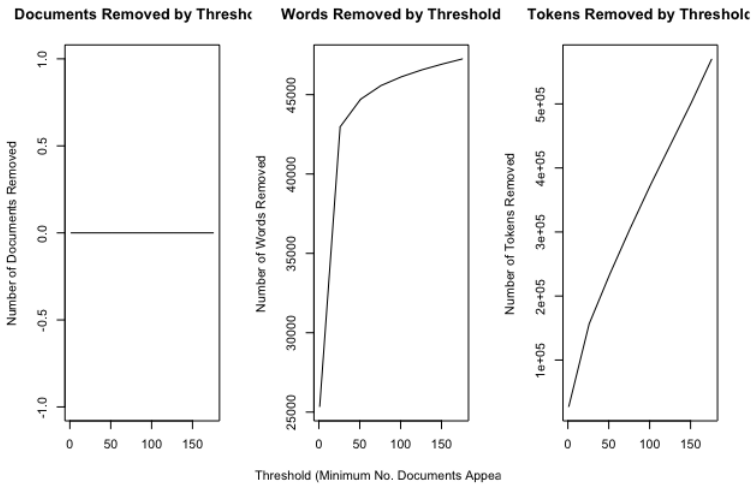
Lab Notes

Scott Patterson

April 30, 2020

1 Update: 30 April 2020

This week, I ran 9 topical content models using a binary EU covariate. The models varied on two dimensions: stop-word threshold and k value, or the number of topics. The '0 threshold' models (t0) do not filter words that appear less than 10 times. It is possible that infrequently occurring words may be important, and these highly sensitive models may pick them up. On the other hand, we may end up including concealing more important words in the process. Computing constraints made it infeasible to fit models at the 30,40, and 50 threshold range, but doing so may also yield insights. My reasoning here is based on the middle figure below, which shows that the 'Words Removed' plot begins to level around a threshold of 50. My suspicion is that our threshold will be around 20, as this is where we see the sharpest increase in removed words. Once we restrict our threshold range, replicating future models will be less computationally intensive.



The k values for this iteration are 12,18, and 24. It will be interesting to see how topics from high k values cluster into more consolidated topics in the lower k models.

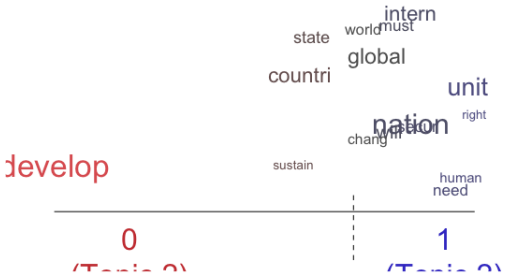
2 Interpretation

The topics can be found at the following link: I will interpret these data in the following manner:

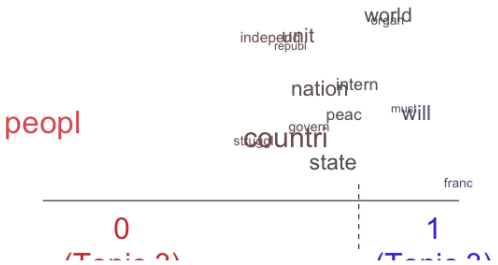
1. Examine the defining words for each topic; compare the topics with similar K values across the different threshold values
2. Examine how topics change across different K values within the same threshold.
3. Examine the Topical Content covariate visualizations (below). These figures demonstrate differences in word choice within topics across the binary EU covariate, where 0 = non-EU states and 1 = EU states. Words at the extremes are more unique to their respective pole. The sides of the word depicts its frequency of use. Words at the dividing line are shared across the covariate.

Figure 1: Topical Content: Thresh = 0, K = 12, Covariate = EU (binary)

(a) Topic 1



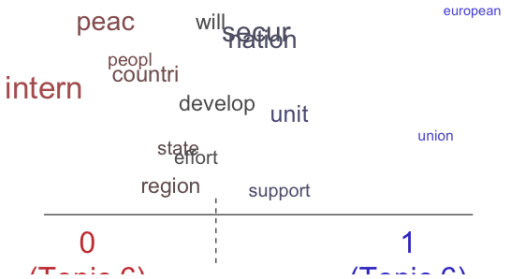
(b) Topic 2



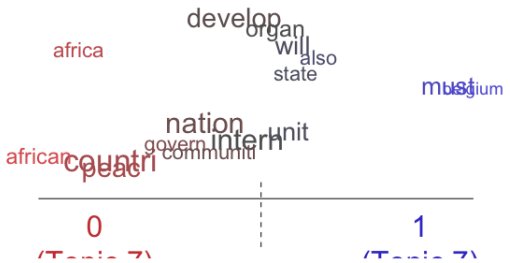
(c) Topic 3

(d) Topic 4

(e) Topic 5



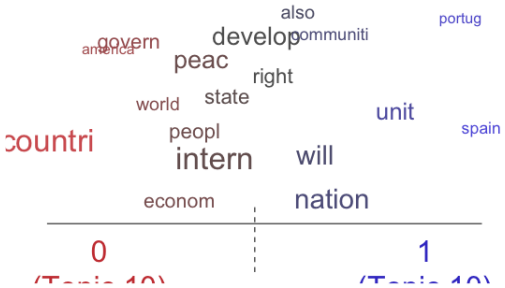
(f) Topic 6



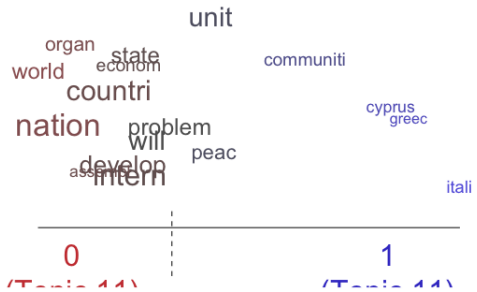
(g) Topic 7

(h) Topic 8

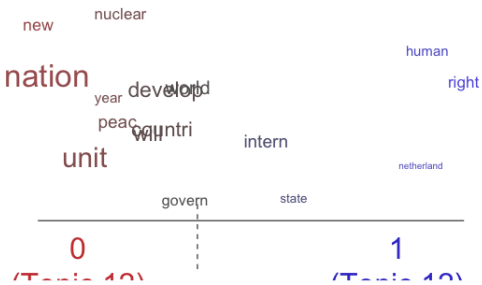
(i) Topic 9



(j) Topic 10



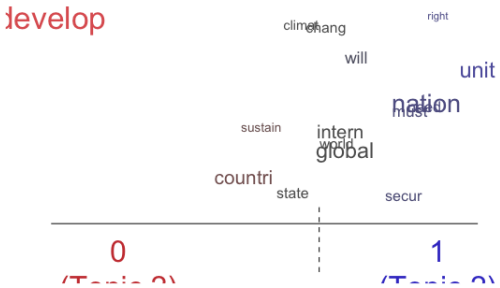
(k) Topic 11



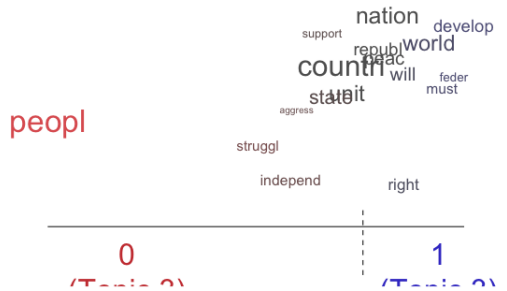
(l) Topic 12

Figure 2: Topical Content: Thresh = 0, K = 18, Covariate = EU (binary)

