0.0001

Table A.21.: F feature selection in real-valued experiments. Rows: features, columns: experiments. Questions (f) subset. Part 2.

	Sobol RFC	Sobol dCor DE	Sobol dCor US	Morris RFC	Morris dCor DE	Morris dCor US	EMO DE	EMO US
f75(f)	0.0000	0.0000	0.0003	0.0009	0.0000	0.0002	0.0886	0.0014
f76(f)	0.0020	0.0005	0.0001	0.0019	0.0003	0.0001	0.0141	0.0009
f77(f)	0.0054	0.0005	0.0057	0.0045	0.0003	0.0009	0.1548	0.0008
f78(f)	0.0029	0.0007	0.0000	0.0027	0.0003	0.0001	0.0230	0.0009
f79(f)	0.0025	0.0001	0.0014	0.0046	0.0001	0.0005	0.1342	0.0003
f80(f)	0.0000	0.0003	0.0001	0.0010	0.0003	0.0001	0.0095	0.0001
f81(f)	0.0007	0.0000	0.0101	0.0018	0.0001	0.0012	0.0006	0.0002
f82(f)	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.1516	0.0000

Table A.22.: Lexical and linguistic feature selection in real-valued experiments. Rows: features, columns: experiments. Questions (f) subset.

	Sobol RFC	Sobol dCor DE	Sobol dCor US	Morris RFC	Morris dCor DE	Morris dCor US	EMO DE	EMO US
ppd(f)	0.0011	0.0013	0.0000	0.0015	0.0005	0.0000	0.0737	0.0006
apd(f)	0.0004	0.0000	0.0024	0.0010	0.0001	0.0006	0.1438	0.0000
adjpd(f)	0.0020	0.0015	0.0055	0.0039	0.0005	0.0009	0.0198	0.0001
advpd(f)	0.0007	0.0004	0.0001	0.0008	0.0003	0.0001	0.0695	0.0000
dpd(f)	0.0000	0.0013	0.0000	0.0003	0.0005	0.0000	0.1339	0.0000
npd(f)	0.0011	0.0085	0.0024	0.0021	0.0013	0.0006	0.0145	0.0000
preppd(f)	0.0029	0.0067	0.0018	0.0014	0.0011	0.0005	0.0576	0.0000
vpd(f)	0.0013	0.0009	0.0043	0.0015	0.0004	0.0008	0.0023	0.0001
R1(f)	0.0042	0.0000	0.0004	0.0115	0.0000	0.0002	0.1183	0.0012
RR(f)	0.0002	0.0001	0.0014	0.0010	0.0001	0.0004	0.0575	0.0004
Q(f)	0.0007	0.0012	0.0000	0.0014	0.0005	0.0001	0.2796	0.0000
alpha(f)	0.0013	0.0019	0.0000	0.0026	0.0006	0.0001	0.0227	0.0002
ATL(f)	0.1700	0.0011	0.0000	0.1840	0.0005	0.0001	0.2105	0.0063
hl(f)	0.0176	0.0008	0.0037	0.0240	0.0004	0.0007	0.0567	0.0003
ttr(f)	0.0263	0.0006	0.0065	0.0414	0.0003	0.0010	0.0115	0.0004
H(f)	0.0022	0.0078	0.0002	0.0024	0.0012	0.0002	0.0076	0.0001
h(f)	0.0002	0.0032	0.0061	0.0003	0.0008	0.0009	0.2063	0.0003
lmbd(f)	0.0009	0.0058	0.0002	0.0004	0.0011	0.0002	0.1297	0.0005
G(f)	0.0149	0.0008	0.0079	0.0242	0.0004	0.0011	0.0304	0.0003
L(f)	0.0022	0.0101	0.0001	0.0018	0.0014	0.0001	0.0002	0.0000
A(f)	0.0049	0.0007	0.0002	0.0054	0.0004	0.0002	0.0854	0.0002
UG(f)	0.0812	0.0000	0.0002	0.0879	0.0001	0.0002	0.0550	0.0001
NDW(f)	0.0502	0.0004	0.0001	0.0454	0.0002	0.0001	0.3425	0.0001
VD(f)	0.0009	0.0126	0.0011	0.0010	0.0016	0.0004	0.2201	0.0001
stc(f)	0.0042	0.0004	0.0034	0.0051	0.0003	0.0007	0.1252	0.0096
ASL(f)	0.0020	0.0000	0.0003	0.0009	0.0000	0.0002	0.0678	0.0000

Table A.23.: Syntactic feature selection in real-valued experiments. Rows: features, columns: experiments. Questions (f) subset. Part 1.

