Using dictionaries allows researchers to scale texts in a rapid and efficient way. Given the dictionary is of high quality, the validity is high, where you measure precisely what you want to measure. Moreover, reliability is high as the measures are consistent across documents irrespective of how large your corpus is. High validity and reliability are great assets in a research process. Still, we are sometimes limited in our ability to study global phenomena that span across different countries with different languages. Off-the-shelf cross-lingual dictionaries that allow for the comparison of particular dimensions across texts written in various languages are rare. Creating your own cross-lingual dictionary is great, but often limited due to language capacities within the research team. Also budgetary constraints prevent researchers from developing their own dictionary. In this paper, I present an approach that requires researchers to master only one language of the corpus. Word-embeddings and automatic translations form the backbone of the approach. I test the method using two different corpora: One, I measure climate change reporting in a corpus of over 103 3617 newspaper articles on climate change. The corpus covers articles in English, German, French, and Spanish. Two, I scale texts according their EU-integration sentiments in 15 languages in a corpus of XXX constitutional court decisions. The method presented here produces valid results on the dimensions of interests in both corpora. This gives confidence that we can analyse multilingual corpora despite mastering only one language and working with a limited budget.

**Scaling multilingual political and legal corpora when mastering only one language: A how-to for cross-language dictionary creation**

Using dictionaries allows researchers to scale texts in a rapid and efficient way. Given the dictionary is of high quality, the validity is high, where you measure precisely what you want to measure. Moreover, reliability is high as the measures are consistent across documents irrespective of how large your corpus is. High validity and reliability are great assets in a research process. Still, we are sometimes limited in our ability to study global phenomena that span across different countries with different languages. Off-the-shelf cross-lingual dictionaries that allow for the comparison of particular dimensions across texts written in various languages are rare. Creating your own cross-lingual dictionary is great, but often limited due to language capacities within the research team. Also budgetary constraints prevent researchers from developing their own dictionary. In this paper, I present an approach that requires researchers to master only one language of the corpus. Word-embeddings and automatic translations form the backbone of the approach. I test the method using two different corpora: One, I measure climate change reporting in a corpus of over 103 3617 newspaper articles on climate change. The corpus covers articles in English, German, French, and Spanish. Two, I scale texts according their EU-integration sentiments in 15 languages in a corpus of XXX constitutional court decisions. The method presented here produces valid results on the dimensions of interests in both corpora. This gives confidence that we can analyse multilingual corpora despite mastering only one language and working with a limited budget.

**Lisa Lechner** is Assistant professor for methods and methodology in political science at the University of Innsbruck. In her research, Lisa studies international treaties, such as trade agreements, bilateral tax treaties, and environmental agreements, as well as national and international jurisdictions by dint of inferential network- and quantitative text-analysis.