(a) Topic 1



(b) Topic 2

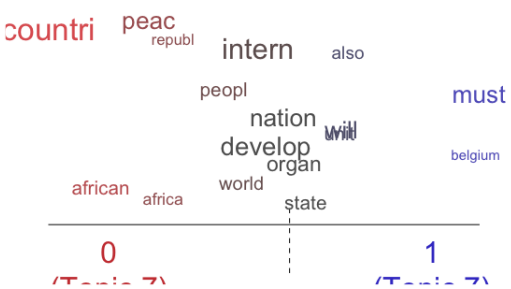


(c) Topic 3

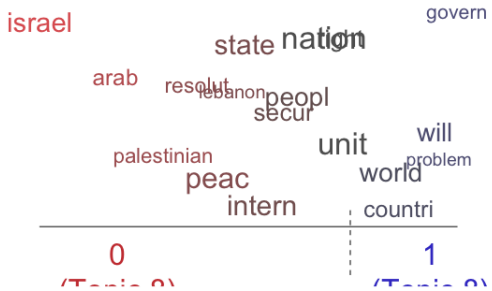
(d) Topic 4

(e) Topic 5

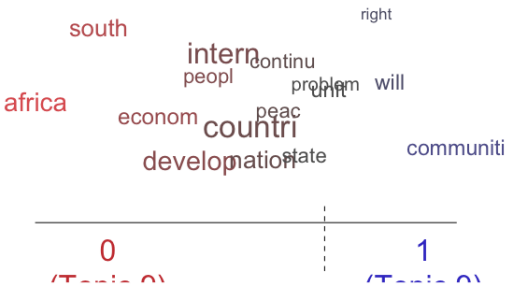
(f) Topic 6



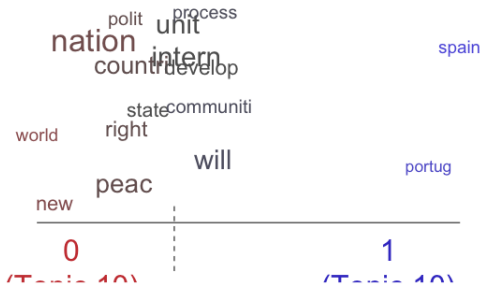
(g) Topic 7



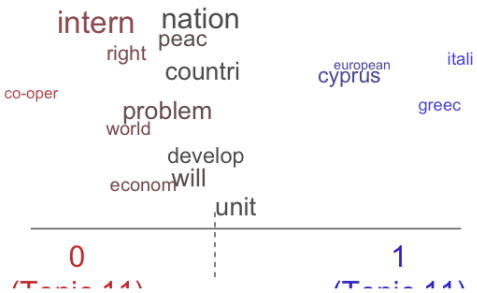
(h) Topic 8



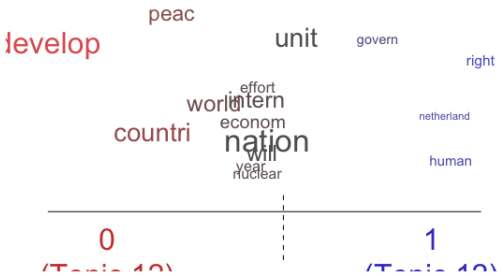
(i) Topic 9



(j) Topic 10

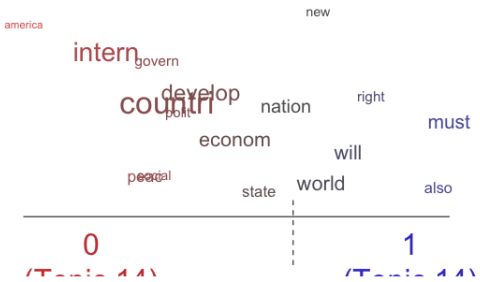


(k) Topic 11



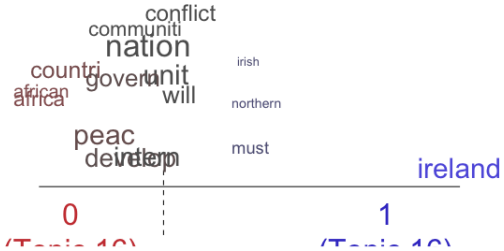
(l) Topic 12

(m) Topic 13

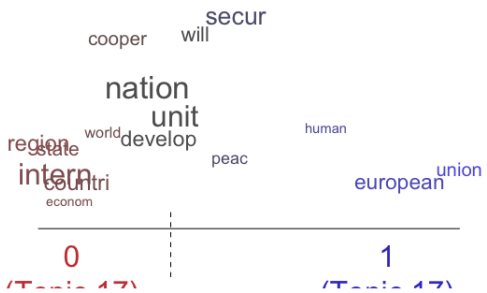


(n) Topic 14

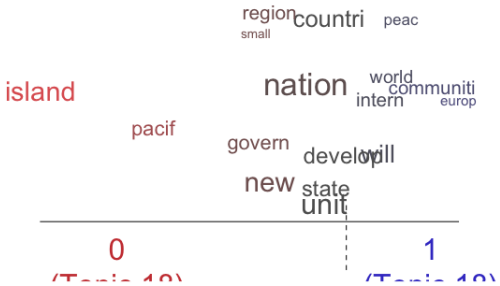
(o) Topic 15



(p) Topic 16

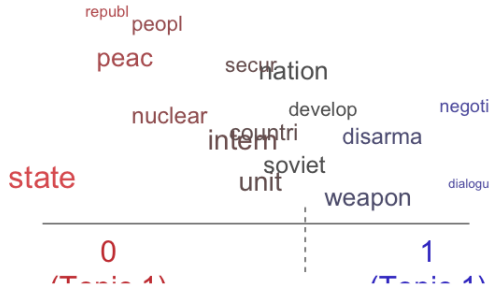


(q) Topic 17

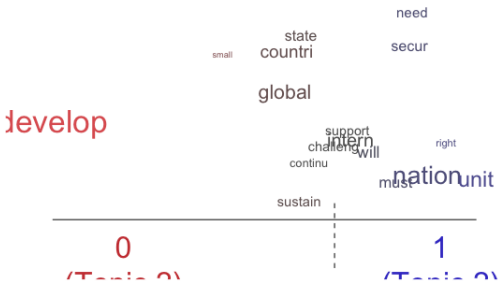


(r) Topic 18

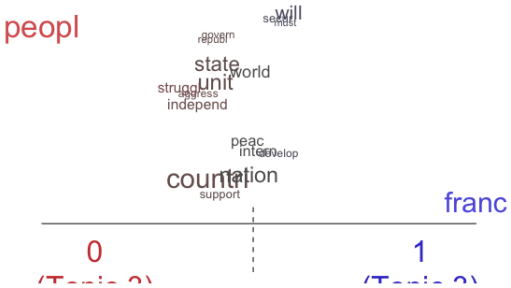
Figure 3: Topical Content: Thresh = 0, K = 24, Covariate = EU (binary)