	Sobol RFC	Sobol dCor DE	Sobol dCor US	Morris RFC	Morris dCor DE	Morris dCor US	EMO DE	EMO US
LDEmu(f)	0.0029	0.0018	0.0000	0.0054	0.0006	0.0000	0.1742	0.0018
depmu(f)	0.0100	0.0006	0.0032	0.0105	0.0003	0.0007	0.0779	0.0383
MDDmu(f)	0.0036	0.0114	0.0016	0.0042	0.0015	0.0005	0.2678	0.0000
DDEmu(f)	0.0007	0.0003	0.0002	0.0014	0.0002	0.0001	0.1318	0.0006
TCImu(f)	0.0025	0.0001	0.0280	0.0037	0.0002	0.0021	0.0053	0.0049
imbmu(f)	0.0004	0.0045	0.1318	0.0009	0.0010	0.0045	0.0754	0.0002
Lmu(f)	0.0036	0.0002	0.0001	0.0037	0.0002	0.0001	0.0708	0.0003
Wmu(f)	0.0031	0.0008	0.0007	0.0032	0.0004	0.0003	0.0283	0.0002
wmu(f)	0.0065	0.0002	0.0029	0.0129	0.0002	0.0007	0.0317	0.0001
lmu(f)	0.0004	0.0027	0.0046	0.0008	0.0007	0.0009	0.9966	0.0045
cmu(f)	0.0042	0.0006	0.0009	0.0060	0.0003	0.0003	0.1475	0.0013
LDEG(f)	0.0000	0.0030	0.0002	0.0010	0.0008	0.0002	0.0604	0.0015
depG(f)	0.0000	0.0001	0.0001	0.0004	0.0002	0.0001	0.0027	0.0003
MDDG(f)	0.0009	0.0003	0.0000	0.0009	0.0002	0.0001	0.0055	0.0001
DDEG(f)	0.0000	0.0012	0.0002	0.0014	0.0005	0.0002	0.0213	0.0006
TCIG(f)	0.0000	0.0069	0.0107	0.0000	0.0011	0.0013	0.0800	0.0001
imbG(f)	0.0000	0.0043	0.0078	0.0000	0.0009	0.0011	0.0748	0.0000
LG(f)	0.0004	0.0008	0.0001	0.0018	0.0003	0.0001	0.0645	0.0001
WG(f)	0.0002	0.0002	0.0001	0.0008	0.0002	0.0001	0.1343	0.0003
wG(f)	0.0004	0.0012	0.0002	0.0004	0.0004	0.0002	0.0872	0.0004
lG(f)	0.0000	0.0025	0.0010	0.0000	0.0007	0.0004	0.1537	0.0001
cG(f)	0.0004	0.0003	0.0001	0.0006	0.0002	0.0001	0.0156	0.0004

Table A.24.: Syntactic feature selection in real-valued experiments. Rows: features, columns: experiments. Questions (f) subset. Part 2.

