Incivility in Austrian parliamentary debates:
A supervised sentiment analysis of parliamentary speeches

Marcelo Jenny[1], Martin Haselmayer[2], Elena Rudkowsky[3],
Matthias Wastian[4], Stefan Emrich[4] and Michael Sedlmair[3]

First draft – Please treat accordingly!

Paper prepared for the Workshop Political Incivility in Parliament,
ECPR Joint Sessions of Workshop,
Nottingham, 25-30 April 2017

Funding:
This research was partially funded by the Hochschuljubiläumsstiftung der Stadt Wien (H-304565/2015) and FFG project 845898 (VALID).

Acknowledgements:
We thank Elisabeth Graf and Lisa Hirsch for their research assistance.

[1] Corresponding author: marcelo.jenny@uibk.ac.at, Department of Political Science, University of Innsbruck.
[2] Department of Government, University of Vienna.
[3] Research Group Visualization and Data Analysis, Faculty of Computer Science, University of Vienna.
[4] Drahtwarenhandlung: Simulation services and Technical Solutions (dwh GmbH), Vienna.

Introduction

Incivility of political communication has become a major topic in public and scientific discourse (e.g. Herbst 2010; Berry and Sobieraj 2013), and it is often seen as a cause of increasing political polarization, lower electoral turnout and voter disaffection with politics and democracy in general (Jamieson 1992; Kahn and Kenny 1999; Mutz and Reeves 2005, Mutz 2007; Brooks and Geer 2007; Lau and Rovner 2009; Harcourt 2012). However, there is no agreement on the definition or measurement of incivility.

Our paper presents an automated sentiment analysis to identify uncivil language and to measure the level of (in)civility in parliamentary speeches. Substantively, we study incivility in the Austrian national parliament during the last two decades (1996-2013) and explore some of the political, institutional and individual factors that affect the level of incivility shown in parliamentary debates. We check whether government/opposition status, the parliamentary role, the type of debate and closeness to the next election has an effect on the level of civility observed in parliament.

Measuring incivility in parliamentary debates

To identify uncivil statements in parliamentary debates we draw on techniques computer scientists use for sentiment analysis, a subfield in natural language processing (Liu 2015). In addition we use crowd-coding, content analysis by lay coders recruited on the internet (Mattes and Redlawsk 2014; Haselmayer and Jenny, forthcoming), to produce human-annotated input data for a supervised learning algorithm that will be used to classify civil and uncivil statements in parliamentary speeches.

First generation sentiment analyses identified the polarity of textual statements as positive or negative. More recent research is aiming for an ordinal measurement of sentiment. We want to build on that. Similar to many studies on political incivility (e.g. Mutz & Reeves 2005; Brooks and Geer 2007; Fridkin and Kenny 2004, 2008, 2011; Sobieraj and Berry 2013; Stryker, Conway and Danielson 2016), we conceptualize incivility as the outer or extreme section of the negativity dimension. Uncivil statements are 'strongly' negative or perceived as such by an audience or by respondents.

Brooks and Geer (2007: 5) define incivility as "claims that are inflammatory and superfluous". The difference in their study is "two strong, pointed words" that transform a 'civil negative' into an 'uncivil negative' statement (Brooks and Geer 2007: 5). Thus the negative sentiment strength of two words is enough to move a 'civil negative' statement into the uncivil section of the negativity dimension. Two vignettes from their study are listed in Table 1 below, with the respective additional words highlighted.

Table 1: Incivility as the more negative section on a negativity dimension

Civil negative	**Uncivil negative**
My opponent does not live by the values of family, as represented by his several divorces, or by community responsibility, as shown by his failure to volunteer. This lack of values will prevent him from being an effective representative.	My <u>unprincipled</u> opponent does not live by the values of family, as represented by his several divorces, or by community responsibility, as shown by his <u>unforgivable</u> failure to volunteer. This lack of values will prevent him from being an effective representative.
My opponent has not been a leader in the past, and will fail to lead this district successfully into the next decade.	My <u>cowardly</u> opponent has not been a leader in the past, and will <u>utterly</u> fall to lead this district successfully into the next decade.

Source: Brooks and Geer (2007), Appendix A.

In these examples it is the sentiment strength of the additional words used, that makes the difference.

We measure incivility in political texts using a form of supervised sentiment analysis. The description of the procedure will be split into the following steps:

1. Creating a crowd-coded training data set
2. Obtaining vector representations of words and sentences
3. Supervised learning of civil and uncivil statements with a neural network classifier
4. Generating incivility scores for sentences and speeches

Creating a training data set through crowd-coding

Extending the data set of crowd-coded statements previously presented in Haselmayer and Jenny (forthcoming) we build a training data set consisting of about 20,600 sentences. These sentences cover two decades of Austrian parliamentary debates and party press releases plus election news reports from the most recent national elections of 2013. Table 2 provides an overview of the composition of the data set.

Table 2: The training data set of crowd-coded sentences

Source	Time period	N	%
Party press releases	1995-2013	14,343	70
Parliamentary debates	1995-2013	3,923	19
Media reports	2013	2,327	11
Total		20,593	100

As the sentences are in German, the anonymous lay coders recruited via the crowd-coding platform CrowdFlower come from Austria, Germany and Switzerland. Each sentence was assigned to at least ten different coders who were asked to rate its negativity on a 5-point scale ranging from 0 (not negative) to 4 (very strongly negative), or to declare it as uncodable.

Individual coder performance is monitored during the coding process to identify cheating or spamming. Each participant has to answer four test questions correctly before the actual coding task starts. In addition, one out of five sentences presented during the coding task is actually a test item. Coding sentiment strength on a five-point ordinal scale is a difficult task (Pang et al. 2002; Hopkins and King 2010), so for the test items we accepted two adjacent options on the five-point scale as correct answers (relative to the reference coding established by some of the authors).

