Lowering the Language Barrier - Investigating Deep Transfer Learning and Machine Translation for Multilingual Analyses of Political Texts

# Appendix

Computational Communication Research

Special Issue on Multilingual Text Analysis

Research Article

Moritz Laurer, Wouter van Atteveldt, Andreu Casas, Kasper Welbers

## Table of Contents

# A. Datasets

Supervised machine learning analyses are highly dependent on the underlying data. We therefore provide additional details on the data distribution for the first and second analysis. The language and country codes follow ISO standards (ISO 639-1 and ISO 3166-1 Alpha-3 respectively). The party family codes follow the Manifesto project abbreviations[1].

## A1. First analysis

**Comparative Manifesto Project** has categorised more than a million quasi-sentences from party manifestos in 48 languages from 50 countries since 1945 in the Manifesto Corpus. We use the topic identification task based on its eight high-level topical domains. We include the "No other category applies" class to make the analysis more realistic, as most classification tasks will have some sort of 'other' class. Note that this class has very little data and makes the task more difficult. We use a 70/30 train-test split stratified by classes and languages.

*Table 1 - Manifesto-8 training data by language*

| | de | en | es | fr | ko | ru | tr |
|---|---|---|---|---|---|---|---|
| **Economy** | 1177 | 1320 | 1289 | 1211 | 750 | 372 | 1733 |
| **External Relations** | 452 | 354 | 335 | 388 | 199 | 80 | 383 |
| **Fabric of Society** | 686 | 625 | 398 | 420 | 163 | 245 | 568 |
| **Freedom and Democracy** | 358 | 263 | 376 | 316 | 214 | 78 | 482 |
| **No other category applies** | 32 | 33 | 50 | 29 | 1 | 25 | 56 |
| **Political System** | 446 | 510 | 711 | 743 | 150 | 149 | 330 |
| **Social Groups** | 509 | 476 | 467 | 586 | 370 | 149 | 465 |
| **Welfare and Quality of Life** | 1497 | 1571 | 1510 | 1447 | 1155 | 300 | 1050 |

*Table 2 - Manifesto-8 test data by language*

| | de | en | es | fr | ko | ru | tr |
|---|---|---|---|---|---|---|---|
| **Economy** | 995 | 1056 | 1039 | 973 | 322 | 158 | 1441 |
| **External Relations** | 388 | 272 | 267 | 334 | 86 | 33 | 312 |
| **Fabric of Society** | 511 | 491 | 304 | 351 | 69 | 106 | 406 |
| **Freedom and Democracy** | 335 | 215 | 312 | 250 | 89 | 33 | 352 |

[1] https://manifesto-project.wzb.eu/down/papers/Manifesto_Project_Party_Family_Handbook.pdf

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **No other category applies** | 29 | 24 | 36 | 20 | 2 | 10 | 43 |
| **Political System** | 348 | 390 | 594 | 554 | 65 | 62 | 260 |
| **Social Groups** | 406 | 376 | 360 | 463 | 160 | 63 | 406 |
| **Welfare and Quality of Life** | 1149 | 1323 | 1229 | 1190 | 501 | 127 | 838 |

For the **PImPo dataset,** we analyse the stances towards immigration. The crowd annotation task consisted of three main steps: first, determine whether a sentence is about one of the two concepts or not; second, if the text is about one of the concepts, determine which of the two concepts is addressed; third, determine whether the sentence is supportive / sceptical / neutral towards the respective concept. This makes the task a stance detection task which is more complex than the topic identification task in the first dataset. The PImPo project required native speaking crowd-workers where possible, but extended the pool of annotators to potential non-native speakers for languages with too few available crowd workers (Zobel & Lehmann, 2018, p. 1067). The dataset is highly imbalanced: Out of the 200 000+ annotated sentences, the vast majority are not about migration or integration (96%) and only a few hundred sentences per language express a stance towards these concepts of interest.

The dataset also includes stances towards immigrant integration, which we included in the "no topic" class for the purpose of our analysis. Note that the PImPo dataset unfortunately only has little data per language for the individual stances, which makes the test set relatively small for some languages. To increase the test-set size, we use a 60/40 train-test split stratified by classes and languages. Moreover, as the PImPo dataset is highly imbalanced (96% of texts belong to the "no topic" class), we downsample the "no topic" class to maximum 2000 texts per language. Note that only little data was available for French (only Canadian Manifestos were analysed in PImPo) and therefore the maximum no_topic sample in the test-set is 1614 instead of 2000 for French. Methods for addressing data imbalance have been addressed in other research (Miller, Linder, & Mebane, 2020) and we leave more advanced sampling techniques to future research.

*Table 3 - PImPo immigration training data by language*

| | da | de | en | es | fi | fr | nl | no | sv |
|---|---|---|---|---|---|---|---|---|---|
| **immigration_neutral** | 49 | 99 | 40 | 71 | 4 | 13 | 39 | 22 | 11 |
| **immigration_sceptical** | 79 | 299 | 73 | 87 | 38 | 8 | 158 | 108 | 10 |
| **immigration_supportive** | 58 | 463 | 90 | 142 | 21 | 20 | 134 | 167 | 36 |
| **no_topic** | 2000 | 2000 | 2000 | 2000 | 2000 | 2000 | 2000 | 2000 | 2000 |

*Table 4 - PImPo immigration test data by language*

| | da | de | en | es | fi | fr | nl | no | sv |
|---|---|---|---|---|---|---|---|---|---|
| **immigration_neutral** | 32 | 66 | 26 | 48 | 2 | 8 | 26 | 15 | 8 |
| **immigration_sceptical** | 52 | 200 | 48 | 58 | 25 | 6 | 105 | 72 | 6 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **immigration_supportive** | 39 | 309 | 60 | 95 | 14 | 14 | 90 | 111 | 24 |
| **no_topic** | 2000 | 2000 | 2000 | 2000 | 2000 | 1614 | 2000 | 2000 | 2000 |

## A2. Second analysis

The tables below display the data distribution per language and by other variables for the second analysis. For this analysis, we increase the training data size successively by including more languages in three steps. First only English, then English and German, then English, German, Swedish and French. For each language we sample up to 500 texts with stances and an equal number of texts that are not about the topic. For example, this leads to a training dataset of 674 texts in English (337 texts with stances and 337 without) or 1000 texts in German. The test set is always the entire remaining corpus minus the training data. As the corpus is very large and imbalanced (200,000+ texts, 96% about "no-topic"), we downsample the "no-topic" class to 50,000 to reduce the required computations.