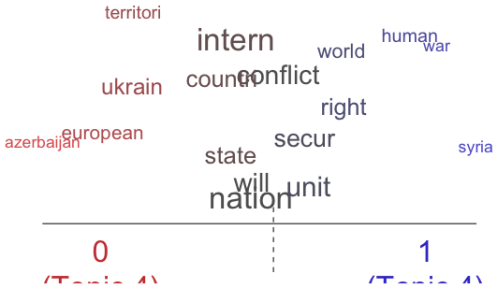
(a) Topic 1



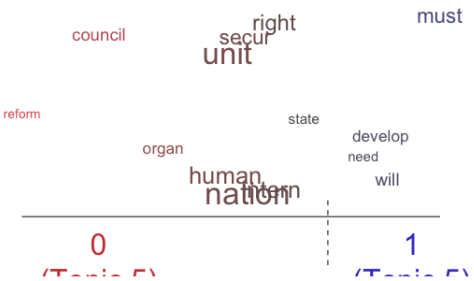
(b) Topic 2



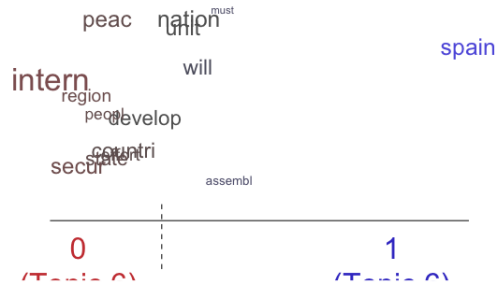
(c) Topic 3



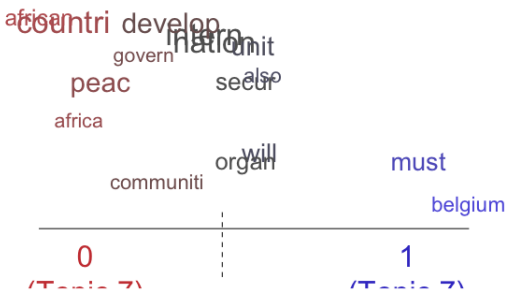
(d) Topic 4



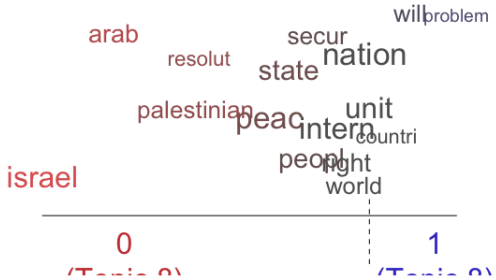
(e) Topic 5



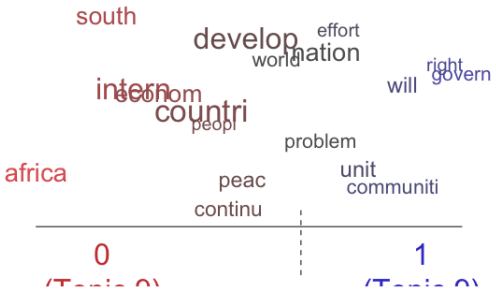
(f) Topic 6



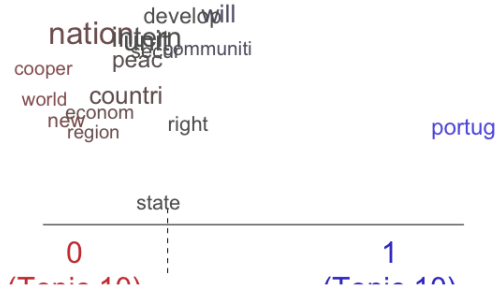
(g) Topic 7



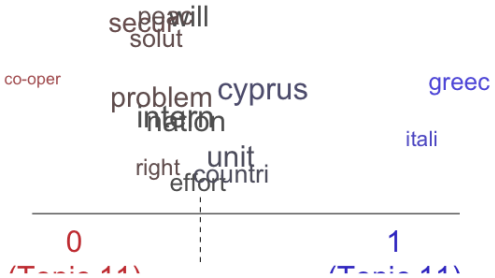
(h) Topic 8



(i) Topic 9



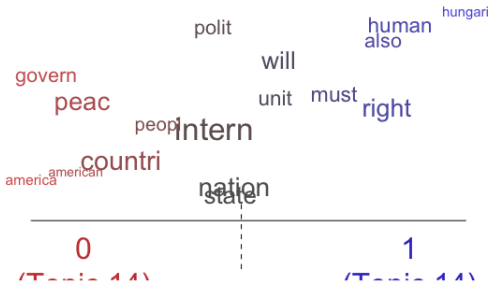
(j) Topic 10



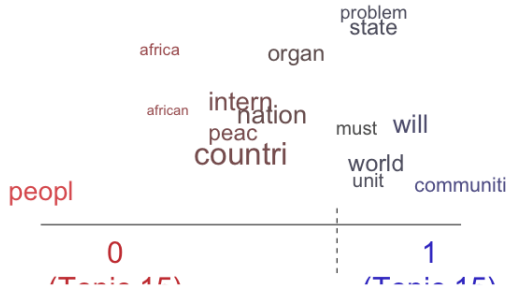
(k) Topic 11

(l) Topic 12

(m) Topic 13

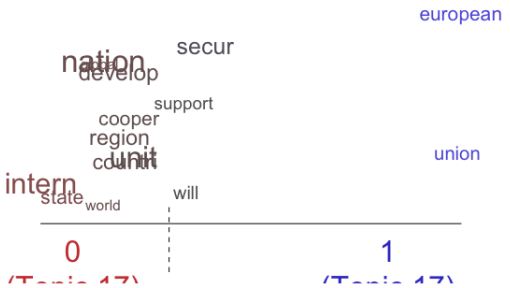


(n) Topic 14



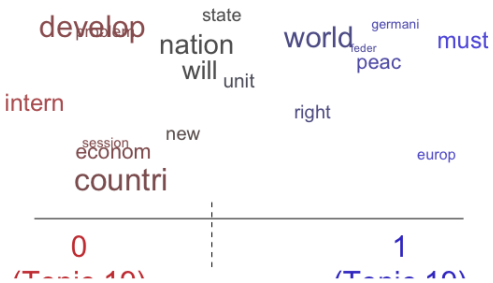
(o) Topic 15

(p) Topic 16



(q) Topic 17

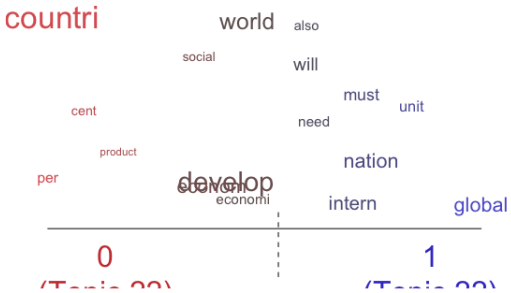
(r) Topic 18



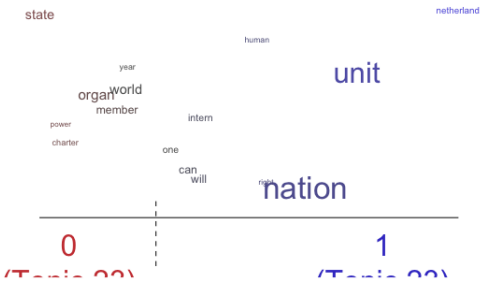
(s) Topic 19

(t) Topic 20

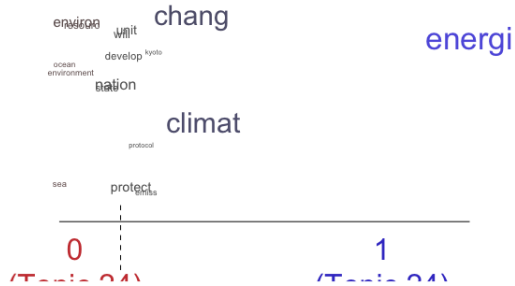
(u) Topic 21



(v) Topic 22



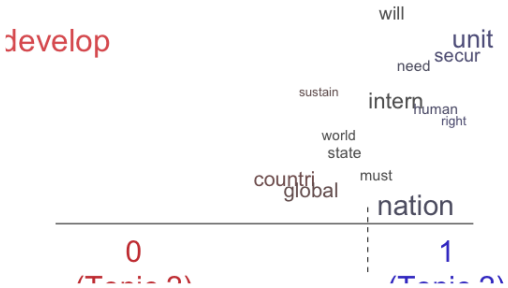
(w) Topic 23



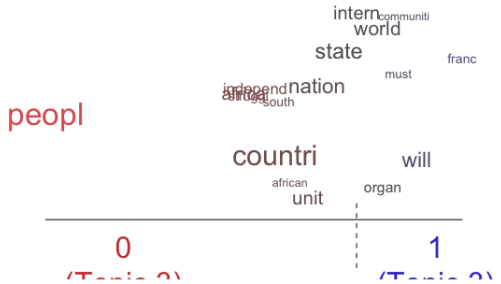
(x) Topic 24

Figure 4: Topical Content: Thresh = 10, K = 12, Covariate = EU (binary)

