Deep learning models for multilingual supervised political text classification

Draft paper for discussion at COMPTEXT; Dublin, 7 May 2022

Tom Nicholls* Pepper D. Culpepper†

Abstract

Comparative computational research in politics is frequently based on large corpora of multilingual news or political speech. A common approach to handling the multiple-language issue is to machine translate to English before downstream modelling; this works well in many cases, but adds an extra step of introduced error. The cost of translation via the DeepL or Google Translate APIs is also high for large datasets. We present a method for supervised classification of large multilingual datasets, using a pre-trained multilingual transformer model.

We fine-tune an XLM-RoBERTA textual model on a large unlabelled corpus, combine it with a final softmax layer for probability estimation of category membership, then train and validate the resulting model with hand-labeled data. Non-English texts are handled directly without producing an intermediate translated representation. We validate the method by analysing a large (N > 1M) corpus of news articles on banking written in English, French, and German.

The classifications investigate aspects of the politics of post-financial crisis banking regulation, are theoretically-informed, and have complex decision boundaries. Results are compared to a conventional machine translation plus Support Vector Machine computational approach, in this case using the publicly available Opus-MT translation model running on local hardware.

Keywords: text-as-data; supervised classification; transformers; deep learning

1 Introduction

Comparative computational research in politics is frequently based on large corpora of multilingual news or political speech. This creates challenges for the researcher, as the standard computational toolkit is designed to work on texts in a common language. With many important questions being cross-national,

*Department of Communication and Media, University of Liverpool; T.E.Nicholls@liverpool.ac.uk

†Blavatnik School of Government, University of Oxford

and therefore often cross-lingual, however, there is a need to overcome these divisions.

One traditional option is to construct parallel models in each language, with the analyst reading the results across. Separate topic models could be produced for each language, for example, and similar topics combined by the researcher on the grounds that they are close enough to group together. This is somewhat unsatisfactory, both because of the degree of subjectivity involved in interpretation (are these two estimates actually of the same construct?) and also in some cases because the analyses, being run separately, cannot draw statistical strength from each other.

There is now an extensive body of work on the translation of texts to a common language (normally English) followed by monolingual analysis (e.g. Courtney et al., 2020; de Vries et al., 2018; Lucas et al., 2015; Reber, 2019). Downstream analyses using bag-of-words modelling discard the structure of the text in any case, and as long as the distribution of words produced by the machine translation systems is reasonably close to those produced by a native speaker covering the same material then we might hope the translated texts will analyse successfully. And this is, in general, what is found.

A significant practical difficulty with this approach for very large datasets, however, is that the well-studied approach relies on machine translation via a state-of-the-art cloud provider, usually Google Translate or DeepL. This is not a viable strategy for the many researchers with more data than funding, given a cost of around $/€20 per million characters for API access on these services[1].

Publicly available machine translation models exist (e.g. Fan et al., 2021; Tiedemann & Thottingal, 2020) and can be run locally. These produce poorer quality translations than the commercial product but, assuming the availability of sufficient hardware (such as a university HPC system) have the great advantage of cost and scalability. What is not clear is whether they produce results which are reliable enough to be used for substantive research. In simply reading these models' output it is clear that they produce less accurate translations than a human translator or a state-of-the-art commercial model. It is not clear whether, once converted to a mere bag-of-words, they are good enough. If they are, this opens the door to cheaply machine translating foreign language texts and then using our well-honed conventional computational toolkit.

A final option for some research designs has recently become feasible with the introduction of multilingual embedding models trained on large parallel and multilingual corpora: that of using the original language documents directly, relying on the model to convert the input texts into a common vector space. These can then produce data for downstream analysis either by using the embeddings as a feature representation for e.g. a classifier, or by training the multilingual model itself to solve the downstream task. This latter, for example, is an option for supervised classification: by adding an additional output layer to a

[1]There are various workarounds to automate translation without an account by abusing the web interface; these are in breach of the terms of service and also somewhat fragile.

transformer-type deep learning model and then training the model on labeled data then classification probabilities for arbitrary input documents can be obtained directly.

There are three obvious sources of error in multilingual supervised classification that do not exist in single-language work. Firstly, that the *translation process* itself can introduce errors into the modelling, as the idiosyncracies of the translation tool itself change the representation of text. Secondly, that the *way in which ideas or events are expressed* will vary by language, in the same way that we would expect there to be variations in textual representation across countries, across speakers, and across audiences. Thirdly, that the *ideas or events themselves* will be different in a similar way. In our banking case, for example, the particular scandals focused on in each country are different, and the ways that German and Swiss German news media conceptualise 'scandal' is quite distinct from that of the US and UK media for example.

