Appendix A- Additional information on sample and coding of Bundestag sentences

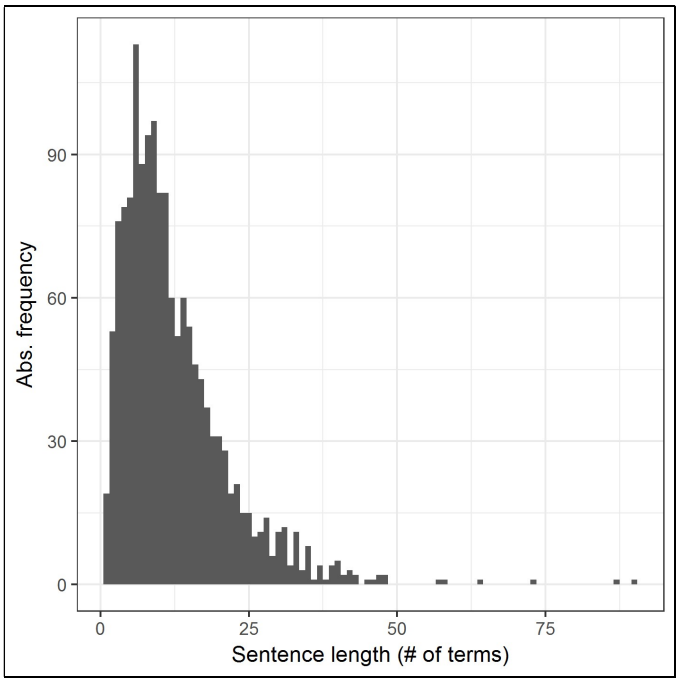

Figure A1: Length of sampled sentences

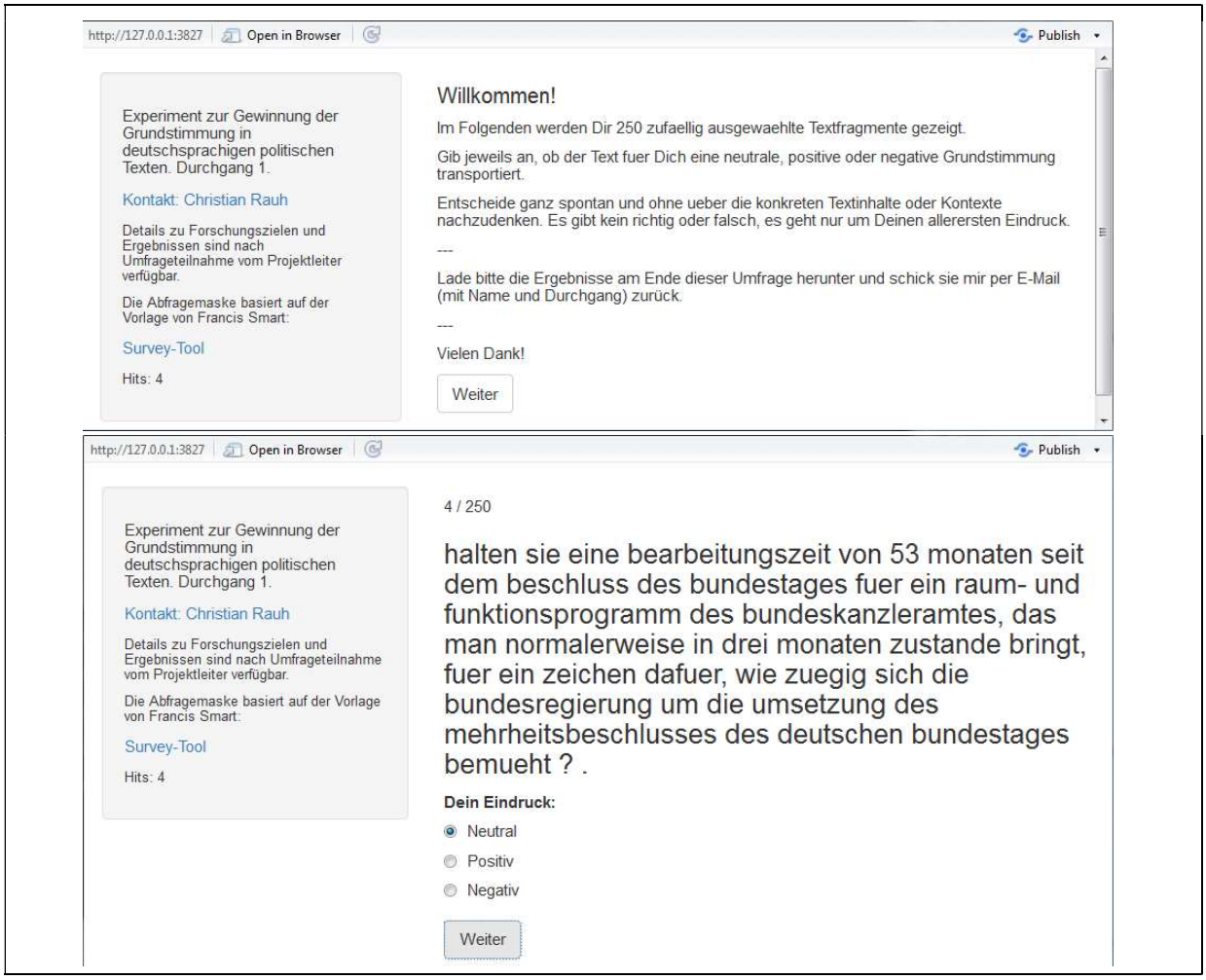

Figure A2: Screenshots of survey tool for human coders

Coders		Percent. agreement	Krippendorff α (ord. data)
Human coders	(3)	59.1	.675
Sent. dictionaries	(3)	77.2	.599
Humans & dictionaries	(6)	24.4	.404

Table A1: Inter-coder agreement
Note: 1.500 randomly sampled sentences from Bundestag speeches coded as 'positive', 'neutral', or 'negative'.

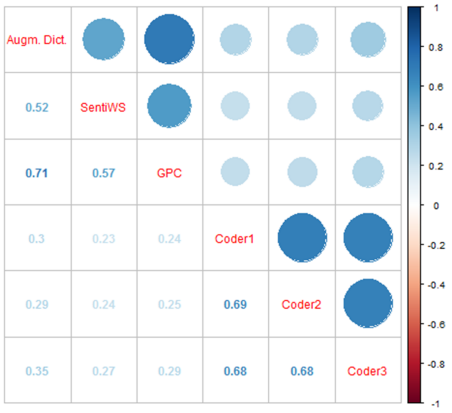

Figure A3: Rank correlations of human and dictionary based sentence classification

Note: Despite their common roots also the sentiment dictionaries achieve no significantly better 'inter-coder reliability' than humans. The univariate distributions shown in Figure 1 in the main text together with the one-standard deviation cut-off-point for text classification suggest that particularly the *SentiWS* dictionary with its much lower spread disagrees frequently with the other two automated procedures. For applied research wishing to reduce automated scores to discrete ordinal scales this means that the robustness of the resulting inferences should be checked against different cut-off points.