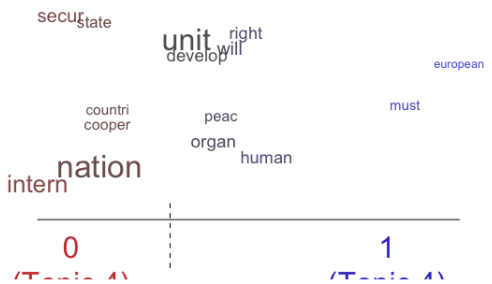
(a) Topic 1



(b) Topic 2



(c) Topic 3

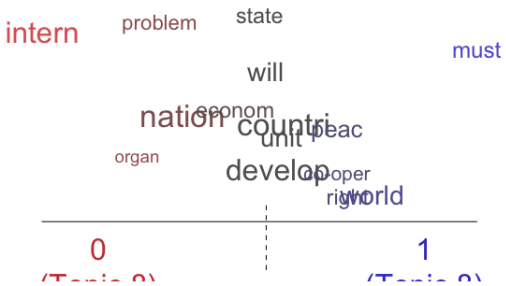


(d) Topic 4

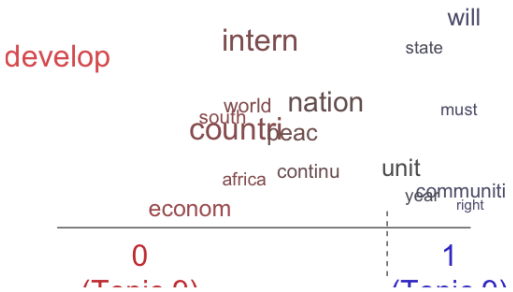
(e) Topic 5

(f) Topic 6

(g) Topic 7

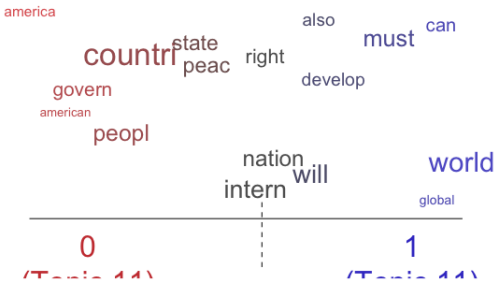


(h) Topic 8



(i) Topic 9

(j) Topic 10



(k) Topic 11

(l) Topic 12

2.1 t0k12contEU

2.2 t0k18contEU

2.3 t0k24contEU

2.4 t10k12contEU

2.5 t10k18contEU

2.6 t10k24contEU

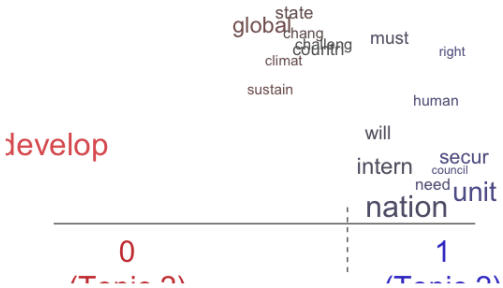
2.7 t20k12contEU

2.8 t20k18contEU

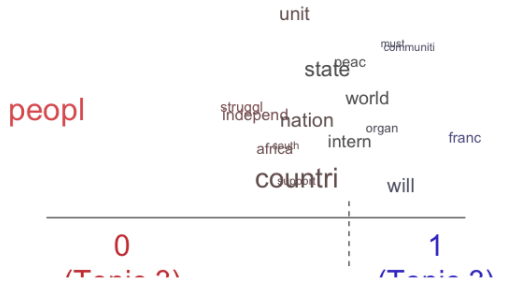
2.9 t20k24contEU

Figure 5: Topical Content: Thresh = 0, K = 18, Covariate = EU (binary)