*Table 5 - PImPo immigration all data by language*

| | **sv** | **no** | **da** | **fi** | **nl** | **es** | **de** | **en** | **fr** |
|---|---|---|---|---|---|---|---|---|---|
| **immigration_neutral** | 19 | 37 | 81 | 6 | 65 | 119 | 165 | 66 | 21 |
| **immigration_sceptical** | 16 | 180 | 131 | 63 | 263 | 145 | 499 | 121 | 14 |
| **immigration_supportive** | 60 | 278 | 97 | 35 | 224 | 237 | 772 | 150 | 34 |
| **no_topic** | 1655 | 7066 | 1240 | 1756 | 6567 | 8596 | 13951 | 8255 | 914 |

*Table 6 - PImPo immigration all data by party family*

| | **ECO** | **LEF** | **SOC** | **LIB** | **CON** | **AGR** | **CHR** | **NAT** | **SIP** | **ETH** |
|---|---|---|---|---|---|---|---|---|---|---|
| **immigration_neutral** | 58 | 75 | 79 | 81 | 62 | 9 | 100 | 65 | 5 | 45 |
| **immigration_sceptical** | 41 | 72 | 143 | 156 | 147 | 26 | 261 | 505 | 44 | 37 |
| **immigration_supportive** | 424 | 255 | 294 | 284 | 144 | 47 | 263 | 59 | 43 | 74 |
| **no_topic** | 5448 | 5828 | 8817 | 5745 | 5975 | 2728 | 7106 | 3187 | 1903 | 3263 |

*Table 7 - PImPo immigration all data by country*

| | **swe** | **nor** | **dnk** | **fin** | **nld** | **esp** | **deu** | **aut** | **che** | **irl** | **usa** | **can** | **aus** | **nzl** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Immigration neutral** | 19 | 37 | 81 | 6 | 65 | 119 | 68 | 74 | 23 | 34 | 9 | 21 | 19 | 4 |
| **Immigration sceptical** | 16 | 180 | 131 | 63 | 263 | 145 | 107 | 297 | 95 | 37 | 42 | 14 | 41 | 1 |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Immigration supportive** | 57 | 278 | 97 | 38 | 224 | 237 | 543 | 151 | 78 | 60 | 38 | 34 | 49 | 3 |
| **no_topic** | 1598 | 7066 | 1240 | 1813 | 6567 | 8596 | 8628 | 4223 | 1100 | 3564 | 1969 | 914 | 1882 | 840 |

## B. Machine translation

All machine translations are implemented by the authors with free, open-source machine translation algorithms. We use two "M2M-100" models by (Fan et al., 2020), which are capable of translating 100 different languages in all language directions. We use these models as they provide good translation quality, while covering all languages and translation directions in our datasets. For the first analysis we tested both the medium sized M2M100_418M[2] model and the large M2M100_1.2B[3] model and the final results were calculated based on translations from the larger model. As the 1.2B model is very slow and we did not notice meaningful differences when we trained our classifiers with texts translated by the medium sized model, we only used the medium sized model for the second analysis. The main advantage of these algorithms is that they are freely available online, cover many translation directions, and are relatively easy to use with the EasyNMT package[4]. Their disadvantage is that they are relatively large encoder-decoder Transformers and require a GPU to translate larger corpora.

## C. BERT variants for supervised classification

While we refer to "BERT" in the main text for simplicity (Devlin, Chang, Lee, & Toutanova, 2019), we actually use "**DeBERTaV3**" (He, Gao, & Chen, 2021) for all experiments, except Sentence-BERT. There are three main differences between BERT and DeBERTaV3: "First, it is pre-trained on more data. The original BERT model is trained on 16GB of text from Wikipedia and Books, while DeBERTaV3 is trained on 78GB of text from Wikipedia, books, web texts like blogs, story-like texts, and news. Second, DeBERTaV3 uses 'disentangled attention', where each token (word) is not only represented as one vector, but as two vectors: one representing the word's content and one representing its position in the text. Third, version three of DeBERTa (hence DeBERTaV3) does not use the classical masked-language-modeling objective for pre-training anymore, but uses replaced-token-detection, which is more effective at creating general language representations" (appendix of Laurer, Van Atteveldt, Casas, & Welbers, 2023). Moreover, a multilingual version of DeBERTaV3 is also available, which makes direct comparisons between an English and multilingual model easier.

[2] https://huggingface.co/facebook/m2m100_418M
[3] https://huggingface.co/facebook/m2m100_1.2B
[4] https://github.com/UKPLab/EasyNMT

The **BERT-NLI** models were trained by the authors. We used (m)DeBERTaV3 as a base model. The English variant was trained on a collection of several hundred thousand English hypothesis-context pairs, creating the English "DeBERTa-v3-base-mnli-fever-anli" in base size and "DeBERTa-v3-large-mnli-fever-anli-ling-wanli" in large size. These models are similar to, but slightly updated versions of the BERT-NLI model used in (Laurer et al., 2023). For more information on the datasets used, please refer to this prior publication and the following Hugging Face model repository.[5] Moreover, for this paper we also created an improved multilingual model. We first took the English hypothesis-context pairs and machine translated a random sample of 105 000 texts for each of 26 different languages. This created a multilingual NLI training dataset of 2 730 000 NLI text pairs. The languages were chosen based on two criteria: (1) They are either included in the list of the 20 most spoken languages on Wikipedia[6] in August 2022 (excluding Telugu and Nigerian Pidgin, for which no machine translation model was available); (2) or they are spoken in polit-economically important countries such as the G20 or Iran and Israel. Based on the information available on Wikipedia, the languages are spoken by more than 4 billion people. The resulting dataset "multilingual-NLI-26lang-2mil7"[7] and model "mDeBERTa-v3-base-xnli-multilingual-nli-2mil7"[8] are also freely available online in Moritz Laurer's Hugging Face repository.