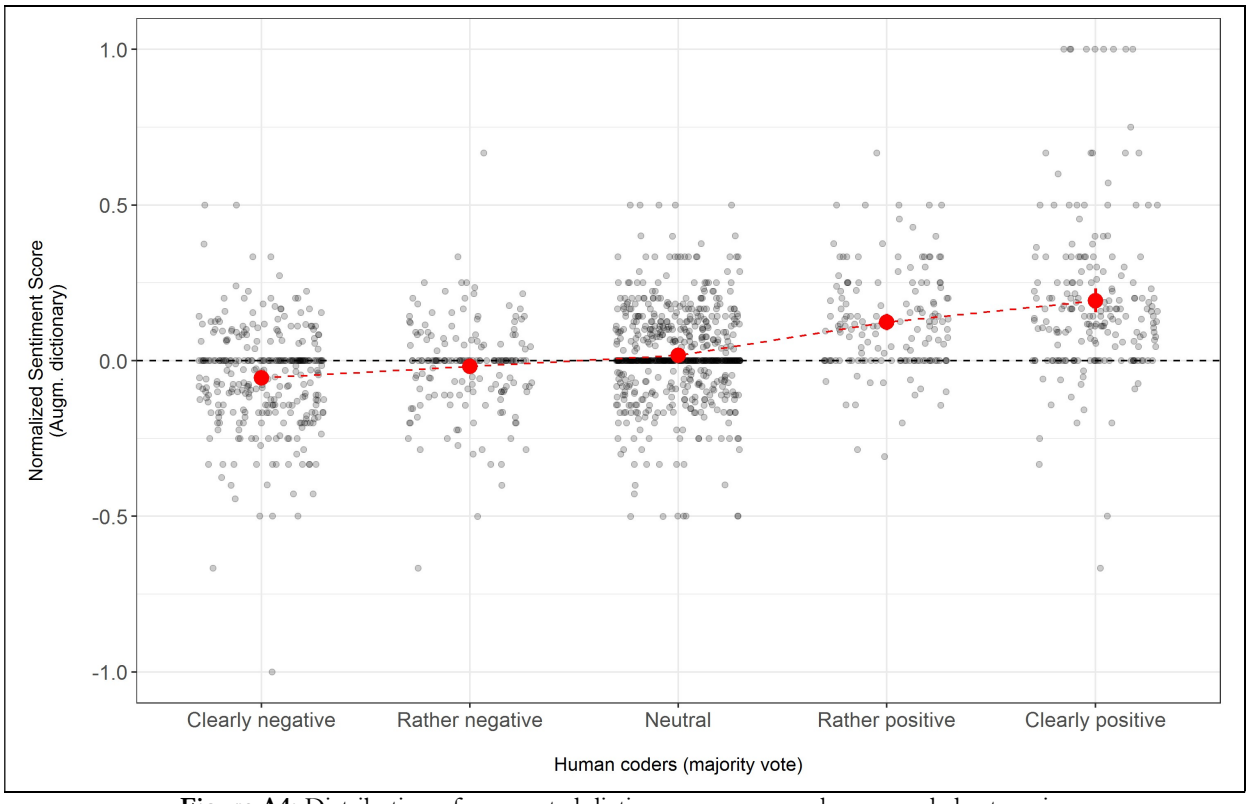

Figure A4: Distribution of augmented dictionary scores over human coded categories