	Sobol RFC	Sobol dCor DE	Sobol dCor US	Morris RFC	Morris dCor DE	Morris dCor US	EMO DE	EMO US
LDEH(f)	0.0016	0.0092	0.0027	0.0019	0.0013	0.0006	0.0291	0.0003
depH(f)	0.0013	0.0089	0.0027	0.0024	0.0013	0.0006	0.0210	0.0001
MDDH(f)	0.0020	0.0069	0.0058	0.0027	0.0012	0.0009	0.0528	0.0002
DDEH(f)	0.0016	0.0090	0.0025	0.0024	0.0013	0.0006	0.0596	0.0001
TCIH(f)	0.0016	0.0010	0.0002	0.0026	0.0004	0.0002	0.0140	0.0000
imbH(f)	0.0007	0.0087	0.0026	0.0027	0.0013	0.0006	0.0428	0.0002
LH(f)	0.0027	0.0087	0.0026	0.0029	0.0013	0.0006	0.1162	0.0003
WH(f)	0.0009	0.0088	0.0027	0.0008	0.0013	0.0006	0.0116	0.0003
wH(f)	0.0011	0.0033	0.0119	0.0019	0.0008	0.0013	0.0337	0.0000
lH(f)	0.0011	0.0071	0.0228	0.0009	0.0012	0.0019	0.0536	0.0185
cH(f)	0.0002	0.0088	0.0026	0.0001	0.0013	0.0006	0.0372	0.0003

Table A.25.: BERT related feature selection in real-valued experiments. Rows: features, columns: experiments. Answers (a) subset.

	Sobol RFC	Sobol dCor DE	Sobol dCor US	Morris RFC	Morris dCor DE	Morris dCor US	EMO DE	EMO US
btH(a)	0.0049	0.0030	0.0039	0.0053	0.0008	0.0008	0.0078	0.0001
btlH(a)	0.0007	0.0008	0.0004	0.0015	0.0004	0.0002	0.0032	0.0087
bth(a)	0.0049	0.0017	0.0006	0.0069	0.0006	0.0003	0.0013	0.0002
btdfa(a)	0.0571	0.0018	0.0001	0.0504	0.0006	0.0001	0.0048	0.0005
btrac(a)	0.0011	0.0000	0.0082	0.0022	0.0000	0.0011	0.1089	0.0434
btradc(a)	0.0022	0.0000	0.0007	0.0018	0.0001	0.0003	0.0005	0.0012
btac1(a)	0.0787	0.0001	0.0003	0.0870	0.0001	0.0002	0.1332	0.0009
btac2(a)	0.0294	0.0046	0.0000	0.0279	0.0009	0.0000	0.1604	0.0001
btac3(a)	0.0058	0.0045	0.0011	0.0062	0.0009	0.0004	0.0474	0.0004
btac4(a)	0.0060	0.0001	0.0002	0.0097	0.0001	0.0002	0.0853	0.0002
btadc1(a)	0.0016	0.0000	0.0002	0.0034	0.0001	0.0002	0.0224	0.0000
btadc2(a)	0.0007	0.0012	0.0008	0.0014	0.0005	0.0003	0.0053	0.0000
btadc3(a)	0.0007	0.0003	0.0003	0.0018	0.0002	0.0002	0.1209	0.0005
btadc4(a)	0.0040	0.0017	0.0025	0.0050	0.0006	0.0006	0.0998	0.0010
btadc5(a)	0.0016	0.0010	0.0054	0.0014	0.0004	0.0009	0.0083	0.0007
btadc6(a)	0.0011	0.0030	0.0005	0.0016	0.0007	0.0003	0.0789	0.0002
btadc7(a)	0.0025	0.0003	0.0000	0.0024	0.0002	0.0001	0.0296	0.0007
btadc8(a)	0.0027	0.0014	0.0001	0.0019	0.0005	0.0001	0.0438	0.0010
btadc9(a)	0.0020	0.0030	0.0001	0.0029	0.0007	0.0001	0.0002	0.0000
btadc10(a)	0.0060	0.0227	0.0003	0.0076	0.0021	0.0002	0.0119	0.0004
bstsim(a)	0.0165	0.0016	0.0017	0.0118	0.0006	0.0005	0.0942	0.0001
bsesim(a)	0.2483	0.0001	0.0000	0.2371	0.0001	0.0000	0.2163	0.0002
bsH(a)	0.1686	0.0001	0.0001	0.2414	0.0001	0.0001	0.0362	0.0004

Table A.26.: F feature selection in real-valued experiments. Rows: features, columns: experiments. Answers (a) subset. Part 1.