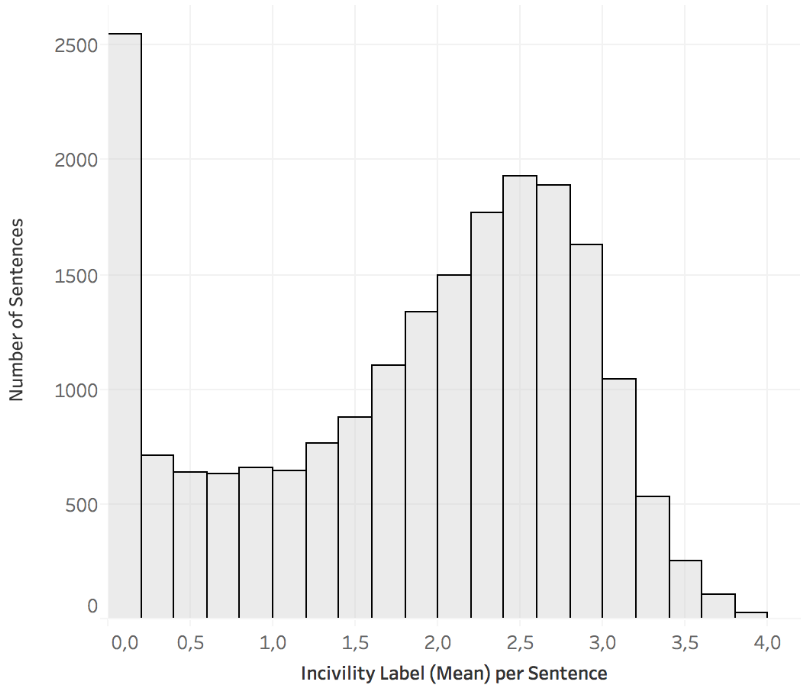
The probability of passing the entry test by guessing correctly is only 4 percent and it gets smaller with each additional test item.[5] During the coding process coders dropping below an

[5] An accuracy threshold of 75 percent means three of the four initial test questions have to answered correctly. Including the “uncodable” answer a coder has six options of which two are accepted as correct. The probability of passing by guessing then is $(\frac{2}{6})^3 = 0.04$.

accuracy threshold of 75 percent correct answers to the test items are stopped from contributing further and their codings are not included in the final training data set.

For each sentence we calculate the arithmetic mean of the ten negativity ratings as its negativity score. Figure 1 shows the distribution of these sentence scores.

Figure 1: Distribution of negativity scores in training data set



Many sentences were identified as neutral. The second mode of the negativity scores was in between moderate and strongly negative. As perceptions of statements differs, we have have a considerable amount of variability in the codings (see Brooks and Geer 2007; Stryker et al. 2016) and averaging individual scores results in very few sentences with very high negativity score.

Obtaining vector representations of words and sentences

To train and predict sentence-level incivility we create distributed sentence embeddings based on averaging the word embeddings related to the words in a sentence. It is crucial to retrieve many word vectors per sentence from a German word embedding corpus to achieve high incivility prediction accuracy for the resulting sentence (average) vectors. Distributed word embeddings enable us to establish the various words' degree of semantic similarity. The famous distributional hypothesis of Harris (1954) states that words that occur in the same or similar contexts also have similar meanings.

We retrieve vector representations for words from the German language part of Polyglot (A-Rfou et al. 2013), a large multi-lingual corpus created from the vocabulary of Wikipedia articles. Based on the word2vec model developed by Mikolov et al. (2013), Polyglot provides word embeddings for the 100,000 most frequent words of the German language Wikipedia pages. They cover 92 percent of the content on the German Wikipedia website.

As described above, sentences in our training data set were sampled from several input sources. It turned out that the initial extraction of sentences from the sources was imperfect and some sentences also contained extraordinary elements. The problems encountered were a) complete sentences written in upper case, e.g. titles of press releases b) unnatural words with hyphens because of surviving end of line hyphens, c) and unnatural compound words because of missing spaces between words originally separated by a line break. While the German language has a lot of compound words, many of the compound words in our training data set do not exist in reality. According to the initial counts the training data set comprised 40,000 unique 'words' (or more exactly strings separated by blanks). After data cleaning the number dropped to about 30,000 words.

Each word (or text string) encountered in a training sentence is checked against the Polyglot corpus. If it is not found there, we apply string modifications in specified order and check after each step again whether a word exists. If a word has been found in Polyglot, the procedure moves on to the next word, if not, the word will simply have no impact on subsequent calculations of sentence vectors. The modifications include setting the word to lower case, setting the word to lower case, capitalizing it, removing hyphens with single and compound word checks, lemmatizing, stemming and checking numerous substring combinations.[6]

About half of the 40,000 words (or text strings) in the training sentences were initially not found in the Polyglot corpus. Pre-processing reduced the number of unknown words to about 10,000. Some of these were person names or other named entities, some were numbers. Many of the unknown words (or strings) appeared only once.

Our final word vector representations provide the building blocks for a vector representation of each sentence in the training set, produced by averaging the vectors of all the identified words in a sentence (Mikolov et al. 2013). Similar sentences with many identical words have similar vector representations.