Appendix B- Additional information on scoring Manifesto units

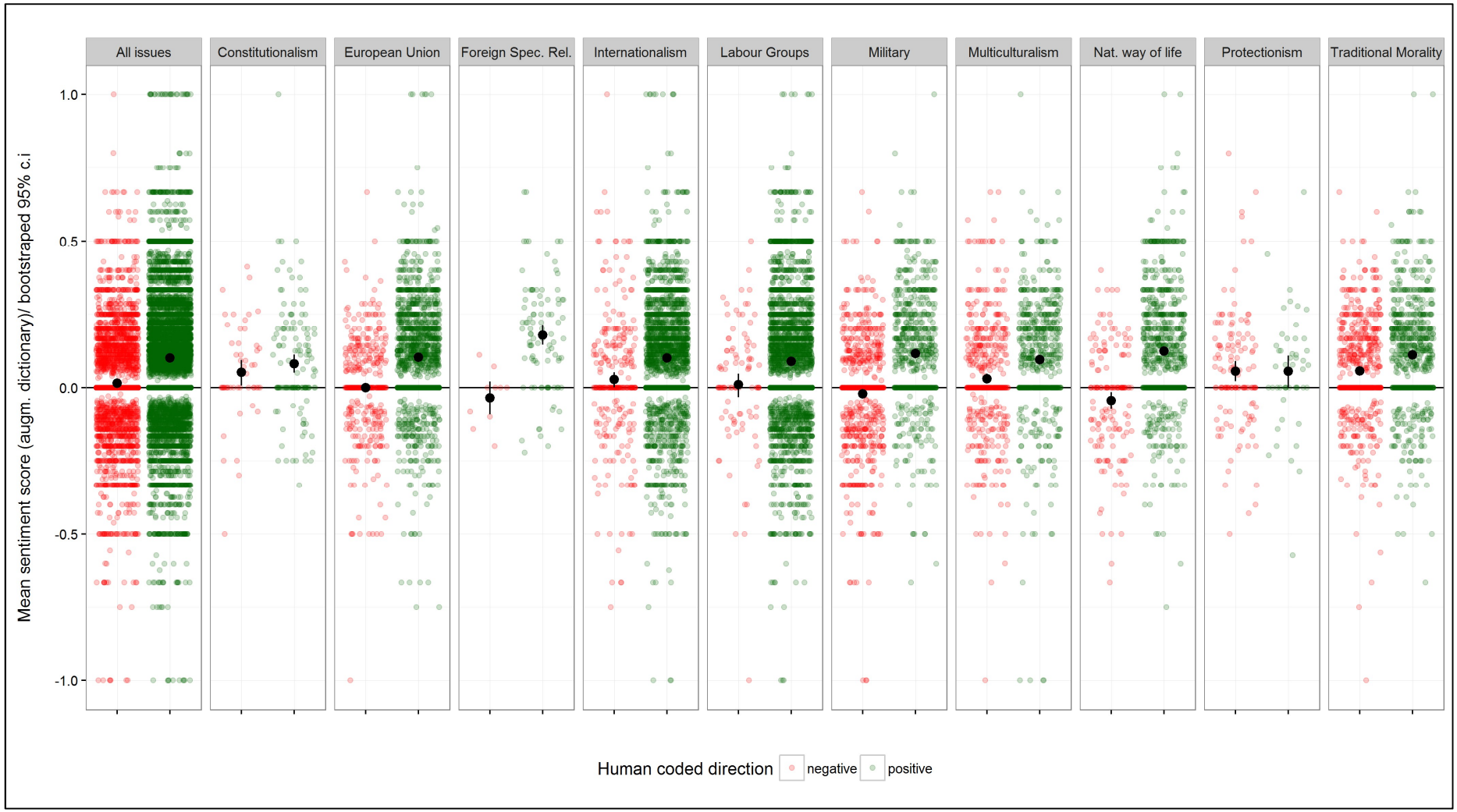

Figure B1: Full distribution of sentiment scores across human codes and Manifesto categories

Appendix C- Additional information on sampling/scoring 'Causa Wulff' media coverage