No **Sentence-BERT** variant based on DeBERTaV3 was available online. We therefore chose the best performing Sentence-BERT variant available on sbert.net which provided both an English and multilingual version (as of August 2022): "paraphrase-multilingual-mpnet-base-v2" as the multilingual variant, and "paraphrase-mpnet-base-v2" as the English variant (Reimers & Gurevych, 2019, 2020).[9] For the **Logistic Regression**, we use the Scikit-learn implementation (Pedregosa et al., 2011).[10]

# D. Details on the analyses

## D1. Combinations of algorithms, MT and data augmentation in analysis 1

The first analysis required a relatively complex analysis pipeline that can handle many different combinations of algorithms, MT and data augmentation strategies. To reduce computation costs, we translated all texts in all other languages once. The downstream analyses were then always on

[5] All models are available at: https://huggingface.co/MoritzLaurer
[6] https://en.wikipedia.org/wiki/List_of_languages_by_total_number_of_speakers
[7] https://huggingface.co/datasets/MoritzLaurer/multilingual-NLI-26lang-2mil7
[8] https://huggingface.co/MoritzLaurer/mDeBERTa-v3-base-xnli-multilingual-nli-2mil7
[9] https://sbert.net/docs/pretrained_models.html
[10] https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html

a subset of this large, augmented dataset. The tables below display which combinations were possible, and which subset of the data was used for training and testing. "Iter" denotes the number of iterations for training and testing necessary for the approach; "lang-original train" denotes the original language of the training data; "lang-trans train" denotes the translated/augmented language of the training data used in the analysis; "lang-original test" denotes the original language of the test data; "lang-trans test" denotes the translated language of the training data used in the analysis.

*Table 8 - Low-resource training data scenario: Combinations of algorithms and machine translation*

| MT Strategy / Algorithms | no-MT | one2anchor | one2many |
|---|---|---|---|
| **TF-IDF + Logistic Regression** | NA | Iterations: 1 train[11], 7 test<br>lang-original train: one (en)<br>lang-trans train: anchor (en)<br>lang-original test: target 7x<br>lang-trans test: anchor (en) | NA |
| **Sent.-BERT-mono + Log. Reg** | NA | Iter: 1 train, 7 test<br>lang-original train: one (en)<br>lang-trans train: anchor (en)<br>lang-original test: target 7x<br>lang-trans test: anchor (en) | NA |
| **Sent.-BERT-multi + Log. Reg.** | Iter: 1 train, 7 test<br>lang-original train: one (en)<br>lang-trans train: - (one)<br>lang-original test: target 7x<br>lang-trans test: - (target 7x) | Iter: 7 train, 7 test<br>lang-original train: one (en)<br>lang-trans train: one (en) + target-lang<br>lang-original test: target 7x<br>lang-trans test: target 7x | Iter: 1 train, 7 test<br>lang-original train: one (en)<br>lang-trans train: one (en) + many<br>lang-original test: target 7x<br>lang-trans test: - (target 7x) |
| **BERT-mono (Both BERT-base and BERT-NLI)** | NA | Iter: 1 train, 7 test<br>lang-original train: one (en)<br>lang-trans train: anchor (en)<br>lang-original test: target 7x<br>lang-trans test: anchor (en) | NA |
| **BERT-multi (Both mBERT-base and mBERT-NLI)** | Iter: 1 train, 7 test<br>lang-original train: one (en)<br>lang-trans train: - (one)<br>lang-original test: target 7x<br>lang-trans test: - (target 7x) | Iter: 7 train, 7 test<br>lang-original train: one (en)<br>lang-trans train: one (en) + target-lang<br>lang-original test: target 7x<br>lang-trans test: target 7x | Iter: 1 train, 7 test<br>lang-original train: one (en)<br>lang-trans train: one (en) + many<br>lang-original test: target 7x<br>lang-trans test: - (target 7x) |

[11] In practice, we needed to vectorizer the train set 7 different times to meet Scikit-learn TF-IDF requirements, when vectorizing the train and test set together.

*Table 9 - Higher-resource training data scenario: Combinations of algorithms and machine translation*

| MT Strategy / Algorithms | no-MT | many2anchor | many2many |
|---|---|---|---|
| **TF-IDF + Logistic Regression** | (separate models per lang)<br>Iterations: 7 train, 7 test<br>lang-original train: one<br>lang-trans train: - (one)<br>lang-original test: one<br>lang-trans test: - (one) | Iterations: 1 train, 7 test<br>lang-original train: many<br>lang-trans train: anchor (en)<br>lang-original test: many<br>lang-trans test: anchor (en) | (separate models per lang)<br>Iterations: 7 train, 7 test<br>lang-original train: many<br>lang-trans train: one<br>lang-original test: one<br>lang-trans test: one |
| **Sent.-BERT-mono + Log. Reg** | (separate models per lang)<br>Iter: 7 train, 7 test<br>lang-original train: one<br>lang-trans train: - (one)<br>lang-original test: one<br>lang-trans test: - (one) | Iter: 1 train, 7 test<br>lang-original train: many<br>lang-trans train: anchor (en)<br>lang-original test: many<br>lang-trans test: anchor (en) | (separate models per lang)<br>Iter: 7 train, 7 test<br>lang-original train: many<br>lang-trans train: one<br>lang-original test: one<br>lang-trans test: one |
| **Sent.-BERT-multi + Log. Reg.** | Iter: 1 train, 7 test<br>lang-original train: many<br>lang-trans train: - (many-original)<br>lang-original test: many<br>lang-trans test: - (many-original) | Iter: 1 train, 7 test<br>lang-original train: many<br>lang-trans train: - many-original +many2anchor<br>lang-original test: many<br>lang-trans test: - (many-original) | Iter: 1 train, 7 test<br>lang-original train: many<br>lang-trans train: many-original + many2many<br>lang-original test: many<br>lang-trans test: - (many-original) |
| **BERT-mono (Both BERT-base and BERT-NLI)** | (separate models per lang)<br>Iter: 7 train, 7 test<br>lang-original train: one<br>lang-trans train: - (one)<br>lang-original test: one<br>lang-trans test: - (one) | Iter: 1 train, 7 test<br>lang-original train: many<br>lang-trans train: anchor (en)<br>lang-original test: many<br>lang-trans test: anchor (en) | (separate models per lang)<br>Iter: 7 train, 7 test<br>lang-original train: many<br>lang-trans train: one<br>lang-original test: one<br>lang-trans test: one |
| **BERT-multi (Both mBERT-base and mBERT-NLI)** | Iter: 1 train, 7 test<br>lang-original train: many<br>lang-trans train: - (many-original)<br>lang-original test: many<br>lang-trans test: - (many-original) | Iter: 1 train, 7 test<br>lang-original train: many<br>lang-trans train: - many-original +many2anchor<br>lang-original test: many<br>lang-trans test: - (many-original) | Iter: 1 train, 7 test<br>lang-original train: many<br>lang-trans train: many-original + many2many<br>lang-original test: many<br>lang-trans test: - (many-original) |

## D2: Exact metrics

The tables below display the exact metrics that were used for figure 2 in the main text.