The second and third of these sources of error is likely to be similar for the latter two approaches, but the first is not. Machine translation is an inherently more complicated process than multilingual embedding; in addition to producing a language-independent intermediate representation of textual content, there is then the further error-introducing stage of trying to render it in the target language. There is some reason to hope, therefore, that classifying directly on a vector representation of a multilingual document embedding will be a less error-prone process, and we will test that hope here.

Supervised classification, as an extremely common computational social science problem, is the motivating problem for this paper. We are engaged in a large multinational research project for which the supervised identification of media frames in news coverage across languages is important.

In this paper, we comparatively test two approaches to supervised multilingual classification: local machine translation followed by a conventional supervised classifier, and a multilingual deep learning model with a classifier head. We outline the collection and preprocessing of our dataset, the methodological choices made for our two modelling approaches, and comparatively analyse the results of the two strategies. Finally, we offer observations and comments for the wider use of these methods in the future.

2 Data

We use a large dataset (N=1,094,077) of news articles on banks and banking from six countries (Australia, France, Germany, Switzerland, UK, and US) published between 2007 and 2018. These data are substantively interesting for us, as researchers interested in the political economy of banking regulation, but also have a number of useful features for comparative analysis of multilingual methods. Firstly, the issues raised by international finance are complex, as is the way coverage of the topic is framed in different newspapers. Consequently, the problem is decidedly non-trivial. Secondly, the problem is genuinely international, as the banking system is globally connected and many of the regulatory challenges are handled with international cooperation. Consequently, we can have reasonable confidence that it *is* sensible to attempt classification of content across countries and languages.

We have collected what we consider to be the most important newspapers in each target country. As the structure of each of the media markets are different, the criteria were different for each, but the sample generally contains the largest and most important national newspapers (including tabloid, broadsheet, and specialist business dailies where possible).

Table 1: Newspaper sources

Country	Newspapers	Source
Australia	The Advertiser/Sunday Mail	LexisNexis
Australia	Australian Financial Review	LexisNexis
Australia	The Australian	LexisNexis
Australia	Canberra Times	LexisNexis
Australia	The Courier Mail/Sunday Mail	LexisNexis
Australia	Herald Sun/Sunday Herald Sun	LexisNexis
Australia	Sydney Morning Herald	LexisNexis
Australia	Daily Telegraph/Sunday Telegraph	LexisNexis
Australia	The Age	LexisNexis
Australia	The West Australian	LexisNexis
France	Le Figaro	LexisNexis
France	Le Monde	LexisNexis
France	Libération	Web crawl
France	Les Echos	LexisNexis
France	20 minutes	Web crawl
Germany	Bild	Web crawl
Germany	Handelsblatt	LexisNexis
Germany	Süddeutsche Zeitung	LexisNexis
Germany	Tageszeitung	LexisNexis
Germany	Frankfurter Allgemeine Zeitung	Web crawl
Germany	Die Welt	LexisNexis/Factiva
Switzerland (French)	20 minutes	Web crawl
Switzerland (French)	24 heures	LexisNexis
Switzerland (French)	Le Matin	LexisNexis
Switzerland (French)	La Tribune de Genève	LexisNexis

Country	Newspapers	Source
Switzerland (German)	20 minuten	Web crawl
Switzerland (German)	Neuer Zürcher Zeitung	LexisNexis
Switzerland (German)	Tages-Anzeiger	LexisNexis
Switzerland (German)	Blick	LexisNexis/Factiva
United Kingdom	Financial Times	LexisNexis
United Kingdom	The Guardian	LexisNexis
United Kingdom	Daily Mail/Mail on Sunday	LexisNexis
United Kingdom	The Mirror/Sunday Mirror	LexisNexis
United Kingdom	The Sun	LexisNexis
United Kingdom	Daily Telegraph	LexisNexis
United Kingdom	The Times	LexisNexis
United States	Chicago Tribune	ProQuest
United States	Los Angeles Times	ProQuest
United States	NY Daily News	LexisNexis
United States	NY Post	LexisNexis
United States	NY Times	LexisNexis
United States	USA Today	LexisNexis
United States	Wall Street Journal	Factiva
United States	Washington Post	LexisNexis

Most articles were sourced from LexisNexis. Where LexisNexis data were unavailable, other sources were used including the ProQuest and Factiva databases and by crawling the online archives of newspapers using the RISJbot (Nicholls, 2018). In all cases documents were been selected by keyword[2].