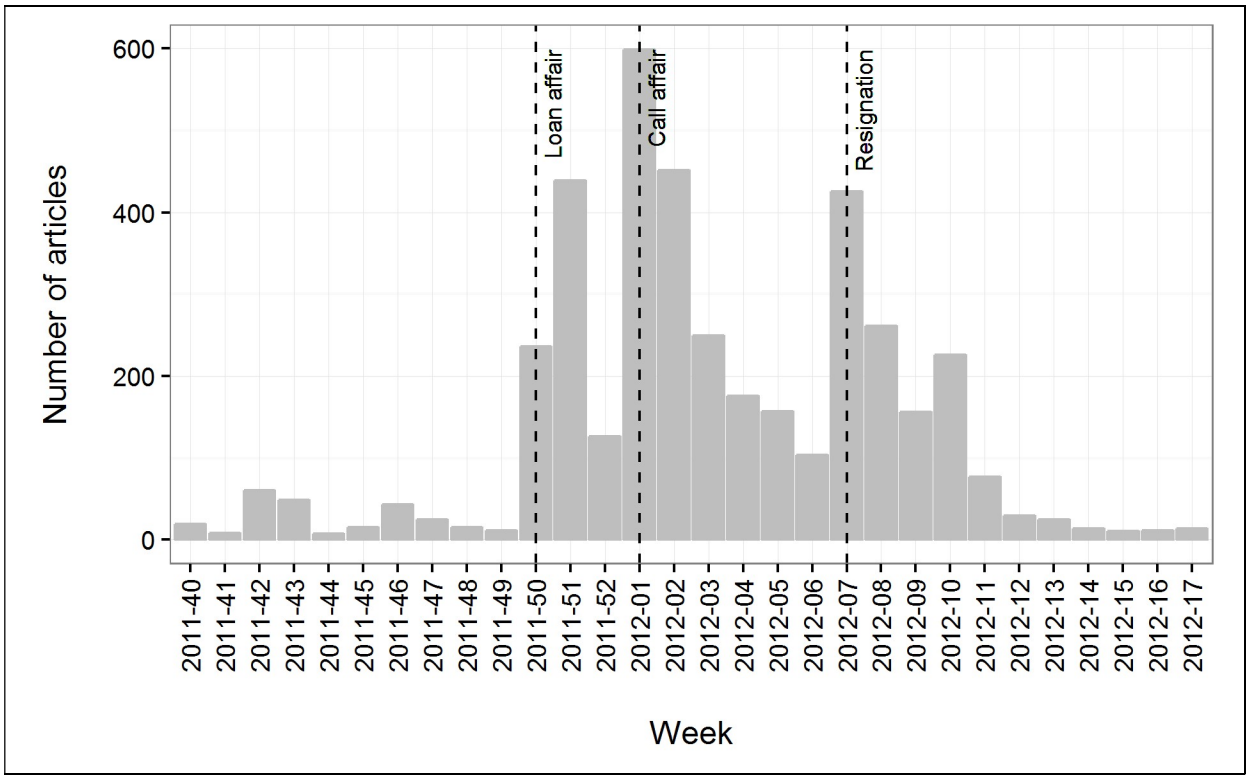

Figure C1: Temporal distribution of newspaper articles in 'Causa Wulff' sample

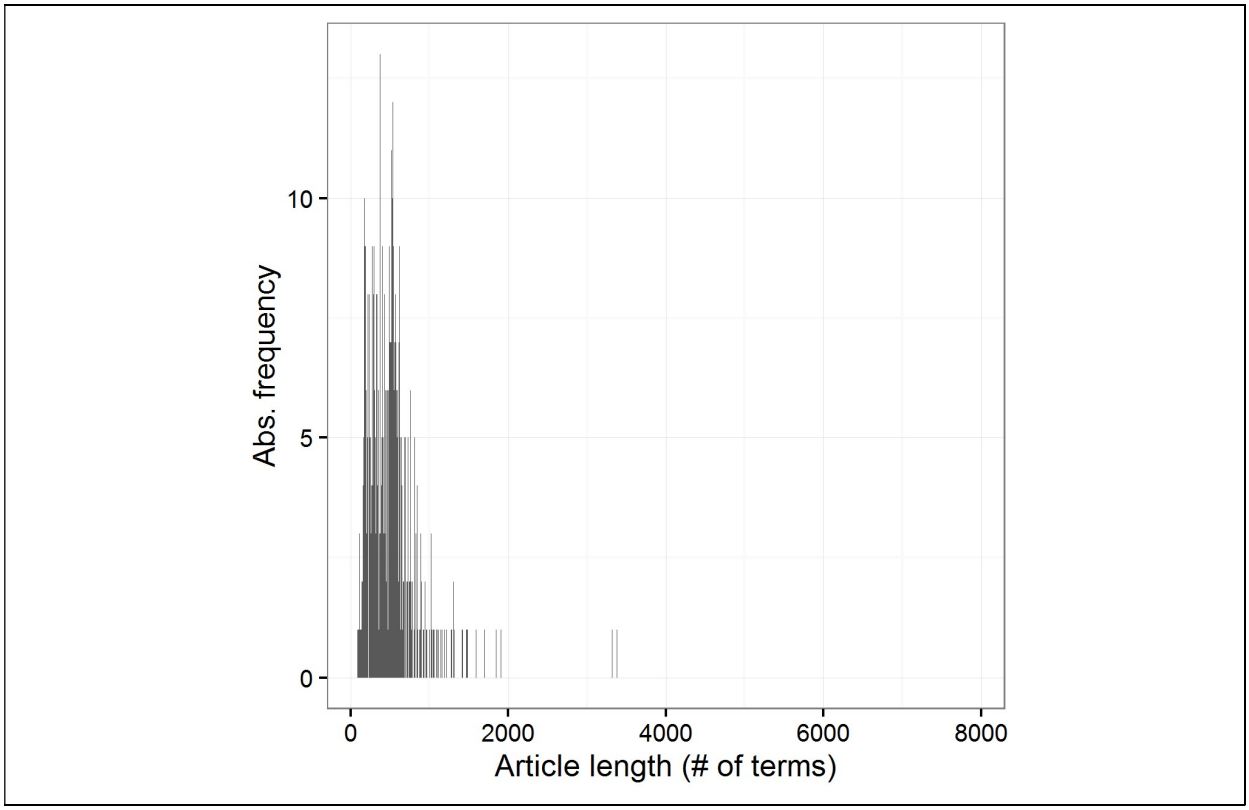