*Table 10 - Low-resource scenario: 500 English training texts, tests on 7 languages*

| | | | Manifesto | | | | PImPo | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **algorithm** | **language representation** | **MT augmentation** | **F1 Macro** | **Acc.** | **F1 Std.** | **Acc. Std.** | **F1 Macro** | **Acc.** | **F1 Std.** | **Acc. Std.** |
| BERT-NLI | multiling-embeddings | one2many | 0,44 | 0,53 | 0,033 | 0,046 | 0,49 | 0,94 | 0,034 | 0,04 |
| | | one2anchor | 0,41 | 0,49 | 0,03 | 0,046 | 0,45 | 0,94 | 0,048 | 0,035 |
| | | no-mt-mono | 0,4 | 0,48 | 0,03 | 0,043 | 0,45 | 0,92 | 0,046 | 0,039 |
| | monoling-embeddings | one2anchor | 0,41 | 0,49 | 0,028 | 0,041 | 0,49 | 0,94 | 0,037 | 0,039 |
| BERT-base | multiling-embeddings | one2many | 0,37 | 0,49 | 0,031 | 0,041 | 0,45 | 0,94 | 0,039 | 0,042 |
| | | one2anchor | 0,33 | 0,45 | 0,023 | 0,028 | 0,45 | 0,93 | 0,031 | 0,042 |
| | | no-mt-mono | 0,27 | 0,4 | 0,028 | 0,055 | 0,42 | 0,93 | 0,029 | 0,046 |
| | monoling-embeddings | one2anchor | 0,34 | 0,44 | 0,027 | 0,036 | 0,44 | 0,94 | 0,039 | 0,043 |
| Sent.-BERT + Log. Reg. | multiling-embeddings | one2many | 0,39 | 0,51 | 0,032 | 0,04 | 0,45 | 0,93 | 0,05 | 0,04 |
| | | one2anchor | 0,38 | 0,49 | 0,032 | 0,043 | 0,44 | 0,94 | 0,041 | 0,041 |
| | | no-mt-mono | 0,39 | 0,5 | 0,027 | 0,037 | 0,42 | 0,93 | 0,033 | 0,045 |
| | monoling-embeddings | one2anchor | 0,36 | 0,46 | 0,024 | 0,032 | 0,44 | 0,94 | 0,045 | 0,044 |
| Logistic Regression | TF-IDF | one2anchor | 0,27 | 0,4 | 0,019 | 0,027 | 0,35 | 0,92 | 0,028 | 0,052 |

*Table 11 - Higher-resource scenario: 500 training texts for 7 languages each*

| | | | Manifesto | | | | PImPo | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **algorithm** | **language representation** | **MT augmentation** | **F1 Macro** | **Acc.** | **F1 Std.** | **Acc. Std.** | **F1 Macro** | **Acc.** | **F1 Std.** | **Acc. Std.** |
| BERT-NLI | multiling-embeddings | many2many | 0,49 | 0,58 | 0,044 | 0,052 | 0,49 | 0,95 | 0,058 | 0,03 |
| | | many2anchor | 0,48 | 0,57 | 0,05 | 0,062 | 0,48 | 0,93 | 0,043 | 0,038 |
| | | no-mt-multi | 0,47 | 0,56 | 0,042 | 0,058 | 0,47 | 0,93 | 0,039 | 0,042 |
| | monoling*-embeddings | many2many* ** | 0,42 | 0,52 | 0,032 | 0,041 | 0,47 | 0,94 | 0,061 | 0,032 |
| | | many2anchor | 0,46 | 0,55 | 0,036 | 0,04 | 0,49 | 0,94 | 0,05 | 0,034 |
| | | no-mt-multi* ** | 0,39 | 0,48 | 0,042 | 0,05 | 0,47 | 0,94 | 0,048 | 0,046 |
| BERT-base | multiling-embeddings | many2many | 0,45 | 0,56 | 0,044 | 0,058 | 0,5 | 0,93 | 0,063 | 0,031 |
| | | many2anchor | 0,47 | 0,58 | 0,048 | 0,062 | 0,54 | 0,95 | 0,046 | 0,033 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | no-mt-multi | 0,47 | 0,58 | 0,052 | 0,067 | 0,52 | 0,94 | 0,047 | 0,037 |
| | monoling*-embeddings | many2many* ** | 0,42 | 0,53 | 0,03 | 0,039 | 0,48 | 0,94 | 0,069 | 0,033 |
| | | many2anchor | 0,45 | 0,55 | 0,036 | 0,038 | 0,53 | 0,95 | 0,043 | 0,035 |
| | | no-mt-multi* ** | 0,36 | 0,46 | 0,05 | 0,061 | 0,42 | 0,93 | 0,068 | 0,053 |
| Sent.-BERT + Log. Reg. | multiling-embeddings | many2many | 0,46 | 0,57 | 0,045 | 0,052 | 0,51 | 0,94 | 0,038 | 0,035 |
| | | many2anchor | 0,47 | 0,58 | 0,046 | 0,054 | 0,51 | 0,95 | 0,041 | 0,034 |
| | | no-mt-multi | 0,47 | 0,59 | 0,046 | 0,053 | 0,51 | 0,95 | 0,044 | 0,033 |
| | monoling*-embeddings | many2many* ** | 0,45 | 0,56 | 0,046 | 0,057 | 0,49 | 0,94 | 0,049 | 0,032 |
| | | many2anchor | 0,44 | 0,55 | 0,036 | 0,039 | 0,52 | 0,95 | 0,046 | 0,034 |
| | | no-mt-multi* ** | 0,44 | 0,55 | 0,066 | 0,077 | 0,46 | 0,94 | 0,053 | 0,047 |
| Logistic Regression | TF-IDF | many2many** | 0,37 | 0,49 | 0,039 | 0,047 | 0,43 | 0,91 | 0,043 | 0,059 |
| | | many2anchor | 0,38 | 0,5 | 0,038 | 0,04 | 0,42 | 0,94 | 0,029 | 0,044 |
| | | no-mt-multi** | 0,33 | 0,45 | 0,058 | 0,073 | 0,39 | 0,94 | 0,068 | 0,048 |

* For these combinations of approaches, we had to use mBERT with monolingual input, as monolingual Transformers do not exist for all languages.
** These approaches required training different models for each language, as the training data is multilingual, but the model is monolingual.