This kind of content selection has the typical disadvantage of being both wide and quite noisy. Many banking words are polysemous (e.g. 'bank' could mean a particular branch of a financial institution, the side of a river, or part of the name 'Sincil Bank', an English football stadium) and there are a few non-banking words which begin 'bank' (e.g. the artist Banksy, or bankdrücken – a German bench press).

Categories for classification were selected for substantive relevance, and are typical of the kinds of complex classifications that are often necessary in practical political communication work. In our case,

[2]Containing at least two words with the stem bank!. In French a similar search was carried out, bank! OR banque OR bancaire OR banquier OR banquière OR financier OR financière. In German, the extensive use of compound nouns required a more complex search strategy, with search terms developed based on analysis of the Bild and Zwanzig Minuten web crawl data: bank! OR bundesbank! OR citibank! OR commerzbank! OR dekabank! OR deutschebank! OR direktbank! OR geschäftsbank! OR genossenschaftsbank! OR gierbank! OR hypovereinsbank OR immobilienbank OR investitionsbank OR investmentbank! OR krisenbank! OR landesbank! OR mittelstandsbank! OR nationalbank! OR nordbank! OR onlinebank! OR pfandbriefbank OR postbank OR privatbank! OR raiffeisenbank! OR targobank OR volksbank! OR zentralbank! OR nord lb OR bayernlb OR consorsbank OR comdirect OR kantonalbank!

they represent different media frames about commercial banks, bankers, and banking that occur in the aftermath of the 2008 financial crisis. For each of the categories, the decision boundaries are complex, with nuances and exclusions, and are thus a challenge for automated classification. These categories were derived inductively by looking at unsupervised categorizations of front-page stories, aggregating the stories about banks into politically meaningful categories. In each of the categories we selected for analysis, the banks were protagonists, not examples or illustrative cases. Classification is multi-membership, in that articles can contain zero of these frames, one, or multiple; consequently, we fit separate models for each category rather than a single multi-category classifier.

The categories, and a simplified summary of what is contained in each, are in the table below.

Table 2: Banking article classifications

Classification	Definition
Scandal	Named banks (allegedly) involved in, or allegedly involved in, *misconduct* that involved settlement/charge/fine/claim etc., or a sense that banks are *ripping off* customers.
Regulation	The making of rules for banks or the regulatory or political oversight of banks
Business as usual	The business of banking as a normal economic activity, including earnings, business strategy, the development of new products, and discussions of senior leaders in banks
Financial crisis/bailouts	The events of the financial crisis of 2008 and the bailouts that resulted from them (including the lingering effects years afterward)
Executive pay	The level of executive pay and bonuses, including shareholder votes on (or revolts about) remuneration of senior executives

3 Method

Data were processed and extensively cleaned in R and Python, using the tm (Feinerer et al., 2008) and quanteda (Benoit et al., 2018) frameworks. As LexisNexis and the other databases can contain multiple copies of each article (normally representing multiple regional editions or different recensions) we de-duplicated articles within each newspaper, using a cosine similarity of 0.65 over a standard bag-of-words vector model as the upper bound for treating articles as unique (see Boumans et al., 2018).

Articles are of wildly varying length, and some cover a range of different topics. Consequently, our unit of analysis is first the (extended) paragraph including the headline and standfirst rather than the whole of the article. We use a modified version of the algorithm used by Gilardi et al. (2020), combining short paragraphs where necessary and aiming for a document length as little as possible above 150 words.

Each of the French and German texts were machine translated to English using the fr-en and de-en pairs from Opus-MT (Tiedemann & Thottingal, 2020), via the EasyNMT library (Ubiquitous Knowledge Processing Lab, 2021).[3]

Because a keyword search on banking related terms includes a large number of articles which are non-bank related (or only feature banks and banking very slightly) a preliminary filter was carried out using a variation of the approach in Benoit et al. (2016): crowd-sourced labeling was used to identify relevant and non-relevant articles, and a linear SVM using tf-idf term features was used to identify and exclude non-relevant articles.[4] The resulting data corpus is a much cleaner representation of our subject of interest than the original collection, though there are still a number of non-banking articles which have slipped through to complicate our downstream analysis.