Supervised learning of civil and uncivil statements with a neural network classifier

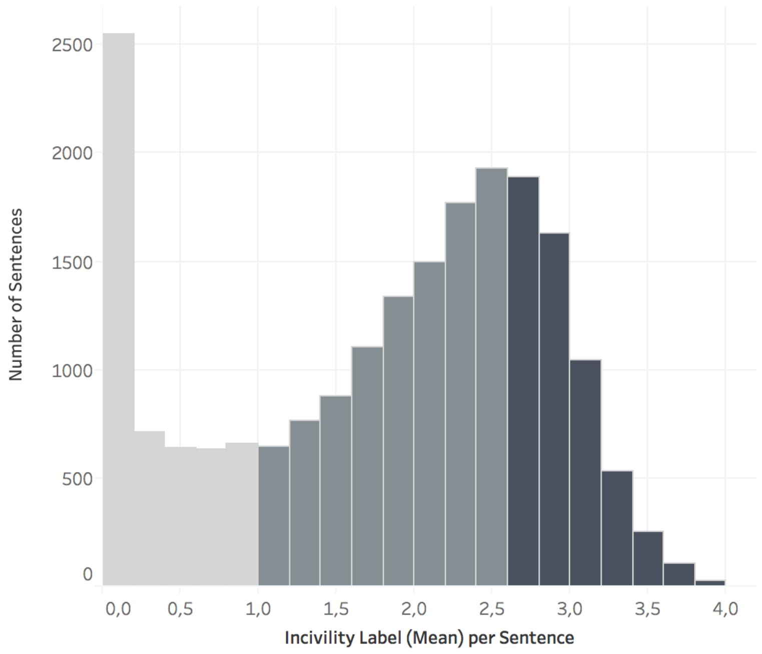
The vector representations of the sentences provide the input features for a neural network-based classification of these sentences as civil or uncivil. A neural network classifier learns to predict for each case the class it belongs to (in our case a recoded sentence negativity score, see below) by processing incoming signals. This is done by processing the sentence vectors through cascading processing layers of neurons that reflect the neural network structure of the human brain. Neural networks have multiple configuration parameters such as the number of layers, number of neurons per layer and the activation (threshold value) function that lets neurons transmit a signal or not. Our final setting was an incoming layer that maps the 64-dimensional sentence vectors to 120 dimensions and an outgoing layer that reduces the complexity to two classes. In between we apply dropout regularization of 20 percent to prevent overfitting by the learning procedure.

We want the algorithm to predict whether a sentence is uncivil or not. Test runs showed that the neural network classifier worked considerably better by training only on extreme cases, discarding the sentences with an intermediate level of negativity. So we trained the neural

[6] We also looked up all substrings of a word, but the error rate thereby introduced was too high. A long German word can often contain substrings which constitute completely different and unrelated separate words.

network classifier on sentences with negativity scores below 1 and sentences with a negativity score above 2,7 (see Figure 2).

Figure 2: Binary classification from subsets of outlying cases



The training data set contains about 5,000 sentences in the lower section of the scale. To get a balanced training set we sampled another 5,000 sentences from the upper part of the scale. In effect, we ended up using only half of the crowd-coded sentences for the training, which left more than enough unused sentences for testing the accuracy of the trained classifier.

We wanted the neural network classifier to exhibit equally good performance in predicting civil and the uncivil statements, as both classes are important to us. The common performance metrics we report are precision and recall, as well as the F-measure as their combination (Fawcett 2006). The trained algorithm is now applied to about 1,400 sentences which were not used during the training phase. This fresh set is balanced: about half of the sentences are from the civil and uncivil class. The results are shown in Table 3.

Table 3: Classifier performance on unused sentences in training data set

	Precision	Recall	F-measure
Civil statements	80%	77%	79%
Uncivil statements	77%	81%	79%

Table 3 shows precision, recall and F-measure are all close to 80 percent, which is a considerable improvement from the 50 percent obtainable by random classification. We aim to improve these metrics further, but want to explore the trained algorithm in its current version to identify incivility in Austrian parliamentary debates.

Generating incivility scores for sentences and speeches

We now have an algorithm to be used in the ‘wild’. Each sentence from a parliamentary speech is classified by the algorithm as civil (0) or uncivil (1). Then we calculate the mean of all sentences contained in a speech to get the incivility score for the complete speech.

A first validity test: Incivility in Austrian parliamentary debates

In this section we undertake an exploratory study to evaluate the validity of the proposed approach for the measurement of the amount and level of incivility in debates in the National Council, the lower house of parliament. We formulate some expectations for which we have strong prior beliefs how the resulting data patterns should look be.

Our substantive analysis uses a data set of 56,701 parliamentary debates covering the last five completed legislative periods (1996-2013), with speeches given by 577 different MPs and ministers. Over this period seven parties (SPÖ, ÖVP, FPÖ, Greens, Liberal Forum, BZÖ, and Team Stronach) were represented in parliament.

Government composition changed from a SPÖ-led grand coalition with the People's Party (ÖVP) in the beginning to centre-right coalitions of ÖVP and Freedom Party (FPÖ). After the latter party's split in 2004, the Alliance for the Future of Austria (BZÖ) replaced the FPÖ in government. After the election 2006 SPÖ and ÖVP renewed their coalition partnership and continued to govern together after the elections of 2008 and 2013. Table 4 provides an overview of the number of parliamentary speeches plus the political context information.

Table 4: Parliamentary debates in the National Council (lower house)

Legislative period	Start-end	Parliamentary speeches	Percent	Parties in parliament	Government coalition
20	1996-1999	11,797	20.8%	5	SPÖ-ÖVP
21	1999-2002	9,164	16.2%	4	ÖVP-FPÖ
22	2002-2006	12,409	21.9%	5	ÖVP-FPÖ/BZÖ
23	2006-2008	5,902	10.4%	5	SPÖ-ÖVP
24	2008-2013	17,429	30.7%	6	SPÖ-ÖVP
Total	1996-2013	56,701	100.00	7	

Hypotheses:

We expect government parties to show lower levels of incivility in parliament than opposition parties. Ministers and MPs from government parties propose and defend the government's bills – which constitute the bulk of legislative proposals in each term –, while MPs from opposition parties criticize these proposals more or less intensively.