	Sobol RFC	Sobol dCor DE	Sobol dCor US	Morris RFC	Morris dCor DE	Morris dCor US	EMO DE	EMO US
f1(a)	0.0018	0.0075	0.0057	0.0019	0.0012	0.0009	0.0002	0.0000
f2(a)	0.1624	0.0001	0.0000	0.2187	0.0001	0.0001	0.0000	0.0003
f3(a)	0.0027	0.0069	0.0092	0.0040	0.0012	0.0012	0.0066	0.0000
f4(a)	0.0018	0.0009	0.0049	0.0034	0.0004	0.0008	0.0006	0.0000
f6(a)	0.0272	0.1013	0.0001	0.0435	0.0045	0.0001	0.0736	0.0000
f7(a)	0.0158	0.0004	0.0138	0.0189	0.0003	0.0014	0.0022	0.0000
f9(a)	0.0011	0.0031	0.0000	0.0016	0.0007	0.0000	0.0405	0.0013
f10(a)	0.0002	0.0001	0.0030	0.0009	0.0001	0.0007	0.1567	0.0000
f15(a)	0.0047	0.0004	0.0003	0.0062	0.0002	0.0002	0.1665	0.0006
f16(a)	0.0009	0.0163	0.0000	0.0012	0.0018	0.0000	0.1089	0.0000
f17(a)	0.0038	0.0002	0.0003	0.0069	0.0002	0.0002	0.0001	0.0000
f18(a)	0.0540	0.0002	0.0075	0.0804	0.0002	0.0011	0.0004	0.0000
f27(a)	0.0489	0.0000	0.0155	0.0693	0.0001	0.0015	0.0279	0.0000
f28(a)	0.0120	0.0001	0.0041	0.0209	0.0001	0.0008	0.0056	0.0001
f29(a)	0.0228	0.0002	0.0066	0.0318	0.0002	0.0010	0.1527	0.0006
f30(a)	0.0000	0.0001	0.0003	0.0001	0.0001	0.0002	0.0040	0.0022
f31(a)	0.0138	0.0000	0.0002	0.0228	0.0001	0.0002	0.1235	0.0009
f32(a)	0.0174	0.0047	0.0001	0.0266	0.0010	0.0001	0.0372	0.0003
f33(a)	0.0033	0.0000	0.0055	0.0048	0.0001	0.0009	0.1113	0.0005
f34(a)	0.0000	0.0002	0.0014	0.0001	0.0002	0.0004	0.0340	0.0006
f71(a)	0.0000	0.0000	0.0026	0.0010	0.0001	0.0006	0.1581	0.0000
f72(a)	0.0000	0.0006	0.0001	0.0004	0.0003	0.0001	0.2699	0.0000
f73(a)	0.0000	0.0222	0.0002	0.0003	0.0021	0.0002	0.5786	0.0000
f74(a)	0.0004	0.0001	0.0008	0.0018	0.0001	0.0004	0.0044	0.0001

Table A.27.: F feature selection in real-valued experiments. Rows: features, columns: experiments. Answers (a) subset. Part 2.

	Sobol RFC	Sobol dCor DE	Sobol dCor US	Morris RFC	Morris dCor DE	Morris dCor US	EMO DE	EMO US
f75(a)	0.0016	0.0002	0.0110	0.0021	0.0002	0.0013	0.0338	0.0003
f76(a)	0.0009	0.0005	0.0020	0.0010	0.0003	0.0005	0.0316	0.0004
f77(a)	0.0011	0.0033	0.0001	0.0010	0.0008	0.0001	0.0014	0.0011
f78(a)	0.0011	0.0006	0.0017	0.0027	0.0004	0.0005	0.1263	0.0006
f79(a)	0.0038	0.0000	0.0003	0.0056	0.0000	0.0002	0.0311	0.0001
f80(a)	0.0042	0.0001	0.0020	0.0072	0.0001	0.0005	0.0733	0.0005
f81(a)	0.0000	0.0003	0.0003	0.0019	0.0002	0.0002	0.1244	0.0011
f82(a)	0.0025	0.0000	0.0023	0.0065	0.0000	0.0006	0.0395	0.0002

Table A.28.: Lexical and linguistic feature selection in real-valued experiments. Rows: features, columns: experiments. Answers (a) subset.