3.1 Labeling

For training, 850 random English documents were selected, together with 400 French and 400 German. For each category, articles were double-coded, with disagreements resolved by discussion. The training data was split into separate training/test (75%) and holdout (25%) samples.

As the category distributions in the data are very unbalanced, top-up samples of likely-positive articles were created for each category. After coding the random sample, a keyword analysis was carried out to identify features associated with positive membership in each category, and a set of English documents were selected for top-up coding using the same procedure as above (scandal: 113 documents; regulation, crisis/bailouts, and executive pay: 250 documents each; business as usual: no top-up). These labels were used exclusively for training/test data, and not for final validation.

3.2 Support Vector Machine classifier

For the conventional approach, a number of standard modelling approaches (including non-linear SVM, logistic regression, and random forests) were trialled, with the best performer being a standard linear SVM model. Two sets of linear Support Vector Machine classifiers were fit using sklearn (pedregosa_scikit-learn_2011). In each case the features used for the models were a conventional tf-idf weighted bag of words, with the hyperparameters chosen using an initial grid search over the data for the "scandal" frame. As the models are monolingual, the Opus-MT machine translated texts are used for

[3]This was carried out using a cluster of machines in the University of Oxford's ARC high-performance computing centre, using GPUs. It is beyond the sensible capacity of a single researcher's desktop computer, but fairly accessible to those with access to modest institutional HPC systems.

[4]Performance was validated using a separate sample of documents coded by the authors, with the French and German content processed in Opus-MT translation. In all cases the performance of both the crowd and the classifier were good but not stellar, with an F1 performance of around 0.8; a full analysis of the merits of crowd-sourced pre-filtering is outside the scope of this paper.

French and German documents.

The first set was trained using the English data only, and validated on the English, French (Opus-MT translated), and German (Opus-MT translated) holdout data separately. The second set was trained on the (Opus-MT translated) French and German training data too, and again validated on the holdout data.

The first set of models reflect the common position that a research team lacks language expertise for each of the countries that they would like to analyse a typical attempt to transfer existing single-language models to a multilingual context, but has obvious theoretical weaknesses. The second represents a more robust approach, though obviously the additional training data used increases the time required for labeling.

3.3 XLM-RoBERTa transformer model with final classifier layer

Transformer models are trained on extremely large corpora of text from sources such as the web. Nevertheless, they are known to perform better for downstream tasks if the pre-trained weights are fine-tuned, in an unsupervised manner, on texts which are specific to the domain and task (devlin_bert_2018).

The feature selection and classification parts of a traditional model are somewhat combined here: transformer models both generate an internal vector representation of the text and classify it (using the additional model layer). The complete model is then fitted using backpropogation and a gradient descent algorithm using the training and test data.

There are a number of pre-trained BERT examples, none of which are perfect for our application: they are trained on general English corpora rather than on news in general or banking news in particular. We have dealt with this in the standard way using *transfer learning*, taking a standard pre-trained model and fine tuning it for our own problem. This is more effective than using the model as-is, and *vastly* cheaper and easier than attempting to train one from scratch (which requires hundreds of gigabytes of training data and millions of dollars worth of computing time).

In our case, we fine-tuned the XLM-RoBERTA model (Conneau et al., 2020) cross-lingual model on all of the documents in our corpus, in their original languages (691,495 English, 178,603 French, and 223,979 German), using the Hugging Face processing framework in Python. Finally, we added a final softmax classification layer, and copied the resulting fine-tuned base classification model to allow it to be trained separately with labeled data for each classification task.

For better comparison with the SVM models, we have again trained two separate models for each classification task: one using the English training data only and the other including the French and German labeled data. In this case, all inputs are in their original languages, and the validation is also carried out on the untranslated holdout samples.

4 Results

We report two separate measures for each classifier. Both are versions of the conventional F1 measure, the harmonic mean of precision and recall. Weighted F1 considers the accuracy of the classification of each document (both positive and negative), taking into account the different numbers of documents in each class. It is useful where both categories are of interest, though the performace of the classifier on smaller classes can be overwhelmed by good performance on a large class where data are strongly unbalanced. Binary F1, on the other hand, is focused on the performance of the classifier for the *positive* class. For a classification problem where the positive class is smaller than the negative (such as here) this is a harder test. Which is more appropriate will depend on the nature of the problem being attempted: for selection tasks, where the aim is to identify wheat in a pile of chaff, the binary measure will best reflect underlying classifier performance. Where the aim is to put documents into two equally-important piles, the better will generally be the weighted measure.