Intra-coalition conflict is quite common in Austria, but empirical analysis suggests that it is less harsh compared to the tone of verbal exchanges across the government-opposition divide (Haselmayer and Jenny forthcoming). Most of the time public intra-coalition conflict is displayed in other arenas, e.g. via press conferences or interviews, rather than in the parliamentary arena.

Government bills require a unanimous cabinet decision before they are introduced to parliament. This means government bills are sent to parliament after the coalition partners have reached a sometimes hard won and painful policy compromise, which they have to defend against multiple criticisms – from within their own party, from opposition parties and from the wider public. This reduces government MPs' incentives, and above all the leeway accorded to them by the parliamentary party group leadership, to criticize a coalition partner's ministers or MPs. The main task of government MPs in legislative debates is to promote and defend the own party's and government's achievements (e.g. Benoit 1999; Walter 2014; Maier & Jansen, forthcoming), In a coalition government this often means government MPs from different parties engage side by side as defenders against a critical parliamentary opposition. Opposition party MPs have much less to lose and are more likely to resort to strong negativism.

Hypothesis 1: Debate speakers from government parties exhibit less incivility than debate speakers from opposition parties.

Parliamentary party groups are characterized by a division of labour and role specialization (Heidar and Koole 2000; Jenny and Müller 2012, Dolezal et al., forthcoming). Parliamentary party group leaders are the conductors of a party's rhetorical performance in parliament. They are not only among a party's most frequent speakers. In any important parliamentary debate the parliamentary party group leaders usually step up first to the podium to declare their party's stance (Müller et al 2001). Only if it is a debate on a government bill, the respective government minister will give the first speech. Party group leaders set a party's tone for the debate on the issue of the day. Their level of civility serves as a guidepost for later speakers from the same party.

We expect cabinet ministers to mostly refrain from uncivil rhetoric. A minister may sometimes be provoked by an uncivil attack coming from the opposition to retort likewise, but in general there is not much of an incentive for ministers to use strongly worded language against others. Distinguishing cabinet members, parliamentary party group leaders and ordinary MPs we expect the following pattern to hold:

Hypothesis 2: Parliamentary party group leaders are most likely to use uncivil statements, followed by ordinary MPs. Cabinet members are least likely to use uncivil statements.

Parliamentary debates differ in type, format and focus. Debates on legislative proposals follow an agenda set by the government. In other types of debates the opposition sets the agenda, selecting topics for debate that will hurt the government parties most. One of these is the Urgent Question with debate in the National Council. Urgent questions to a minister can be introduced without any prior warning at the start of a parliamentary day's session. The minister is then forced to respond to the question within a few hours and on the afternoon of the same day at the latest, followed by a debate. Urgent questions became such a strong instrument in the hands of skilful opposition party leaders in the 1990s that the government parties at some point decided to change the standing orders and impose quantitative limits on the instrument's use.

We expect the level of incivility in Urgent Question debates to systematically differ from the level found in other parliamentary debates.

Hypothesis 3: Urgent Question debates exhibit higher levels of incivility than other parliamentary debates.

Another factor discussed in the literature is the closeness to elections (Damore 2002; Walter et al. 2014. An approaching election intensifies parliamentary disputes that foreshadow the parties' actual election campaigns. Coalition parties want to emphasize where and how they differ from their government partner, claim a record of accomplishments and absolve themselves from any failures, while the opposition parties intensify the efforts to remind voters of the government's unfulfilled promises and failures. For a number of reasons we should expect increasing rhetorical heat and more instances of incivility when the next election approaches.

Of course, in any legislative term there are some government bills that are more controversial than others, so the patterns of incivility should exhibit a few peaks throughout the term, but we should see an upturn in the level of incivility closer to an election.

Hypothesis 4: In parliamentary debates close to the next parliamentary election, the the level of incivility will be higher on average than in the rest of the legislative term.

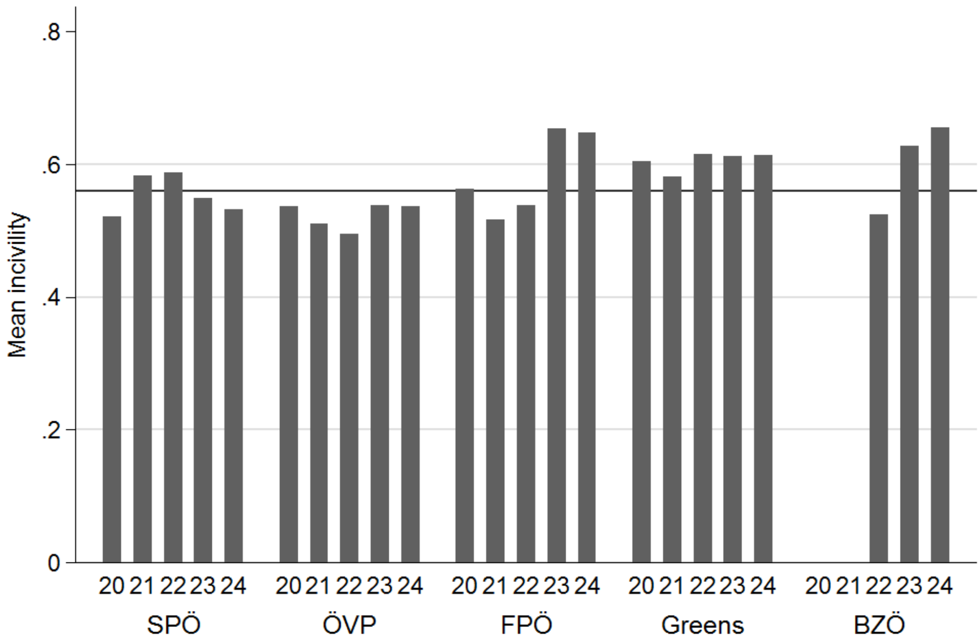
Finally we test the common hypothesis that incivility in politics has become worse over time.