In addition, the tables below displays the F1 score for each individual class separately. Note that these values are averages over multiple random seeds and multiple languages, as the table would have been too long otherwise. The tables display the numeric class labels. These can be mapped to the following labels:

- Manifesto: {0: 'Economy', 1: 'External Relations', 2: 'Fabric of Society', 3: 'Freedom and Democracy', 4: 'No other category applies', 5: 'Political System', 6: 'Social Groups', 7: 'Welfare and Quality of Life'}
- PImPo: {0: 'immigration_neutral', 1: 'immigration_sceptical', 2: 'immigration_supportive', 3: 'no_topic'}

The column “F1-lang-std” refers to the cross-language standard deviation of F1-macro scores.

We observe that, for the Manifesto datasets, the most difficult class is the “No other category applies” class, which contains text that does not fit into the other classes. We included this category to increase the realism of the task. For the PImPo dataset, we observe that the immigration_neutral class is the most difficult class, while the “no_topic” class is easy to learn and constitutes the majority of all texts. These observations underline the importance of metrics like F1-macro, which attribute equal importance to all classes independently of their size.

*Table 12 - Manifesto dataset per-class F1 metrics*

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | vectorizer | augmentation | method | F1-lang-std |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.66 | 0.64 | 0.51 | 0.49 | 0.07 | 0.49 | 0.64 | 0.39 | embeddings-multi | many2many | nli | 0.04 |
| 0.66 | 0.65 | 0.5 | 0.52 | 0.06 | 0.49 | 0.64 | 0.39 | embeddings-multi | many2many | nli | 0.05 |
| 0.66 | 0.61 | 0.49 | 0.47 | 0.05 | 0.48 | 0.64 | 0.39 | embeddings-multi | many2anchor | nli | 0.05 |
| 0.66 | 0.62 | 0.5 | 0.46 | 0.06 | 0.49 | 0.64 | 0.42 | embeddings-multi | many2anchor | nli | 0.05 |
| 0.63 | 0.6 | 0.44 | 0.44 | 0.03 | 0.43 | 0.6 | 0.37 | embeddings-multi | one2many | nli | 0.03 |
| 0.62 | 0.59 | 0.43 | 0.43 | 0.03 | 0.44 | 0.61 | 0.36 | embeddings-multi | one2many | nli | 0.04 |
| 0.61 | 0.55 | 0.4 | 0.39 | 0.04 | 0.39 | 0.57 | 0.35 | embeddings-multi | one2anchor | nli | 0.03 |
| 0.6 | 0.54 | 0.39 | 0.4 | 0.02 | 0.39 | 0.57 | 0.35 | embeddings-multi | one2anchor | nli | 0.03 |
| 0.66 | 0.61 | 0.48 | 0.44 | 0.02 | 0.47 | 0.63 | 0.41 | embeddings-multi | no-nmt-many | nli | 0.04 |
| 0.65 | 0.61 | 0.48 | 0.44 | 0.02 | 0.47 | 0.63 | 0.42 | embeddings-multi | no-nmt-many | nli | 0.04 |
| 0.59 | 0.54 | 0.4 | 0.38 | 0.03 | 0.37 | 0.55 | 0.36 | embeddings-multi | no-nmt-single | nli | 0.03 |
| 0.59 | 0.54 | 0.4 | 0.38 | 0.03 | 0.37 | 0.55 | 0.36 | embeddings-multi | no-nmt-single | nli | 0.03 |
| 0.62 | 0.56 | 0.43 | 0.45 | 0.03 | 0.41 | 0.59 | 0.35 | embeddings-en | many2many | nli | 0.04 |
| 0.62 | 0.56 | 0.42 | 0.43 | 0.02 | 0.42 | 0.59 | 0.34 | embeddings-en | many2many | nli | 0.03 |
| 0.64 | 0.58 | 0.45 | 0.48 | 0.04 | 0.45 | 0.63 | 0.39 | embeddings-en | many2anchor | nli | 0.04 |
| 0.65 | 0.61 | 0.46 | 0.48 | 0.04 | 0.45 | 0.63 | 0.4 | embeddings-en | many2anchor | nli | 0.03 |
| 0.59 | 0.52 | 0.41 | 0.43 | 0.03 | 0.4 | 0.56 | 0.36 | embeddings-en | one2anchor | nli | 0.03 |
| 0.59 | 0.5 | 0.39 | 0.42 | 0.03 | 0.4 | 0.55 | 0.36 | embeddings-en | one2anchor | nli | 0.03 |
| 0.59 | 0.51 | 0.39 | 0.4 | 0.04 | 0.37 | 0.54 | 0.3 | embeddings-en | no-nmt-many | nli | 0.04 |
| 0.59 | 0.51 | 0.39 | 0.4 | 0.04 | 0.37 | 0.54 | 0.3 | embeddings-en | no-nmt-many | nli | 0.04 |
| 0.65 | 0.61 | 0.46 | 0.46 | 0.03 | 0.37 | 0.43 | 0.63 | embeddings-multi | many2many | standard_dl | 0.04 |
| 0.65 | 0.63 | 0.47 | 0.46 | 0.02 | 0.39 | 0.43 | 0.62 | embeddings-multi | many2many | standard_dl | 0.05 |
| 0.66 | 0.64 | 0.48 | 0.48 | 0.02 | 0.41 | 0.44 | 0.64 | embeddings-multi | many2anchor | standard_dl | 0.05 |
| 0.66 | 0.63 | 0.49 | 0.49 | 0.05 | 0.38 | 0.42 | 0.65 | embeddings-multi | many2anchor | standard_dl | 0.06 |
| 0.6 | 0.49 | 0.39 | 0.31 | 0.01 | 0.31 | 0.29 | 0.56 | embeddings-multi | one2many | standard_dl | 0.03 |
| 0.58 | 0.51 | 0.38 | 0.35 | 0 | 0.26 | 0.26 | 0.54 | embeddings-multi | one2many | standard_dl | 0.03 |
| 0.55 | 0.49 | 0.34 | 0.27 | 0 | 0.28 | 0.2 | 0.53 | embeddings-multi | one2anchor | standard_dl | 0.02 |
| 0.57 | 0.47 | 0.33 | 0.27 | 0 | 0.3 | 0.21 | 0.54 | embeddings-multi | one2anchor | standard_dl | 0.03 |