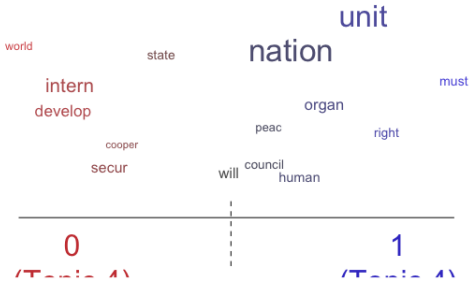
(a) Topic 1



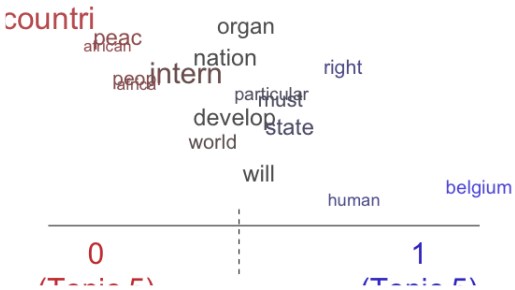
(b) Topic 2



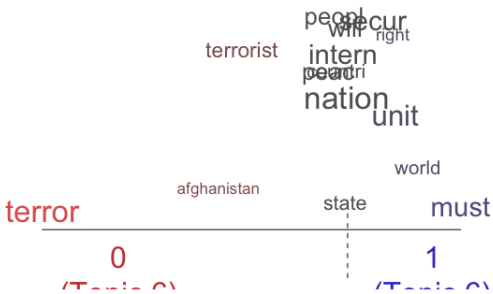
(c) Topic 3



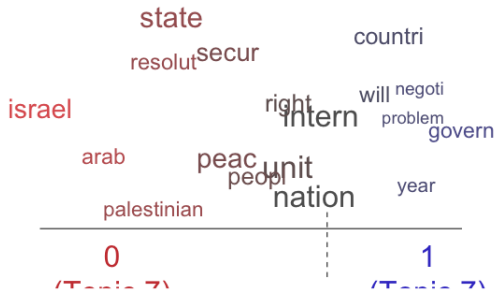
(d) Topic 4



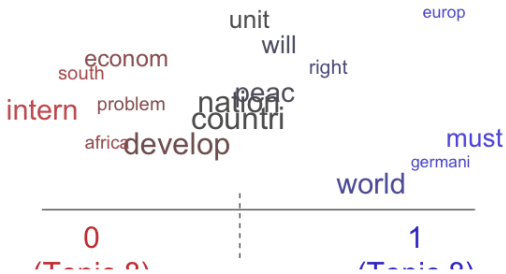
(e) Topic 5



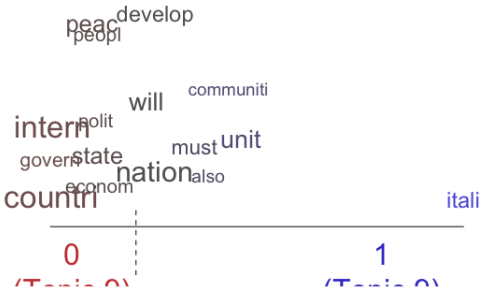
(f) Topic 6



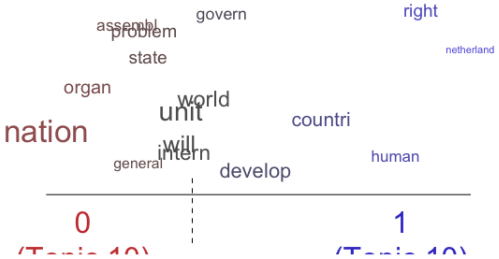
(g) Topic 7



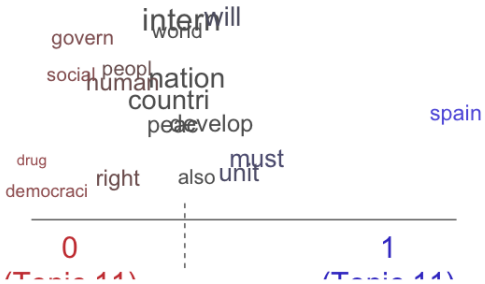
(h) Topic 8



(i) Topic 9

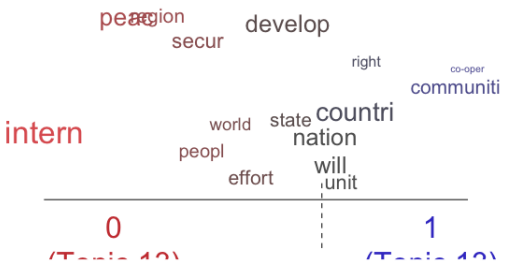


(j) Topic 10



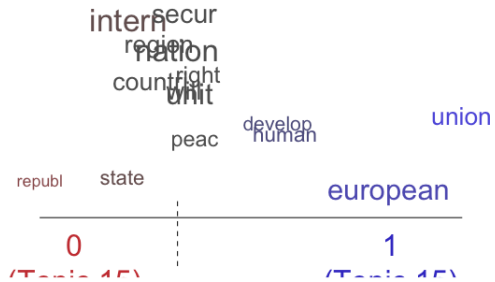
(k) Topic 11

(l) Topic 12

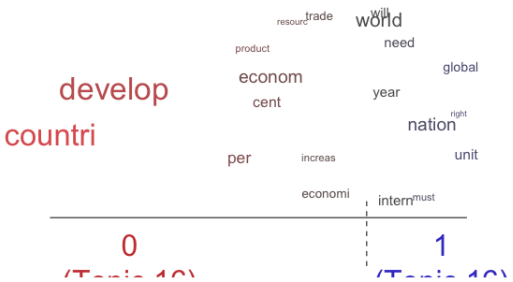


(m) Topic 13

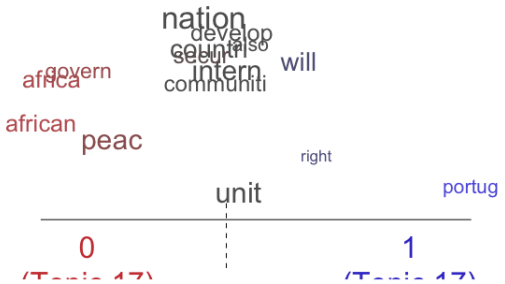
(n) Topic 14



(o) Topic 15



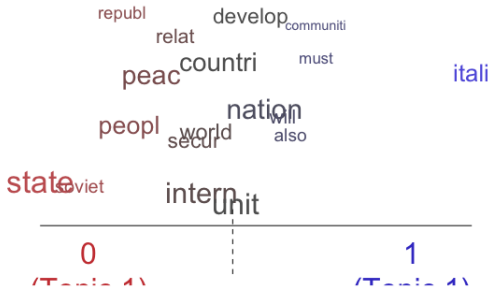
(p) Topic 16



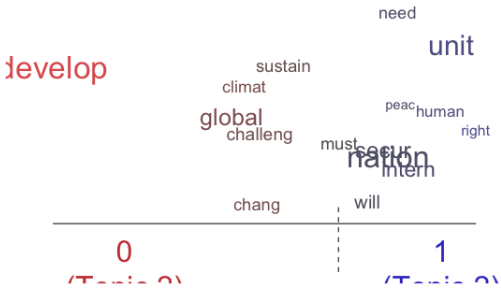
(q) Topic 17

(r) Topic 18

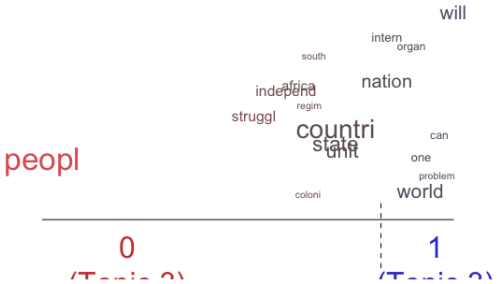
Figure 6: Topical Content: Thresh = 0, K = 24, Covariate = EU (binary)