Hypothesis 5: The level of incivility in parliamentary debates has increased over time.

Results

Figure 4 shows the mean of the incivility scores for all speakers from each of the parties represented in parliament over at least three of the five legislative terms analyzed.

Figure 3: Mean incivility of parliamentary parties[1] in last five legislative terms



Note: [1]Figure 4 shows only the parliamentary parties represented in three or more terms.

It is worth remembering that two parties switched sides – from government to opposition and back again or the reverse way: These were the parties SPÖ and FPÖ. The SPÖ was in opposition during the 21st and 22nd legislative term. The FPÖ was a government party in the 21st and parts of the 22nd term, until it was left almost without parliamentary representation as its split-off BZÖ – basically the ministers and almost all MPs remained in government. Two parties always held the same status. The ÖVP was always a government party, the Greens were always an opposition party. The BZÖ was a government party during the 22nd and an opposition party in the later terms. The patterns observed lend credence to the hypothesis that opposition parties exhibit higher levels of incivility.

Table 6 shows the speakers with the lowest and highest mean incivility scores in each of the five periods studied. Heinz Fischer (SPÖ), then the president of parliament, had the lowest score which fits perfectly with his role as a neutral arbiter (Jenny and Müller 1995). The list with the highest incivility scores includes two parliamentary party group leaders (Josef Cap and Herbert Kickl), both leading an opposition party at the time.

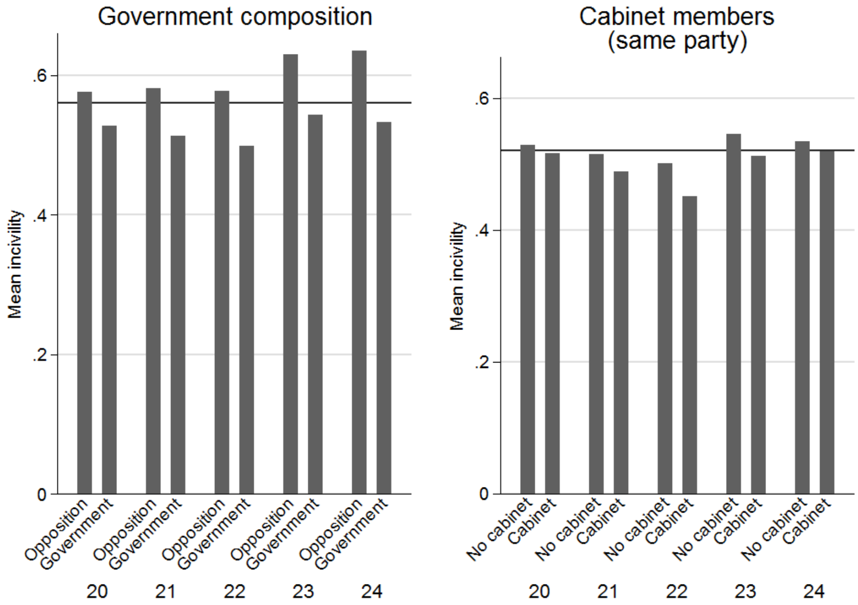
Table 5: Speakers with lowest and highest incivility scores of each term

Legislative term	Lowest	Score	Highest	Score
GP 20	Heinz Fischer (SPÖ)	0.24	Liane Höbinger-Lehrer (FPÖ)	0.67
GP 21	Peter Haubner (ÖVP	0.23	Josef Cap (SPÖ)	0.72
GP 22	Erwin Hornek (ÖVP)	0.31	Josef Cap (SPÖ)	0.70
GP 23	Johann Hell (SPÖ)	0.30	Hermann Lautenschlager (FPÖ)	0.76
GP 24	Rosa Lohfeyer (SPÖ)	0.32	Herbert Kickl (FPÖ)	0.80

Note: Numbers are the individual’s mean incivility scores per legislative term.

The left panel of Figure 4 provides a direct answer to our first hypothesis, that opposition parties exhibit higher levels of incivility than government parties. The pattern is the same over the five terms.

Figure 4: Mean incivility scores by party status, and for minister and MPs

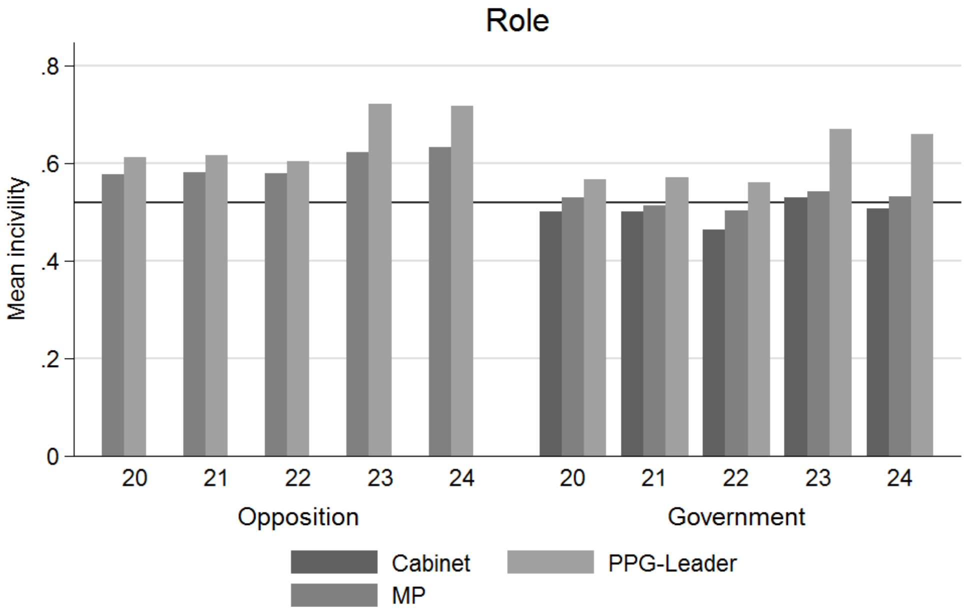


Note: Mean incivility scores aggregated for each legislative term. The left panel shows differences for parties in and out of government. The right panel illustrates how cabinet members differ from ordinary MPs in their own parties.