Table 3: Validation set category membership

Language	N	Scandal	Regulation	Business as usual	Crisis/bailouts	Exec pay
English	210	31	61	83	30	6
French	200	31	33	89	47	7
German	200	26	58	91	48	7

Table 4: SVM classifiers, English-only training, weighted F1

Language	Scandal	Regulation	Business as usual	Crisis/bailouts	Exec pay
English	0.91	0.88	0.74	0.88	0.97
French	0.92	0.80	0.56	0.81	0.97
German	0.90	0.79	0.51	0.76	0.97

Table 5: SVM classifiers, All-language (translated) training, weighted F1

Language	Scandal	Regulation	Business as usual	Crisis/bailouts	Exec pay
English	0.94	0.88	0.79	0.90	0.96
French	0.89	0.87	0.80	0.84	0.97
German	0.90	0.80	0.72	0.76	0.97

Table 6: SVM classifiers, English-only training, binary F1

Language	Scandal	Regulation	Business as usual	Crisis/bailouts	Exec pay
English	0.68	0.77	0.65	0.55	0.44
French	0.71	0.55	0.33	0.57	0.44
German	0.55	0.67	0.26	0.43	0.44

Table 7: SVM classifiers, All-language (translated) training, binary F1

Language	Scandal	Regulation	Business as usual	Crisis/bailouts	Exec pay
English	0.78	0.78	0.74	0.64	0.25
French	0.63	0.62	0.75	0.59	0.44
German	0.56	0.62	0.64	0.41	0.44

For the SVM models, the baseline performance is generally reasonable when looking at the weighted measure, but poorer when looking at the performance on the (smaller) positive group using the binary measure. The advantage of using the additional (Opus-MT translated) French and German training data is not large; it helps performance on some tasks and hurts on others, with perhaps a slight reversion to the mean overall.

In general, performance is best on the frames with the clearest semantic markers (scandal and regulation), most challenging on executive pay (which is wildly unbalanced: only around 3% of articles in the sample are positive for this category) and most mediocre for business as usual (which co-occurs with many of the other frames and has the widest definition and the loosest boundaries).

Table 8: XLM-RoBERTa classifiers, English-only training, weighted F1

Language	Scandal	Regulation	Business as usual	Crisis/bailouts	Exec pay
English	0.94	0.90	0.81	0.90	0.96
French	0.94	0.83	0.84	0.84	0.97
German	0.91	0.83	0.79	0.82	0.97

Table 9: XLM-RoBERTa classifiers, All-language training, weighted F1 measure

Language	Scandal	Regulation	Business as usual	Crisis/bailouts	Exec pay
English	0.94	0.88	0.86	0.92	0.96
French	0.92	0.87	0.86	0.87	0.96
German	0.90	0.86	0.81	0.87	0.96

Table 10: XLM-RoBERTa classifiers, English-only training, binary F1

Language	Scandal	Regulation	Business as usual	Crisis/bailouts	Exec pay
English	0.68	0.77	0.78	0.66	0.48
French	0.71	0.55	0.81	0.62	0.63
German	0.55	0.67	0.77	0.58	0.59

Table 11: XLM-RoBERTa classifiers, All-language training, binary F1

Language	Scandal	Regulation	Business as usual	Crisis/bailouts	Exec pay
English	0.81	0.79	0.81	0.72	0.47
French	0.81	0.74	0.83	0.70	0.56
German	0.65	0.60	0.79	0.72	0.56

The XLM-RoBERTa classification model results are generally at least equivalent to those of the SVM. It has scored radically better on some of the hardest classification problems, especially the complex decision boundaries of business as usual and crisis/bailouts. Performance is also much better for executive pay, where the amount of training data is particularly low, and the additional gain from the all-language training is somewhat higher overall.

5 Conclusions

Both approaches have been effective for most tasks, and the English-only models perform better than expected. Although both models here are more computationally intensive than a simple monolingual bag-of-words, they offer opportunities for expanding the scope of research involving supervised classification approaches to cover multiple languages at relatively low cost.