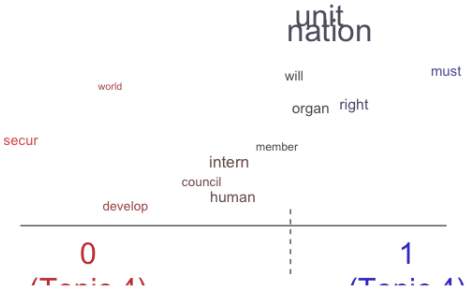
(a) Topic 1



(b) Topic 2



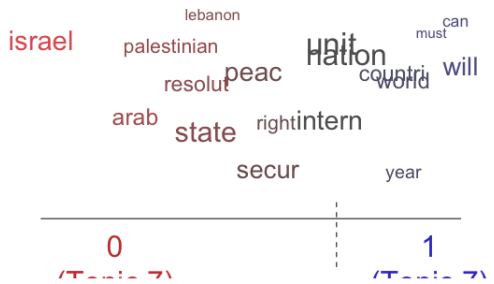
(c) Topic 3



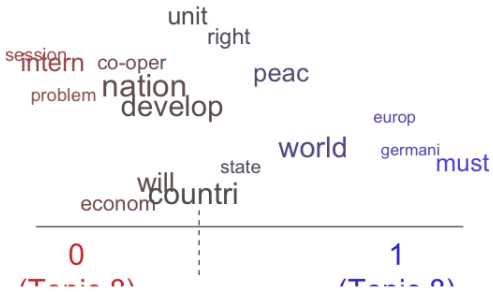
(d) Topic 4

(e) Topic 5

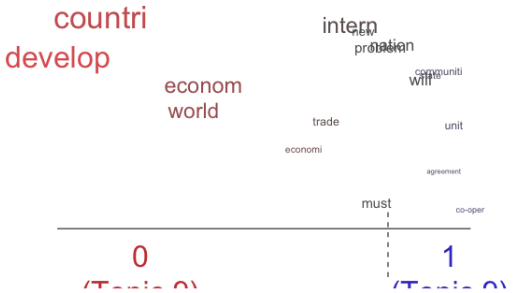
(f) Topic 6



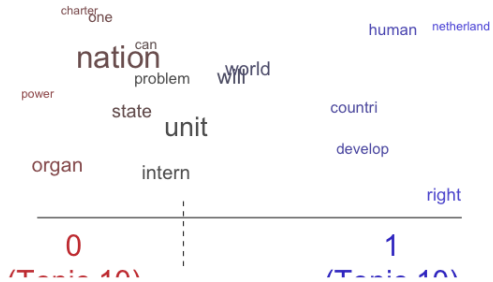
(g) Topic 7



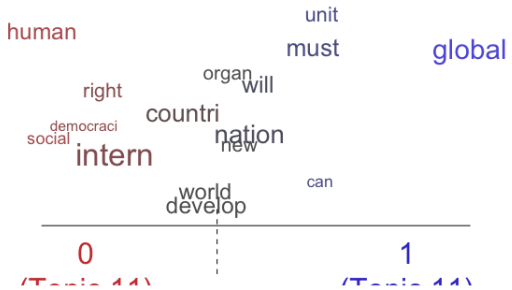
(h) Topic 8



(i) Topic 9

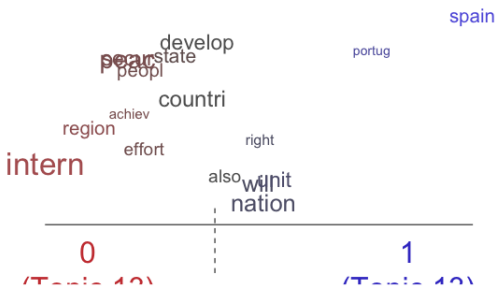


(j) Topic 10

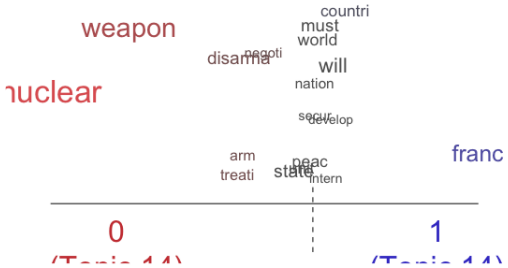


(k) Topic 11

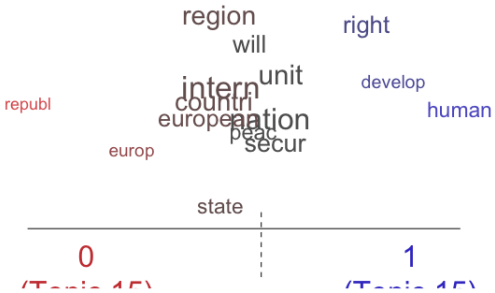
(l) Topic 12



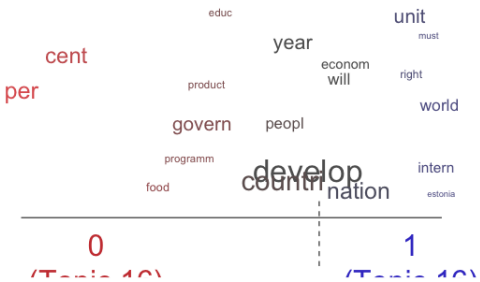
(m) Topic 13



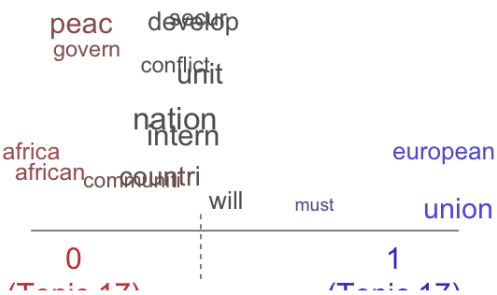
(n) Topic 14



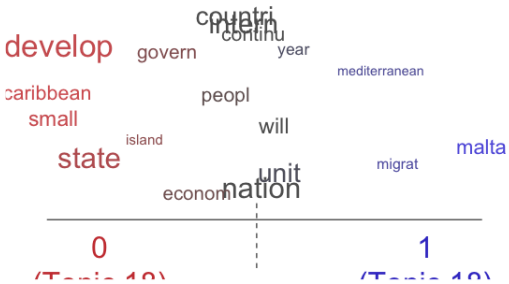
(o) Topic 15



(p) Topic 16



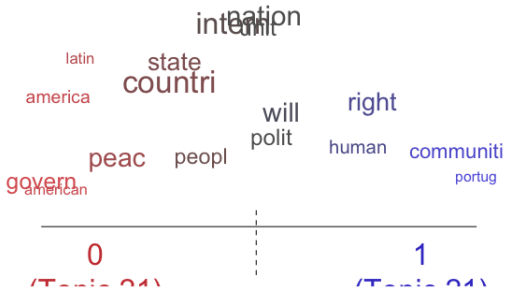
(q) Topic 17



(r) Topic 18

(s) Topic 19

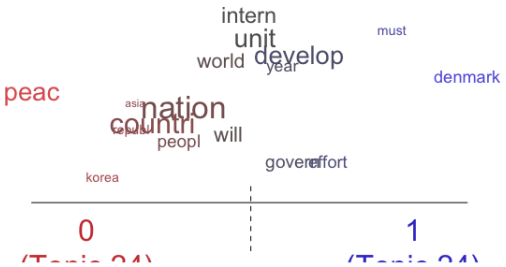
(t) Topic 20



(u) Topic 21

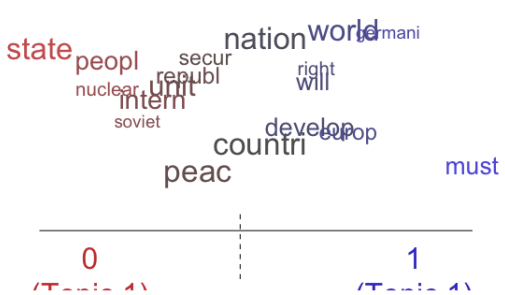
(v) Topic 22

(w) Topic 23

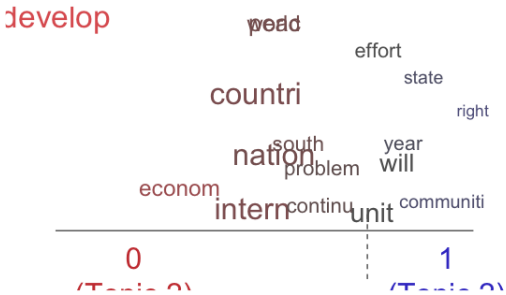


(x) Topic 24

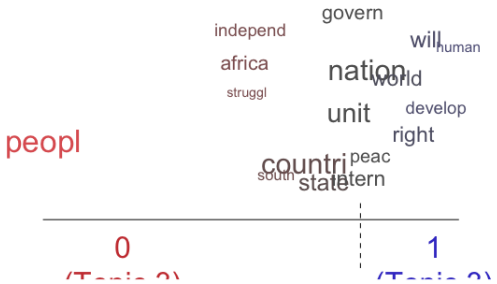
Figure 7: Topical Content: Thresh = 20, K = 12, Covariate = EU (binary)