| 0.66 | 0.63 | 0.49 | 0.49 | 0.03 | 0.37 | 0.44 | 0.64 | embeddings-multi | no-nmt-many | standard_dl | 0.05 |
| 0.66 | 0.64 | 0.48 | 0.48 | 0.02 | 0.36 | 0.46 | 0.64 | embeddings-multi | no-nmt-many | standard_dl | 0.05 |
| 0.53 | 0.4 | 0.3 | 0.13 | 0 | 0.25 | 0.11 | 0.47 | embeddings-multi | no-nmt-single | standard_dl | 0.03 |
| 0.54 | 0.44 | 0.3 | 0.13 | 0 | 0.21 | 0.11 | 0.47 | embeddings-multi | no-nmt-single | standard_dl | 0.03 |
| 0.62 | 0.58 | 0.43 | 0.43 | 0.02 | 0.34 | 0.39 | 0.59 | embeddings-en | many2many | standard_dl | 0.03 |
| 0.62 | 0.56 | 0.45 | 0.44 | 0.01 | 0.33 | 0.39 | 0.6 | embeddings-en | many2many | standard_dl | 0.03 |
| 0.65 | 0.63 | 0.49 | 0.48 | 0.02 | 0.4 | 0.41 | 0.64 | embeddings-en | many2anchor | standard_dl | 0.04 |
| 0.63 | 0.61 | 0.45 | 0.47 | 0.01 | 0.39 | 0.42 | 0.63 | embeddings-en | many2anchor | standard_dl | 0.04 |
| 0.55 | 0.46 | 0.34 | 0.3 | 0 | 0.3 | 0.2 | 0.53 | embeddings-en | one2anchor | standard_dl | 0.03 |
| 0.55 | 0.49 | 0.36 | 0.31 | 0 | 0.31 | 0.25 | 0.53 | embeddings-en | one2anchor | standard_dl | 0.03 |
| 0.54 | 0.48 | 0.38 | 0.36 | 0.01 | 0.29 | 0.29 | 0.54 | embeddings-en | no-nmt-many | standard_dl | 0.05 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.54 | 0.48 | 0.39 | 0.35 | 0.01 | 0.3 | 0.29 | 0.54 | embeddings-en | no-nmt-many | standard_dl | 0.05 |
| 0.65 | 0.62 | 0.47 | 0.49 | 0.01 | 0.4 | 0.43 | 0.63 | embeddings-multi | many2many | classical_ml | 0.05 |
| 0.65 | 0.61 | 0.47 | 0.48 | 0.01 | 0.4 | 0.43 | 0.64 | embeddings-multi | many2many | classical_ml | 0.04 |
| 0.66 | 0.63 | 0.48 | 0.5 | 0.01 | 0.41 | 0.43 | 0.64 | embeddings-multi | many2anchor | classical_ml | 0.05 |
| 0.66 | 0.63 | 0.48 | 0.5 | 0.01 | 0.41 | 0.44 | 0.64 | embeddings-multi | many2anchor | classical_ml | 0.05 |
| 0.58 | 0.49 | 0.39 | 0.38 | 0 | 0.33 | 0.31 | 0.57 | embeddings-multi | one2many | classical_ml | 0.03 |
| 0.6 | 0.51 | 0.41 | 0.37 | 0 | 0.35 | 0.31 | 0.58 | embeddings-multi | one2many | classical_ml | 0.03 |
| 0.58 | 0.5 | 0.4 | 0.37 | 0 | 0.34 | 0.32 | 0.57 | embeddings-multi | one2anchor | classical_ml | 0.03 |
| 0.59 | 0.5 | 0.4 | 0.37 | 0 | 0.35 | 0.31 | 0.57 | embeddings-multi | one2anchor | classical_ml | 0.03 |
| 0.66 | 0.63 | 0.49 | 0.51 | 0.01 | 0.41 | 0.44 | 0.64 | embeddings-multi | no-nmt-many | classical_ml | 0.05 |
| 0.66 | 0.63 | 0.49 | 0.51 | 0.01 | 0.41 | 0.44 | 0.64 | embeddings-multi | no-nmt-many | classical_ml | 0.05 |
| 0.59 | 0.51 | 0.4 | 0.37 | 0 | 0.35 | 0.3 | 0.57 | embeddings-multi | no-nmt-single | classical_ml | 0.03 |
| 0.59 | 0.51 | 0.4 | 0.37 | 0 | 0.35 | 0.3 | 0.57 | embeddings-multi | no-nmt-single | classical_ml | 0.03 |
| 0.64 | 0.6 | 0.45 | 0.48 | 0.01 | 0.39 | 0.41 | 0.62 | embeddings-en | many2many | classical_ml | 0.05 |
| 0.64 | 0.61 | 0.45 | 0.48 | 0.02 | 0.39 | 0.43 | 0.62 | embeddings-en | many2many | classical_ml | 0.05 |
| 0.63 | 0.59 | 0.44 | 0.46 | 0 | 0.38 | 0.39 | 0.6 | embeddings-en | many2anchor | classical_ml | 0.04 |
| 0.63 | 0.62 | 0.46 | 0.47 | 0 | 0.39 | 0.41 | 0.62 | embeddings-en | many2anchor | classical_ml | 0.04 |
| 0.55 | 0.46 | 0.37 | 0.35 | 0 | 0.31 | 0.31 | 0.52 | embeddings-en | one2anchor | classical_ml | 0.02 |
| 0.56 | 0.48 | 0.39 | 0.36 | 0 | 0.31 | 0.32 | 0.53 | embeddings-en | one2anchor | classical_ml | 0.02 |
| 0.63 | 0.59 | 0.45 | 0.47 | 0.01 | 0.37 | 0.39 | 0.59 | embeddings-en | no-nmt-many | classical_ml | 0.07 |
| 0.63 | 0.59 | 0.45 | 0.47 | 0.01 | 0.37 | 0.4 | 0.59 | embeddings-en | no-nmt-many | classical_ml | 0.07 |
| 0.56 | 0.45 | 0.38 | 0.41 | 0.02 | 0.28 | 0.34 | 0.56 | tfidf | many2many | classical_ml | 0.05 |
| 0.57 | 0.43 | 0.36 | 0.4 | 0.01 | 0.28 | 0.33 | 0.55 | tfidf | many2many | classical_ml | 0.04 |
| 0.59 | 0.48 | 0.37 | 0.39 | 0 | 0.3 | 0.36 | 0.58 | tfidf | many2anchor | classical_ml | 0.04 |
| 0.58 | 0.45 | 0.37 | 0.39 | 0 | 0.3 | 0.34 | 0.57 | tfidf | many2anchor | classical_ml | 0.04 |
| 0.49 | 0.26 | 0.27 | 0.21 | 0.01 | 0.21 | 0.26 | 0.49 | tfidf | one2anchor | classical_ml | 0.02 |
| 0.49 | 0.26 | 0.27 | 0.23 | 0.01 | 0.21 | 0.26 | 0.49 | tfidf | one2anchor | classical_ml | 0.02 |
| 0.53 | 0.34 | 0.33 | 0.33 | 0.01 | 0.27 | 0.29 | 0.52 | tfidf | no-nmt-many | classical_ml | 0.06 |
| 0.53 | 0.34 | 0.33 | 0.33 | 0.01 | 0.27 | 0.29 | 0.52 | tfidf | no-nmt-many | classical_ml | 0.06 |