	Sobol RFC	Sobol dCor DE	Sobol dCor US	Morris RFC	Morris dCor DE	Morris dCor US	EMO DE	EMO US
ppd(a)	0.0000	0.0002	0.0005	0.0008	0.0002	0.0003	0.0547	0.0001
apd(a)	0.0000	0.0002	0.0013	0.0003	0.0002	0.0004	0.1560	0.0003
adjpd(a)	0.0018	0.0008	0.0007	0.0024	0.0004	0.0003	0.0457	0.0001
advpd(a)	0.0000	0.0023	0.0003	0.0000	0.0007	0.0002	0.0043	0.0000
dpd(a)	0.0000	0.0002	0.0005	0.0000	0.0002	0.0003	0.1095	0.0004
npd(a)	0.0042	0.0038	0.0017	0.0050	0.0009	0.0005	0.0008	0.0000
preppd(a)	0.0042	0.0374	0.0098	0.0084	0.0027	0.0012	0.7387	0.0020
vpd(a)	0.0045	0.0010	0.0040	0.0078	0.0005	0.0008	0.0058	0.0000
R1(a)	0.0094	0.0017	0.0000	0.0109	0.0006	0.0001	0.0102	0.0004
RR(a)	0.0020	0.0012	0.0018	0.0024	0.0005	0.0005	0.0242	0.0005
Q(a)	0.0054	0.0001	0.0002	0.0087	0.0001	0.0001	0.2167	0.0002
alpha(a)	0.0020	0.0000	0.0002	0.0033	0.0001	0.0002	0.1160	0.0001
ATL(a)	0.0729	0.0011	0.0001	0.1031	0.0004	0.0001	0.0150	0.0000
hl(a)	0.0042	0.0016	0.0004	0.0048	0.0006	0.0002	0.0903	0.0000
ttr(a)	0.0009	0.0001	0.0004	0.0016	0.0001	0.0002	0.1133	0.0005
H(a)	0.0016	0.0002	0.0041	0.0014	0.0002	0.0008	0.0111	0.0002
h(a)	0.0018	0.0006	0.0032	0.0040	0.0003	0.0007	0.0513	0.0005
lmbd(a)	0.0007	0.0037	0.0004	0.0014	0.0008	0.0002	0.0694	0.0002
G(a)	0.0027	0.0001	0.0016	0.0043	0.0001	0.0005	0.0847	0.0002
L(a)	0.0040	0.0005	0.0051	0.0056	0.0003	0.0009	0.0000	0.0000
A(a)	0.0018	0.0019	0.0004	0.0022	0.0006	0.0002	0.0003	0.0001
UG(a)	0.0245	0.0008	0.0010	0.0314	0.0004	0.0004	0.1004	0.0003
NDW(a)	0.0033	0.0112	0.0000	0.0046	0.0015	0.0000	0.0766	0.0001
VD(a)	0.0029	0.0004	0.0001	0.0062	0.0003	0.0001	0.0009	0.0000
stc(a)	0.0016	0.0001	0.0160	0.0026	0.0001	0.0016	0.0193	0.0004
ASL(a)	0.0201	0.0004	0.0111	0.0350	0.0003	0.0013	0.5499	0.0000

Table A.29.: Syntactic feature selection in real-valued experiments. Rows: features, columns: experiments. Answers (a) subset. Part 1.