The right panel compares the incivility scores of ministers with the incivility scores of the coalition parties' MPs. We find the expected pattern, but in some periods (20th and 24th) the difference between members of the executive and government party legislators is rather small.

Figure 5 provides a more fine-grained role differentiation: ministers, parliamentary group leaders, and ordinary MPs for government parties, and the latter two groups only for the opposition parties.

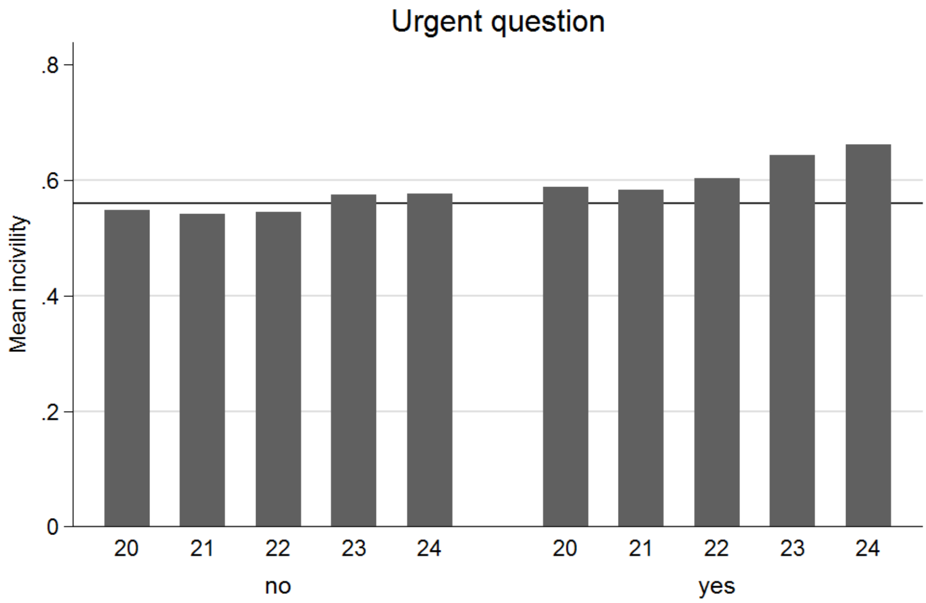
Figure 5: Mean incivility scores by party status and parliamentary role



The pattern is again as expected. Parliamentary party group leaders dole out stronger attacks than the average MP. Ministers tend to exhibit even more rhetorical restraint.

In our third hypothesis we expected higher levels of incivility in Urgent Question debates, because these are strong tools in the hands of opposition parties. Figure 6 confirms the expected pattern.

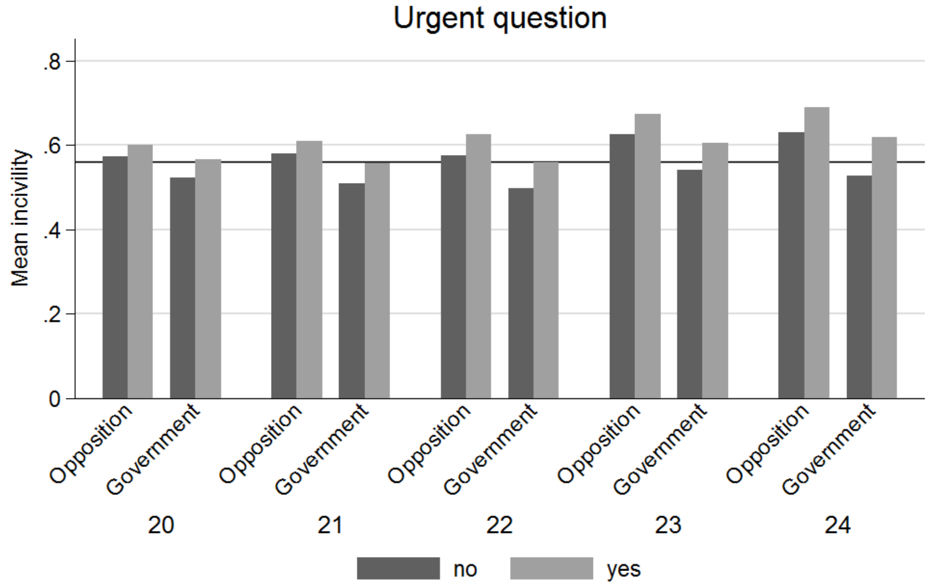
Figure 6: Mean incivility scores for Urgent Question debates and other debates



Note: Mean incivility scores for debates over Urgent question debates and other debates per term