*Table 13 - PImPo dataset per-class F1 metrics*

| 0 | 1 | 2 | 3 | vectorizer | augmentation | method | F1-lang-std |
|---|---|---|---|---|---|---|---|
| 0.04 | 0.44 | 0.51 | 0.98 | embeddings-multi | many2many | nli | 0.06 |
| 0.11 | 0.36 | 0.47 | 0.98 | embeddings-multi | many2anchor | nli | 0.04 |
| 0.15 | 0.4 | 0.44 | 0.98 | embeddings-multi | one2many | nli | 0.03 |
| 0.08 | 0.29 | 0.44 | 0.98 | embeddings-multi | one2anchor | nli | 0.05 |
| 0.08 | 0.42 | 0.4 | 0.98 | embeddings-multi | no-nmt-many | nli | 0.04 |
| 0.13 | 0.34 | 0.34 | 0.98 | embeddings-multi | no-nmt-single | nli | 0.05 |
| 0.06 | 0.4 | 0.43 | 0.98 | embeddings-en | many2many | nli | 0.06 |
| 0.04 | 0.44 | 0.5 | 0.98 | embeddings-en | many2anchor | nli | 0.05 |
| 0.12 | 0.38 | 0.49 | 0.98 | embeddings-en | one2anchor | nli | 0.04 |
| 0.08 | 0.37 | 0.44 | 0.98 | embeddings-en | no-nmt-many | nli | 0.05 |
| 0.14 | 0.38 | 0.49 | 0.98 | embeddings-multi | many2many | standard_dl | 0.06 |
| 0.21 | 0.44 | 0.53 | 0.98 | embeddings-multi | many2anchor | standard_dl | 0.05 |
| 0.1 | 0.28 | 0.44 | 0.98 | embeddings-multi | one2many | standard_dl | 0.04 |
| 0.14 | 0.26 | 0.43 | 0.98 | embeddings-multi | one2anchor | standard_dl | 0.03 |
| 0.18 | 0.43 | 0.47 | 0.98 | embeddings-multi | no-nmt-many | standard_dl | 0.05 |
| 0.09 | 0.29 | 0.31 | 0.97 | embeddings-multi | no-nmt-single | standard_dl | 0.03 |
| 0.15 | 0.36 | 0.45 | 0.98 | embeddings-en | many2many | standard_dl | 0.07 |
| 0.16 | 0.46 | 0.5 | 0.98 | embeddings-en | many2anchor | standard_dl | 0.04 |
| 0.1 | 0.31 | 0.37 | 0.98 | embeddings-en | one2anchor | standard_dl | 0.04 |
| 0.09 | 0.3 | 0.32 | 0.97 | embeddings-en | no-nmt-many | standard_dl | 0.07 |
| 0.14 | 0.43 | 0.49 | 0.98 | embeddings-multi | many2many | classical_ml | 0.04 |
| 0.13 | 0.43 | 0.51 | 0.98 | embeddings-multi | many2anchor | classical_ml | 0.04 |
| 0.1 | 0.29 | 0.42 | 0.97 | embeddings-multi | one2many | classical_ml | 0.05 |
| 0.1 | 0.28 | 0.42 | 0.97 | embeddings-multi | one2anchor | classical_ml | 0.04 |
| 0.11 | 0.43 | 0.51 | 0.98 | embeddings-multi | no-nmt-many | classical_ml | 0.04 |
| 0.06 | 0.31 | 0.35 | 0.97 | embeddings-multi | no-nmt-single | classical_ml | 0.03 |
| 0.11 | 0.4 | 0.47 | 0.98 | embeddings-en | many2many | classical_ml | 0.05 |
| 0.14 | 0.43 | 0.51 | 0.98 | embeddings-en | many2anchor | classical_ml | 0.05 |
| 0.11 | 0.31 | 0.38 | 0.98 | embeddings-en | one2anchor | classical_ml | 0.05 |
| 0.11 | 0.34 | 0.42 | 0.98 | embeddings-en | no-nmt-many | classical_ml | 0.05 |
| 0.1 | 0.32 | 0.36 | 0.97 | tfidf | many2many | classical_ml | 0.04 |
| 0.04 | 0.29 | 0.37 | 0.97 | tfidf | many2anchor | classical_ml | 0.03 |
| 0.04 | 0.12 | 0.26 | 0.96 | tfidf | one2anchor | classical_ml | 0.03 |
| 0.08 | 0.27 | 0.25 | 0.97 | tfidf | no-nmt-many | classical_ml | 0.07 |

## D3. Preprocessing, randomness and hyperparameters

**Text preprocessing** can have an important impact on computational text analyses methods and different methods need to be handled differently. For TF-IDF, we used lemmatization for each language (except Turkish, for which we did not find an open-source lemmatizer), stop-word removal with publicly available stop-word lists for each language. Our main source of information was the SpaCy library (Montani et al., 2022). During hyperparameter search, we also tested different thresholds for the removal of very (in)frequent words and for different n-gram ranges on the word- or character-level using Scikit-learn (Pedregosa et al., 2011). For the BERT models, text processing was easier, as neither English nor multilingual BERT models require special text pre-processing except for the out-of-the-box tokenizer. Only for BERT-NLI did we add the "The quote is about {text}" string to each text, as described in (Laurer et al., 2023). Moreover, for all different algorithms we used only single quasi-sentences for the first analysis and a concatenation of the preceding+target+following quasi-sentence for the second analysis. Concatenating texts with the noisy Manifesto corpus improves performance and PImPo is based on the Manifesto corpus. We did not conduct concatenation for the first analysis as it would have further increased the data augmentation complexity of the many scenarios tested in the first analysis.