	Sobol RFC	Sobol dCor DE	Sobol dCor US	Morris RFC	Morris dCor DE	Morris dCor US	EMO DE	EMO US
LDEmu(a)	0.0020	0.0002	0.0023	0.0066	0.0002	0.0006	0.0037	0.0003
depmu(a)	0.0056	0.0001	0.0004	0.0057	0.0001	0.0002	0.0210	0.0007
MDDmu(a)	0.0054	0.0050	0.0002	0.0045	0.0010	0.0001	0.0006	0.0006
DDEmu(a)	0.0040	0.0139	0.0044	0.0034	0.0017	0.0008	0.2513	0.0004
TCImu(a)	0.0047	0.0011	0.0021	0.0060	0.0005	0.0005	0.5198	0.0002
imbmu(a)	0.0011	0.0020	0.0013	0.0012	0.0006	0.0004	0.1942	0.0007
Lmu(a)	0.0013	0.0066	0.0001	0.0027	0.0011	0.0001	0.0158	0.0001
Wmu(a)	0.0022	0.0063	0.0026	0.0046	0.0011	0.0006	0.2607	0.0003
wmu(a)	0.0025	0.0011	0.0006	0.0036	0.0004	0.0003	0.0011	0.0001
lmu(a)	0.0000	0.0001	0.0027	0.0000	0.0001	0.0006	0.0056	0.0116
cmu(a)	0.0018	0.0090	0.0104	0.0043	0.0013	0.0013	0.0505	0.0006
LDEG(a)	0.0002	0.0005	0.0020	0.0022	0.0003	0.0005	0.0393	0.0003
depG(a)	0.0016	0.0022	0.0060	0.0039	0.0007	0.0010	0.0139	0.1261
MDDG(a)	0.0016	0.0012	0.0003	0.0021	0.0005	0.0002	0.0310	0.0000
DDEG(a)	0.0076	0.0002	0.0046	0.0143	0.0002	0.0008	0.0010	0.0002
TCIG(a)	0.0031	0.0000	0.0003	0.0039	0.0001	0.0002	0.0147	0.0004
imbG(a)	0.0011	0.0004	0.0000	0.0037	0.0003	0.0001	0.0569	0.0014
LG(a)	0.0000	0.0045	0.0002	0.0009	0.0010	0.0002	0.1286	0.0001
WG(a)	0.0000	0.0007	0.0002	0.0001	0.0004	0.0001	0.0335	0.0012
wG(a)	0.0009	0.0004	0.0007	0.0019	0.0002	0.0003	0.0001	0.0001
lG(a)	0.0002	0.0046	0.0000	0.0009	0.0010	0.0000	0.1279	0.0003
cG(a)	0.0018	0.0005	0.0001	0.0014	0.0003	0.0001	0.1335	0.0001

Table A.30.: Syntactic feature selection in real-valued experiments. Rows: features, columns: experiments. Answers (a) subset. Part 2.

	Sobol RFC	Sobol dCor DE	Sobol dCor US	Morris RFC	Morris dCor DE	Morris dCor US	EMO DE	EMO US
LDEH(a)	0.0071	0.0001	0.0001	0.0069	0.0001	0.0001	0.0889	0.0007
depH(a)	0.0031	0.0002	0.0003	0.0054	0.0002	0.0002	0.0143	0.0011
MDDH(a)	0.0009	0.0000	0.0020	0.0029	0.0001	0.0005	0.0154	0.0059
DDEH(a)	0.0045	0.0002	0.0002	0.0115	0.0002	0.0002	0.0002	0.0011
TCIH(a)	0.0306	0.0000	0.0027	0.0410	0.0001	0.0006	0.0922	0.0002
imbH(a)	0.0029	0.0002	0.0002	0.0070	0.0002	0.0001	0.0011	0.0010
LH(a)	0.0047	0.0002	0.0002	0.0075	0.0002	0.0002	0.1897	0.0005
WH(a)	0.0105	0.0002	0.0002	0.0158	0.0002	0.0002	0.0217	0.0005
wH(a)	0.0141	0.0007	0.0042	0.0156	0.0004	0.0008	0.1414	0.0012
lH(a)	0.0105	0.0002	0.0053	0.0072	0.0002	0.0009	0.1164	0.0057
cH(a)	0.0031	0.0002	0.0002	0.0042	0.0002	0.0002	0.2942	0.0002