Figure C2: Histogram of article lengths in 'Causa Wulff' sample

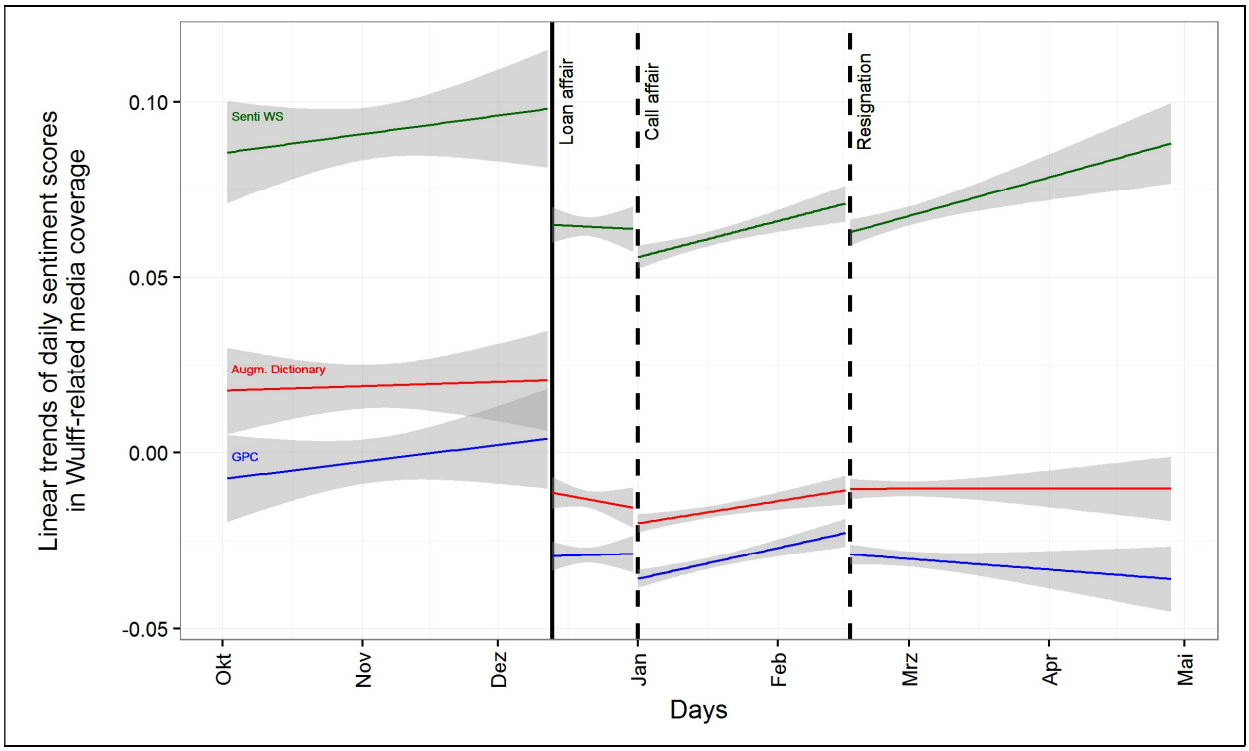

Figure C3: Comparing different dictionaries during the 'Causa Wulff'