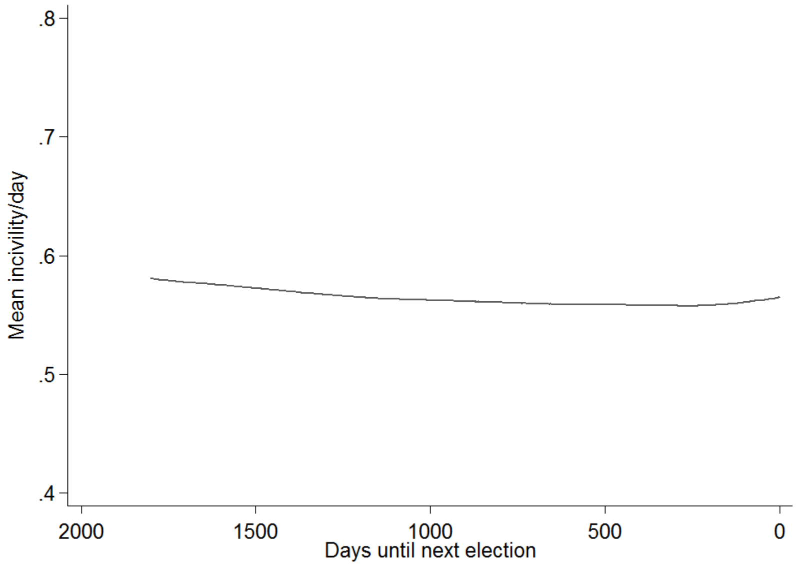
Figure 7 shows that the binary nature of Urgent Question debates leads to higher incivility on both sides. Speakers from government parties have higher incivility scores in Urgent Question debates than in other debates, though speakers from opposition parties will generally trump them in that regard.

Figure 7: Urgent Question debates and the government/opposition divide



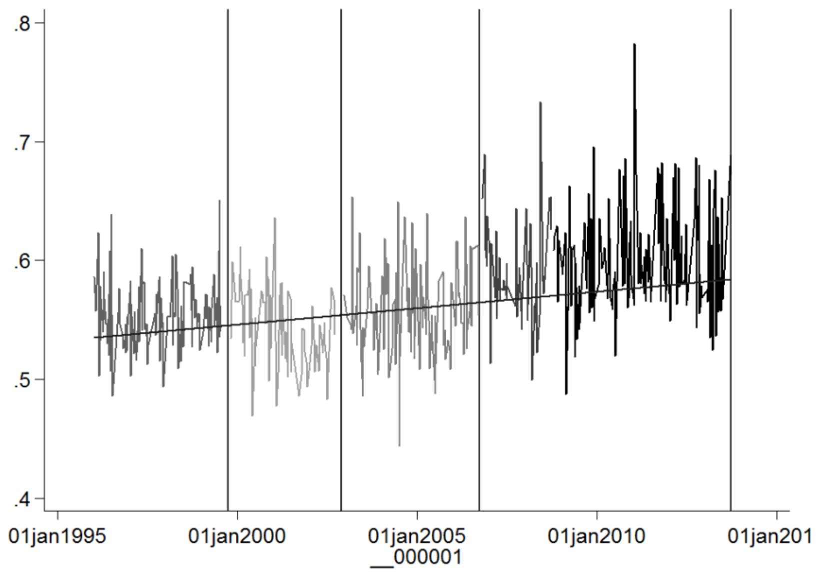
Does an upcoming election lead to more incivility in parliament? The graphical format of Figure 9 may not provide a good answer yet, but the answer is a tentative yes. However, the graph also reveals that incivility was above average in the aftermath of elections.

Figure 8: Mean incivility scores by party status and parliamentary role



The last hypothesis was that courteousness in parliamentary debates is on the way down and incivility is increasing? According to figure 10 the expectation appears confirmed.

Figure 9: Daily mean incivility, 1996-2013



Note: Vertical lines denote elections.

Conclusions

The paper describes a measurement procedure for establishing levels and degrees of incivility in parliamentary debates with the help of crowd-coded negativity scores as training data for a sentiment analysis with a supervised learning algorithm. Once trained we used that algorithm to measure incivility in parliamentary speeches given during the last five legislative terms of the Austrian parliament.

The empirical results presented here are preliminary results. We set up hypotheses where we had strong expectations to serve as validity checks for measurement procedure outlined in the first part of the paper. The preliminary results confirmed almost all our expectations and thus provide some reassuring evidence that our automated measurement provides valuable insights into incivility in the Austrian parliament during the past two decades. We have found mostly sensible variation at the level of individual speakers, across political roles, party status and other contextual factors.

The data patterns provide surprises, too. One of them is that the level of incivility actually dropped during the first ÖVP-FPÖ coalition government. The inauguration of the first 'black-blue' coalition in 2000 had faced a strongly critical public reaction and the new bipolar pattern of party competition between a right-wing government and a left-wing parliamentary opposition (Müller and Fallend 2004) could have led to an increase in parliamentary incivility. What we find instead is that the rhetorical taming of the FPÖ, after its dream of getting into government came true, overall had a stronger effect on the levels of parliamentary incivility.

We acknowledge that we need to do a more thorough validation of the measurement procedure by directly comparing incivility scoring from manual content analysis with the automated scores for a sample of parliamentary speeches (Grimmer and Stewart 2013). In the process we will also have to confront the issue of calibration head on and deal in more detail where to draw the boundary between ‘civil negative’ and ‘uncivil negative’ statements. In the current draft we have simply set the boundary at values above 2.5 in our five point negativity scale, when we devised the binary classifier.

References

Al-Rfou, Rami, Bryan Perozzi and Steven Skiena (2013). "Polyglot: Distributed word representations for multilingual NLP". In: Association for Computational Linguistics (Ed.), *Proceedings of the Seventeenth Conference on Computational Natural Language Learning* (pp. 183-192). Sofia, Bulgaria, August 8-9, 2013.

Ansolabehere, Stephen and Shanto Iyengar (1997). *Going negative: how attack ads shrink and polarize the electorate*. New York: Free Press.