To handle **randomness and increase the robustness** of our metrics, each metric is based on three runs with three different random seeds in the first analysis. For the second analysis, we could not use different random samples, as we only used the number of texts available in one language and used the entire remaining corpus as the test set. Moreover, the number of texts available in PImPo per language was too small to create meaningful variation through different random training samples.

We conducted extensive **hyperparameter searches** for the Logistic Regression with both TF-IDF or Sentence-BERT. The exact hyperparameters tested are available in our hyperparameter search files in GitHub.[12] We did not conduct hyperparameter searches for training the BERT and BERT-NLI Transformers to save compute. Instead, we used the recommended hyperparameters from (Laurer et al., 2023) for both the first and second analysis.

- For BERT-base, the hyperparameters are: 'lr_scheduler_type': 'constant', 'learning_rate': 2e-5, 'warmup_ratio': 0.06, 'seed': 42, 'per_device_train_batch_size': 32, 'weight_decay': 0.05, 'num_train_epochs': 50,
- For BERT-NLI, the hyperparameters are: 'lr_scheduler_type': linear, 'learning_rate': 2e-5, 'warmup_ratio': 0.40, 'seed': 42, 'per_device_train_batch_size': 32, 'weight_decay': 0.05, 'num_train_epochs': 20.
- Note that after initial tests, we dynamically adapted the number of epochs, as the training time would have otherwise varied widely depending on the different data augmentation strategies (ranging from only 500 texts to 24500 texts). The number of epochs indicated above are therefore reduced as the amount of data grows.[13]

[12] https://github.com/MoritzLaurer/language-barrier-multilingual-transfer

[13] https://github.com/MoritzLaurer/language-barrier-multilingual-transfer/blob/f5488b5b738a79adcf3d2626f0ad06ab964d5b91/analysis-transf-run-a1.py#L405

Another important hyperparameter for **BERT-NLI** are the hypotheses against which each text is tested. We used the same text formatting strategy as in (Laurer et al., 2023). The tables below display the hypotheses used for the two datasets.

*Table 14 - Hypotheses used with BERT-NLI for Manifesto-8 dataset*

| Label | Hypothesis |
|---|---|
| Economy | The quote is about topics like economy, or technology, or infrastructure, or free market |
| External Relations | The quote is about topics like international relations, or foreign policy, or military |
| Fabric of Society | The quote is about topics like law and order, or multiculturalism, or national way of life, or traditional morality |
| Freedom and Democracy | The quote is about topics like democracy, or freedom, or human rights, or constitutionalism |
| Political System | The quote is about topics like governmental efficiency, or political authority, or decentralisation, or corruption |
| Social Groups | The quote is about topics like agriculture, or social groups, or labour groups, or minorities |
| Welfare and Quality of Life | The quote is about topics like welfare, or education, or environment, or equality, or culture |
| No other category applies | The quote is about something other than the topics economy, international relations, society, freedom and democracy, political system, social groups, welfare. It is about none of these topics |

*Table 15 - Hypotheses used with BERT-NLI for PImPo dataset*

| Label | Hypothesis |
|---|---|
| Immigration neutral | The quote describes immigration neutrally without implied value judgement or describes the status quo of immigration, for example only stating facts or using technocratic language about immigration |
| Immigration sceptical | The quote describes immigration sceptically / disapprovingly. For example, the quote could mention the costs of immigration, be against migrant workers, state that foreign labour decreases natives' wages, that there are already enough refugees, refugees are actually economic migrants, be in favour of stricter immigration controls, exceptions to the freedom of movement in the EU. |
| Immigration supportive | The quote describes immigration favourably / supportively. For example, the quote could mention the benefits of immigration, the need for migrant workers, international obligations to take in refugees, protection of human rights, in favour of family reunification or freedom of movement in the EU. |
| no_topic | The quote is not about immigration. |

# Bibliography

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *ArXiv:1810.04805 [Cs]*. Retrieved from http://arxiv.org/abs/1810.04805

Fan, A., Bhosale, S., Schwenk, H., Ma, Z., El-Kishky, A., Goyal, S., … Joulin, A. (2020, October 21). *Beyond English-Centric Multilingual Machine Translation*. arXiv. https://doi.org/10.48550/arXiv.2010.11125

He, P., Gao, J., & Chen, W. (2021). DeBERTaV3: Improving DeBERTa using ELECTRA-Style Pre-Training with Gradient-Disentangled Embedding Sharing. *ArXiv:2111.09543 [Cs]*. Retrieved from http://arxiv.org/abs/2111.09543

Laurer, M., Van Atteveldt, W., Casas, A. S., & Welbers, K. (2023). Less Annotating, More Classifying – Addressing the Data Scarcity Issue of Supervised Machine Learning with Deep Transfer Learning and BERT - NLI. *Political Analysis*. Retrieved from https://osf.io/74b8k

Miller, B., Linder, F., & Mebane, W. R. (2020). Active Learning Approaches for Labeling Text: Review and Assessment of the Performance of Active Learning Approaches. *Political Analysis*, *28*(4), 532–551. https://doi.org/10.1017/pan.2020.4

Montani, I., Honnibal, M., Honnibal, M., Van Landeghem, S., Boyd, A., Peters, H., … Baumgartner, P. (2022). *explosion/spaCy: V3.2.4*. Zenodo. https://doi.org/10.5281/ZENODO.6394862

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*(85), 2825–2830.

Reimers, N., & Gurevych, I. (2019, August 27). *Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks*. arXiv. Retrieved from http://arxiv.org/abs/1908.10084

Reimers, N., & Gurevych, I. (2020, October 5). *Making Monolingual Sentence Embeddings Multilingual using Knowledge Distillation*. arXiv. https://doi.org/10.48550/arXiv.2004.09813

Zobel, M., & Lehmann, P. (2018). Positions and saliency of immigration in party manifestos: A novel dataset using crowd coding: POSITIONS AND SALIENCY OF IMMIGRATION IN PARTY MANIFESTOS. *European Journal of Political Research*, *57*(4), 1056–1083. https://doi.org/10.1111/1475-6765.12266