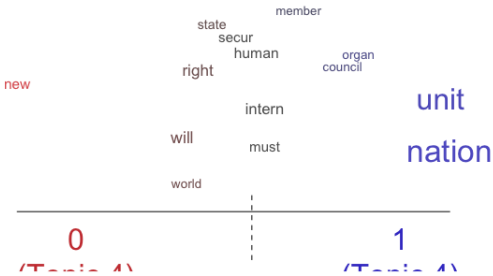
(a) Topic 1



(b) Topic 2

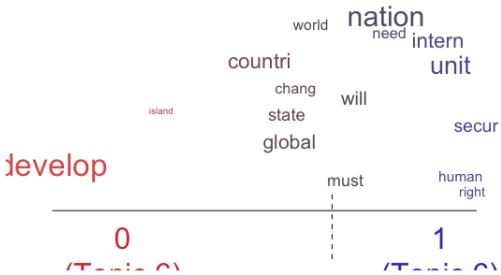


(c) Topic 3

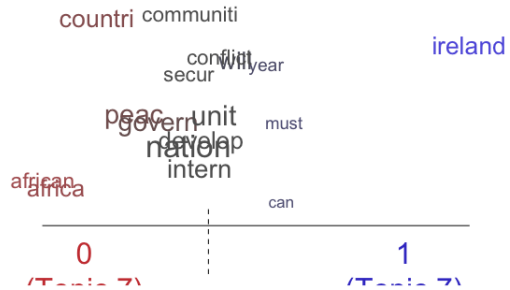


(d) Topic 4

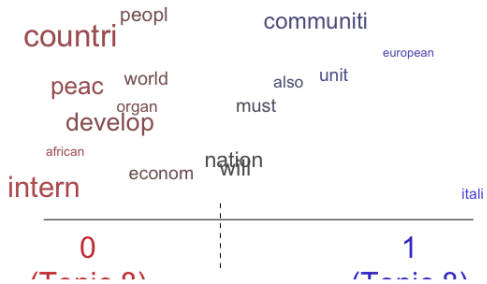
(e) Topic 5



(f) Topic 6



(g) Topic 7

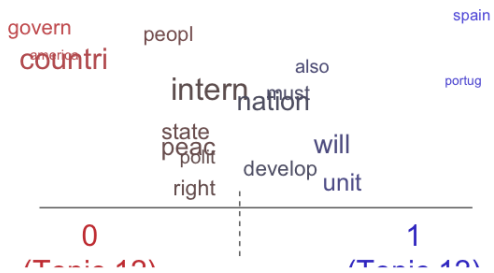


(h) Topic 8

(i) Topic 9

(j) Topic 10

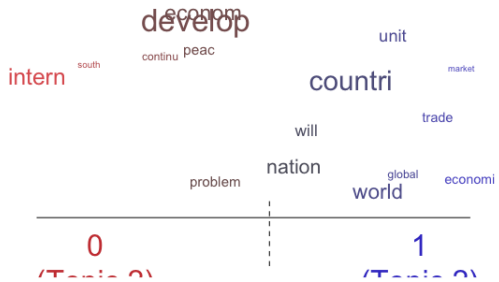
(k) Topic 11



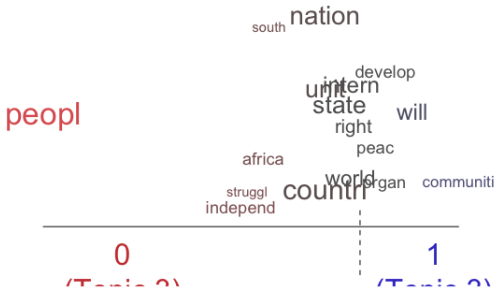
(l) Topic 12

Figure 8: Topical Content: Thresh = 0, K = 18, Covariate = EU (binary)

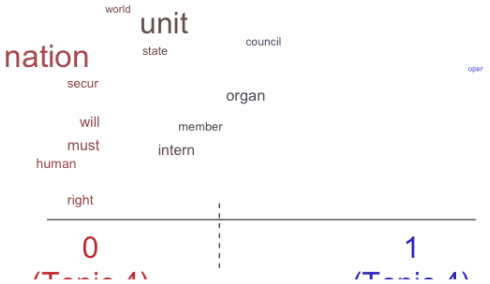
(a) Topic 1



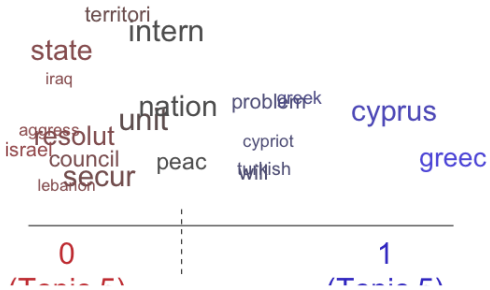
(b) Topic 2



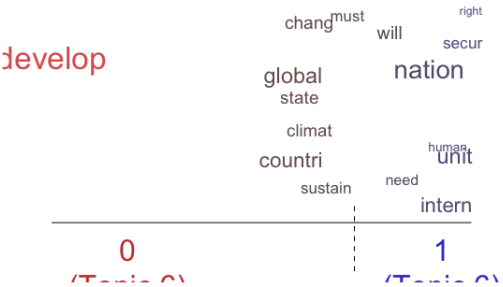
(c) Topic 3



(d) Topic 4



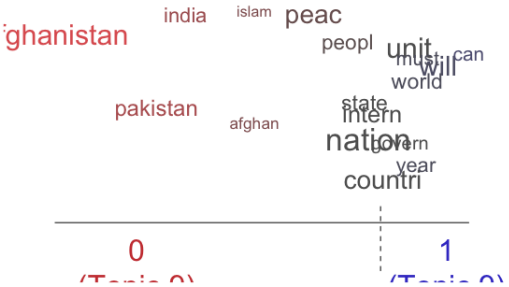
(e) Topic 5



(f) Topic 6

(g) Topic 7

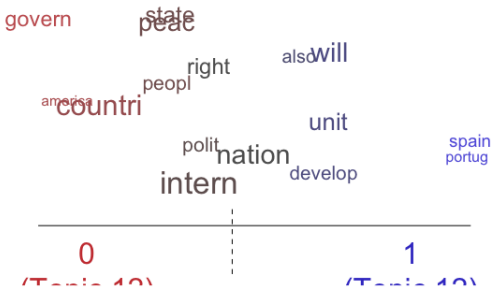
(h) Topic 8



(i) Topic 9

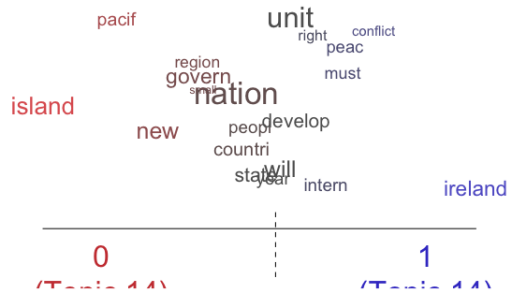
(j) Topic 10

(k) Topic 11

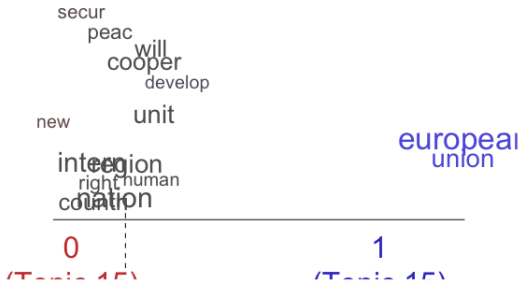


(l) Topic 12

(m) Topic 13

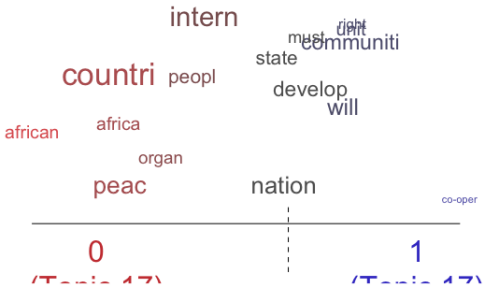


(n) Topic 14



(o) Topic 15

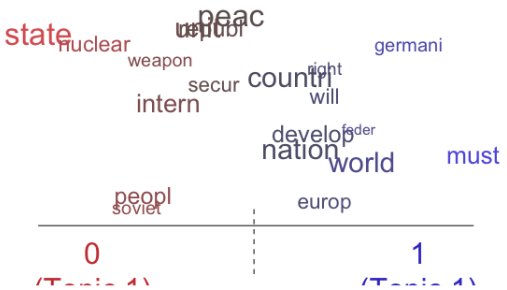
(p) Topic 16



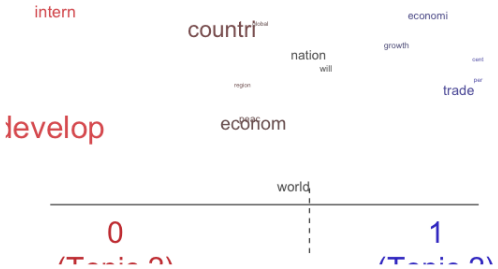
(q) Topic 17

(r) Topic 18

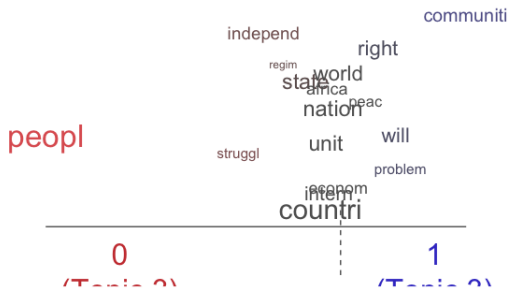
Figure 9: Topical Content: Thresh = 0, K = 24, Covariate = EU (binary)