Benoit, William L. 1999. *Seeing spots: A functional analysis of presidential television advertisements, 1952-1996*. Westport, CT: Praeger.

Berry, Jeffrey M. and Sarah Sobieraj (2013*). The Outrage Industry: Political Opinion Media and the New Incivility*. New York: Oxford University Press.

Brooks, Deborah J. And John G. Geer (2007). "Beyond negativity: The effects of incivility on the electorate". *American Journal of Political Science* 51 (1): 1-16.

Damore, David. F. (2002). "Candidate Strategy and the Decision to Go Negative". *Political Research Quarterly 55*(3): 669-685.

Dolezal, Martin, Laurenz Ennser-Jedenastik and Wolfgang C. Müller (forthcoming). „Who Will Attack the Competitors? How Political Parties Resolve Strategic and Collective Action Dilemmas in Negative Campaigning". *Party Politics*, Online first. doi: 10.1177/1354068815619832.

Fawcett, Tom (2006). "An introduction to ROC analysis". *Pattern Recognition Letters* 27: 861-874.

Fridkin, Kim L and Patrick J. Kenney (2011). "Variability in citizens' reactions to different types of negative campaigns". *American Journal of Political Science* 55 (2): 307-325.

Fridkin, Kim L. and Patrick J. Kenney (2008). "The dimensions of negative messages". *American Politics Research* 36 (5): 694-723.

Grimmer, Justin and Brandon M. Stewart (2013). "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts". Political Analysis 21 (3): 267-297.

Harris, Zellig (1954). "Distributional structure". *Word* 10 (23): 146-162.

Haselmayer, Martin and Marcelo Jenny (2017). "Friendly fire? Negative campaigning for electoral success and government participation". Working Paper.

Haselmayer, Martin and Marcelo Jenny Marcelo (forthcoming). "Sentiment analysis of political communication: Combining a dictionary approach with crowd-coding". *Quality & Quantity*. First online: 21 Sept 2016, 1-24. doi:10.1007/s11135-016-0412-4.

Heidar, Knut and Ruud Koole (eds.) (2000). *Parliamentary Party Groups in European Democracies: Political parties behind closed doors*. London: Routledge.

Herbst, Susan (2010). *Rude Democracy: Civility and Incivility in American Politics*. Philadelphia: Temple University Press.

Hopkins, Daniel J. and Gary King (2010). “A Method of Automated Nonparametric Content Analysis for Social Science”. *American Journal of Political Science* 54(1): 229-247.

Jamieson, Kathleen H. (1992). *Dirty politics*. New York: Oxford University Press.

Jenny, Marcelo and Wolfgang C. Müller, Wolfgang (1995). „Presidents of Parliament: Neutral Chairmen or Assets of the Majority”. In Döring, Herbert (ed.), *Parliaments and Majority rule in Western Europe*. Frankfurt: Campus Verlag, 326-364.

Jenny, Marcelo and Wolfgang C. Müller (2012). “Parliamentary roles of MPs in sharp and soft focus: Interviews and behavioural record compared”. In: Magnus Blomgren and Olivier Rozenberg (eds.), *Parliamentary Roles in Modern Legislatures*. London: Routledge, 145-161.

Kahn, Kim F. and Patrick J. Kenney (1999). „Do negative campaigns mobilize or suppress turnout? Clarifying the relationship between negativity and participation”. *American Political Science Review* 93 (4): 877-889.

Kahn, Kim Fridkin and Patrick J. Kenney (2004). *No holds barred: Negativity in US Senate campaigns*. Upper Saddle River, NJ: Prentice Hall.

Lau, Richard R. and Ivy B. Rovner (2009). “Negative Campaigning”. *Annual Review of Political Science* 12: 285-306.

Liu, Bing (2015). *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*. Cambridge: Cambridge University Press.

Mattes, Kyle and David P. Redlawsk (2014). *The Positive Case for Negative Campaigning*. Chicago: Chicago University Press.

Mikolov, Tomas, Ilya Sutskeyer, Chen Kai, Greg S. Corrado, and Jeff Dean (2013). “Distributed representations of words and phrases and their compositionality”. In: C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani and K. Q. Weinberger (eds.). *Advances in Neural Information Processing Systems* 26 (NIPS 2013), 3111-3119.

Müller, Wolfgang C. and Franz Fallend (2004). „Changing Patterns of Party Competition in Austria: From Multipolar to Bipolar System”. *West European Politics* 27 (5): 801-835.

Müller, Wolfgang C., Marcelo Jenny, Martin Dolezal, Barbara Steininger, Wilfried Philipp, and Sabine Westphal (2001). *Die österreichischen Abgeordneten: Individuelle Präferenzen und politisches Verhalten*. Wien: WUV Universitätsverlag.

Mutz, Diana C. (2007). Effects of “in-your-face” television discourse on perceptions of a legitimate opposition”. *American Political Science Review* 101 (4): 621-635.

Mutz, Diana C. and Byron Reeves (2005). “The new videomalaise: Effects of televised incivility on political trust”. *American Political Science Review* 99 (1): 1-15.

Pang, Bo, Lillian Lee and Shivakumar Vaithyanathan (2002). “Thumbs Up? Sentiment Classification Using Machine Learning Techniques”, In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (EMNLP-2002), Philadelphia, Pennsylvania, July 6-7, 2002, 79-86.

Stryker, Robin, Bethany Anne Conway and J. Taylor Danielson (2016). “What is political incivility?” *Communication Monographs* 83 (4): 535-556.

Walter, Annemarie S., Wouter van der Brug and Philipp van Praag (2014). „When the stakes are high: Party competition and negative campaigning”. *Comparative Political Studies* 47 (4): 550-573.