How emotional tone varies by chamber

Comparing parliaments · Five corpora · 1988–2026

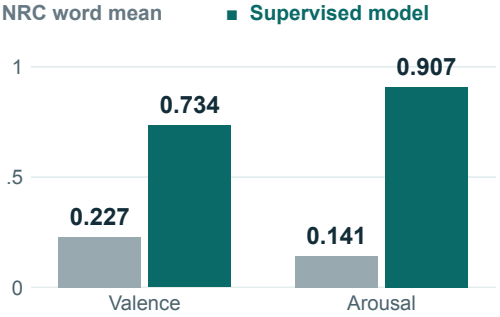
Problem and challenge

How does emotional tone differ across parliaments? Building on Russell (1980) and Pipal et al. (2024), we distinguish valence (negative to positive) from arousal (calm to activated). We use synthetic whole-speech labels to train a 70/30 gradient boosting and TF-IDF Ridge blend selected by grouped five-fold cross-validation. After testing on 5,000 held-out labels, we refit on 10,000 labels and scored 1,681,102 speeches. Future work could apply this method to large collections of public social media posts.

Key findings

01 Speech-level predictions improve

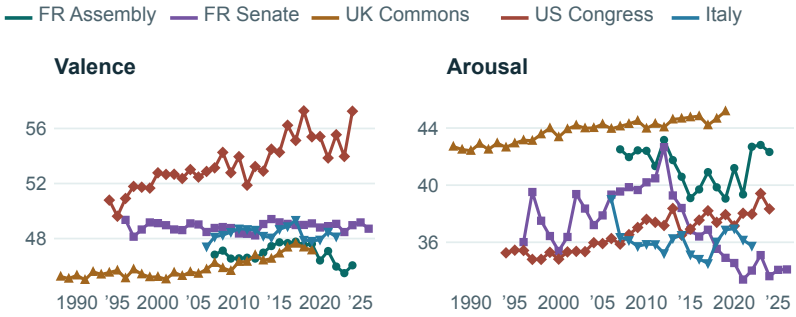
Lin's CCC against 5,000 held-out synthetic labels



Fit on 5,000 synthetic training labels. Over NRC word means, CCC rises by 0.51 for valence and 0.77 for arousal.

02 How tone evolved across chambers

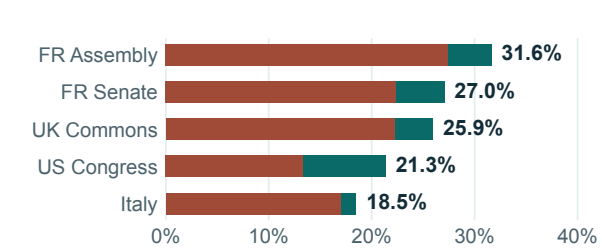
Unadjusted annual mean predictions, 0–100 scale



All 1,680,870 dated speeches (232 lack a usable date). French Senate arousal fell from 42.7 in 2012 to 34.9 in 2019.

03 High arousal, different valence

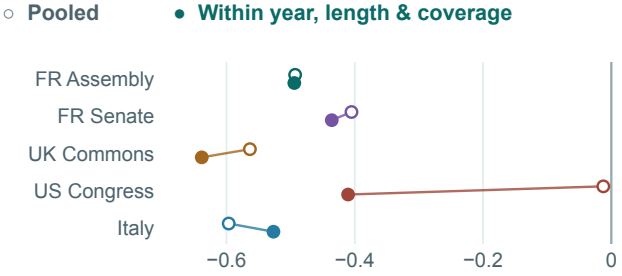
Share with arousal ≥ 48.5 (75th percentile)



In the 2008–2019 matched sample, 31.6% of Assembly speeches and 21.3% of US speeches qualify. Among those, 13% and 37% respectively score ≥50 on valence.

04 Speech length masks a US pattern

Pearson correlation of predicted valence and arousal



In Congress, longer speeches scored higher on both axes. Correlation shifts from −0.01 pooled to −0.41 within profiles. Figures 3–4 use 806,867 matched speeches.

Societal value

A larger share of Assembly speeches reach high arousal. Within that group, Congress more often reaches the valence midpoint. Reporting both scores keeps intensity from being mistaken for hostility.

Real-world use case

A newsroom could compare how chambers discuss the same policy, use speech scores to find the passages driving a shift, and check the claims against transcripts and independent ratings.

Russell (1980) Pipal et al. (2024)

Key takeaway

Synthetic speech labels let us compare valence and arousal at scale. In Congress, speech length masks a negative link between the two scores among comparable speeches.

The predictor was trained and tested on synthetic labels. CCC measures agreement with those ratings. Independent human validation is pending. The planned open-weight labeling run targets 5,000 training and 1,000 test speeches and includes Italy. The Italian predictions shown precede that refit.