The XLM-RoBERTa model has performed particularly well in this test. Offering classification performance beyond the SVM baseline on the English dataset alone, its ability to operate multilingually comes 'for free'. If researchers are able to work with each of the materials in their original language, and have no need to have their data translated for other research purposes, it has the additional advantage of being computationally cheaper than the full Opus-MT machine translation model. It may, on the other hand, perform less well with smaller datasets that provide less scope for fine-tuning the base model. XLM-RoBERTa handles a wide range of languages beyond English, French, and German, and other language models trained on particular language combinations are also available.

Work on transformers and other deep learning models is still progressing at a rapid pace and both Opus-MT and XLM-RoBERTa have already been joined by newer and more accurate language models. Nevertheless, the ecosystem is now robust enough that models are fairly straightforward to obtain, train, and use for inference, given suitable hardware. We endorse their wider use by computational social scientists.

6 Acknowledgements

We would like to thank Robin Hsieh and Clemence Hautefort for their research assistance, and to all those other colleagues who have given useful comments and advice during this research. We would also like to acknowledge the use of the University of Oxford Advanced Research Computing (ARC) facility in carrying out this work. http://dx.doi.org/10.5281/zenodo.22558

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 787887). This paper reflects only the authors' views; the European Research Council Executive Agency is not responsible for any use that may be made of the information it contains.

References

Benoit, K., Conway, D., Lauderdale, B. E., Laver, M., & Mikhaylov, S. (2016). Crowd-sourced Text Analysis: Reproducible and Agile Production of Political Data. *American Political Science Review*, *110*(2), 278–295. https://doi.org/10.1017/S0003055416000058

Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., & Matsuo, A. (2018). Quanteda: An R package for the quantitative analysis of textual data. *Journal of Open Source Software*, *3*(30), 774. https://doi.org/10.21105/joss.00774

Boumans, J., Trilling, D., Vliegenthart, R., & Boomgaarden, H. (2018). The agency makes the (online) news world go round: The impact of news agency content on print and online news. *International Journal of Communication*, *12*, 1768–1789. Retrieved 11/05/2018, from http://ijoc.org/index.php/ijoc/article/view/7109

Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., & Stoyanov, V. (2020). *Unsupervised Cross-lingual Representation Learning at Scale*. Retrieved 05/05/2022, from http://arxiv.org/abs/1911.02116

Courtney, M., Breen, M., McMenamin, I., & McNulty, G. (2020). Automatic translation, context, and supervised learning in comparative politics. *Journal of Information Technology & Politics*, *0*(0), 1–10. https://doi.org/10.1080/19331681.2020.1731245

de Vries, E., Schoonvelde, M., & Schumacher, G. (2018). No Longer Lost in Translation: Evidence that Google Translate Works for Comparative Bag-of-Words Text Applications. *Political Analysis*, *26*(4), 417–430. https://doi.org/10.1017/pan.2018.26

Fan, A., Bhosale, S., Schwenk, H., Ma, Z., El-Kishky, A., Goyal, S., Baines, M., Celebi, O., Wenzek, G., & Chaudhary, V. (2021). Beyond english-centric multilingual machine translation. *Journal of Machine Learning Research*, *22*(107), 1–48.

Feinerer, I., Hornik, K., & Meyer, D. (2008). Text mining infrastructure in R. *Journal of Statistical Software*, *25*(5), 1–54. http://www.jstatsoft.org/v25/i05/

Gilardi, F., Shipan, C. R., & Wüest, B. (2020). Policy Diffusion: The Issue-Definition Stage. *American Journal of Political Science*. https://doi.org/10.1111/ajps.12521

Lucas, C., Nielsen, R. A., Roberts, M. E., Stewart, B. M., Storer, A., & Tingley, D. (2015). Computer-assisted text analysis for comparative politics. *Political Analysis*, *23*(2), 254–277. https://doi.org/10.1093/pan/mpu019

Nicholls, T. (2018). *RISJbot: A scrapy project to extract the text and metadata of articles from news websites* (Version v1.1.0). https://doi.org/10.5281/zenodo.1341873

Reber, U. (2019). Overcoming Language Barriers: Assessing the Potential of Machine Translation and Topic Modeling for the Comparative Analysis of Multilingual Text Corpora. *Communication Methods and Measures*, *13*(2), 102–125. https://doi.org/10.1080/19312458.2018.1555798

Tiedemann, J., & Thottingal, S. (2020). OPUS-MT–Building open translation services for the World. *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation*, 479–480.

Ubiquitous Knowledge Processing Lab. (2021). *UKPLab/EasyNMT*. GitHub. Retrieved 27/04/2021, from https://github.com/UKPLab/EasyNMT