(a) Topic 1



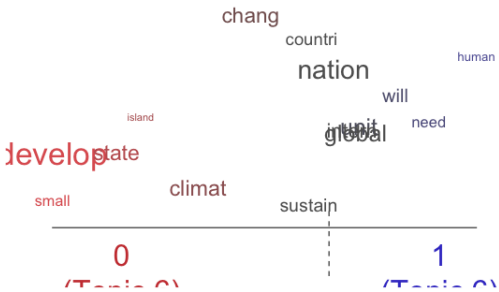
(b) Topic 2



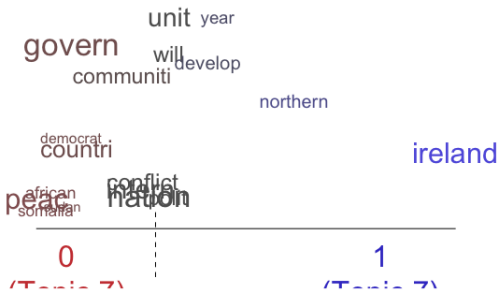
(c) Topic 3

(d) Topic 4

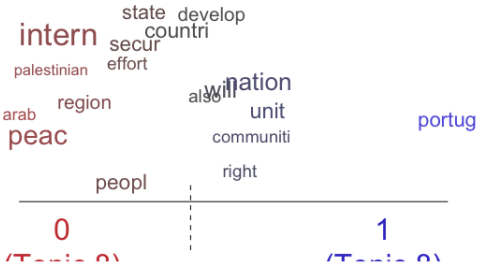
(e) Topic 5



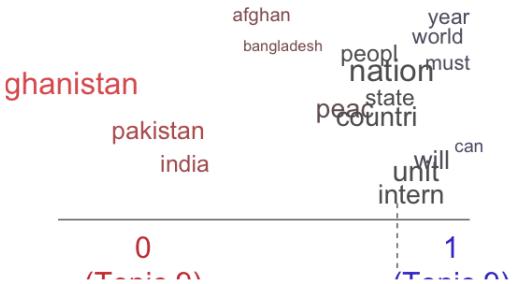
(f) Topic 6



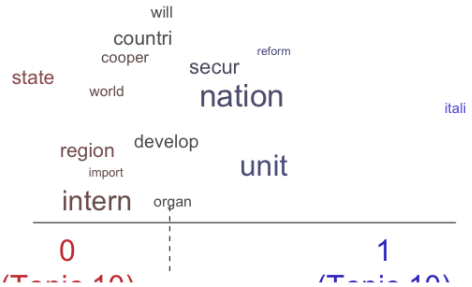
(g) Topic 7



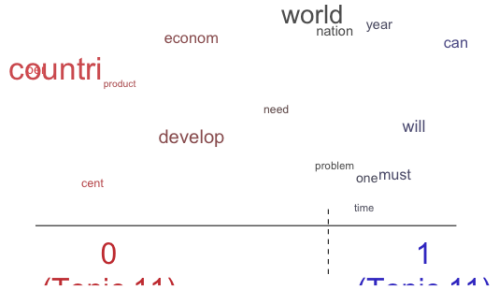
(h) Topic 8



(i) Topic 9



(j) Topic 10

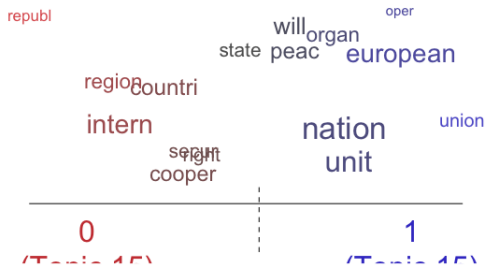


(k) Topic 11

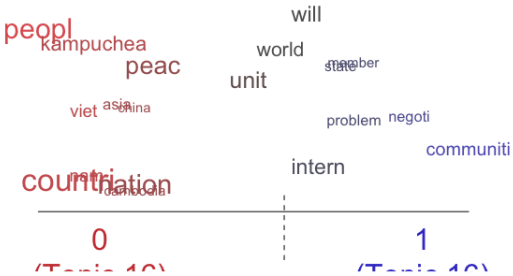
(l) Topic 12

(m) Topic 13

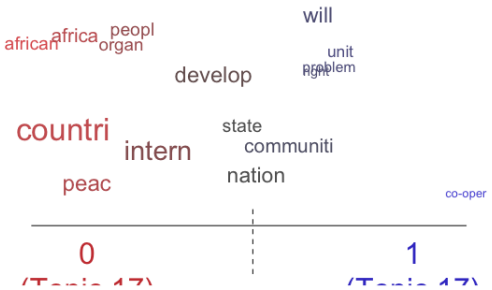
(n) Topic 14



(o) Topic 15

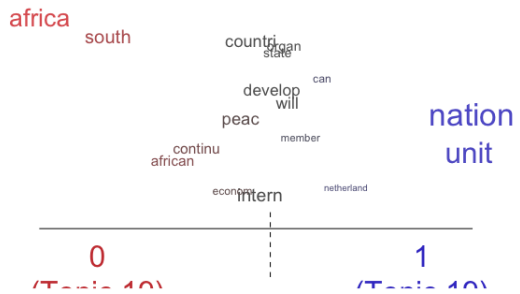


(p) Topic 16

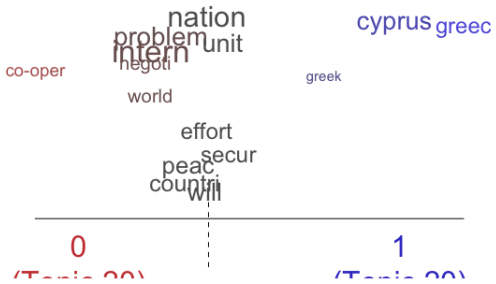


(q) Topic 17

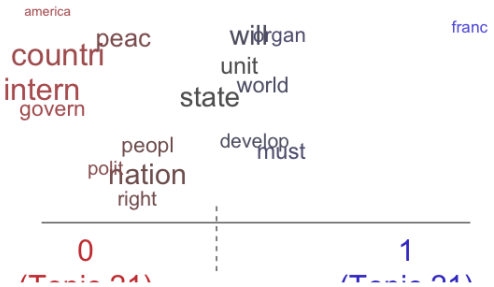
(r) Topic 18



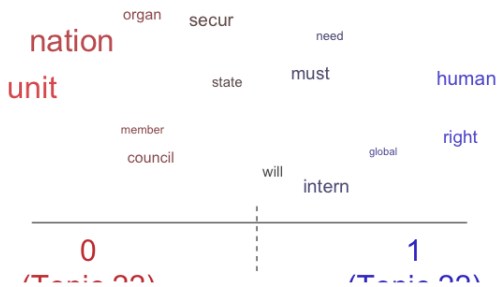
(s) Topic 19



(t) Topic 20

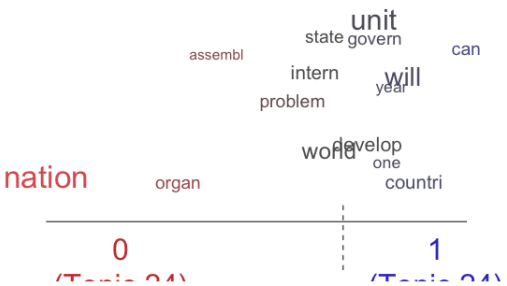


(u) Topic 21



(v) Topic 22

(w) Topic 23



(x) Topic 24