

# The Promise and Pitfalls of Automated Text-Scaling Techniques for the Analysis of Judicial Opinions

Arthur Dyevre  
KU Leuven  
Tiensestraat 45  
3000 Leuven  
Belgium  
arthur.dyevre@law.kuleuven.be

February 29, 2016

## Abstract

The present paper assesses the performances of text-scaling methods for the analysis of judicial opinions using a well-documented body of case-law: the 16 opinions rendered by the German Federal Constitutional Court on European integration. Unsupervised text-scaling (Wordfish) turns out to generate reliable results. Judicial position estimates on what is interpreted as a pro/anti-integration dimension are broadly in line with standard doctrinal accounts in EU law textbooks and law review articles. Meanwhile, results from supervised text-scaling (Wordscores) prove to be highly sensitive to the choice of reference texts. These findings, it is argued, demonstrate that the basic assumption underpinning the application of text-scaling techniques to party manifestos and legislative speeches—that word choice is driven by rhetorical considerations—is generalizable to the specialised form of political communication embodied in judicial opinions.

# 1 Introduction

Automated content analysis methods have enjoyed growing popularity among social scientists as part of the text-as-data movement. The surge of interest in computerised approaches stems in part from the increasing availability of politically relevant documents in electronic format. But it has also been spurred by the development of new methods that seem to hold great promise for social science research (Grimmer and Stewart, 2013). Among the techniques that have emerged over the last decade, two latent trait scaling algorithms, Wordscores (Laver et al., 2003) and Wordfish (Slapin and Proksch, 2008) have proved remarkably popular with political scientists. Scholars have used Wordscores to estimate party positions from Irish, British, French, US and Danish election manifestos (Laver et al., 2003, Klemmensen et al., 2007, Benoit and Laver, 2003, Laver et al., 2006, Kidd, 2008) and to measure the preferences of delegates at the Convention on the Future of Europe (Benoit et al., 2005). Wordfish, meanwhile, has variously served to measure the content of electoral pledges in German and Japanese elections (Slapin and Proksch, 2008, Proksch and Slapin, 2009, Proksch et al., 2011); to map the policy stances of interest groups at EU level (Klüver, 2009); and to analyse thousands of EU Parliament speeches (Proksch and Slapin, 2010).

Attempts to apply these techniques to the domain of law and courts, though, have been few and far between. A rare exception to the dearth of applications to the legal field is Evans et al. (2007). Their study evaluated the performances of Wordscores, along with that of various Naïve Bayesian classifiers, in recovering the ideological direction of litigant and amicus curiae briefs submitted in the *Bakke* (1978) and *Bollinger* (2003) affirmative action cases. But while such briefs typically invoke legal language and authorities, proving that text-scaling models work for the briefs is not quite the same as proving that they work for the judicial opinions they purport to influence. So, to date, McGuire and Vanberg (2005) remains the only attempt to apply computer-assisted text-scaling to the analysis of judicial opinions. This lack of interest in text-scaling approaches on the part of law and court scholars may, at first sight, come as a surprise. After all, the written word represents the primary manifestation of judicial behaviour. Judicial opinions are the instrument through which courts, in the course of settling disputes, make, announce and justify policies (Friedman, 2006, Lax and Cameron, 2007). While empirical research in the United States has traditionally focused on the voting behaviour of judges, there have been repeated calls for

judicial politics scholars to adopt a more comprehensive conception of judicial decision-making and, in particular, to pay more attention to what the judges say in their opinions (Lax, 2011, Epstein and Knight, 2013, Friedman, 2006). Automated text-scaling methods promise to offer interval-scale measures that, because they are based on the actual content of judicial opinions, are potentially more reliable than alternative methods seeking to infer the direction of opinions from concurrence behaviour or patterns of citations between cases (Clark and Lauderdale, 2010, Carrubba et al., 2012). That the relevant texts are now available at a few clicks for courts all around the world should have made automated text-scaling approaches more attractive still. Nor is their neglect reflective of a broader trend in the discipline to shun automated content analysis altogether. Indeed, a cursory look at the literature reveals a rapidly expanding body of research employing a varied range of techniques, including dictionary approaches (Corley and Wedeking, 2014, Owens and Wedeking, 2011, Hinkle et al., 2012), topic models (Lauderdale and Clark, 2014) along with applications relying on plagiarism software (Corley et al., 2011, Corley, 2008).

So why has this family of automated content analysis techniques failed to gain the favour of judicial scholars? Taking a closer look, three reasons may explain why judicial politics research has yet to embrace text-scaling approaches. First, because public support for the judicial branch is linked to the perception of judges as neutral and impartial guardians of the law (Gibson et al., 1998), judges seem to have a strong incentive to stick to legal, technical jargon and to steer clear of the more flowery rhetoric politicians commonly resort to in public debates. Whereas electoral competition gives partisan organizations an incentive to engage in product differentiation (Myerson, 1999, p. 676), judicial legitimacy builds on the belief that courts serve the interests of multiple, non-overlapping constituencies (Gibson et al., 1998). Political science studies using automated text-scaling typically presume that ideology is the dominant dimension in the textual data under analysis. But to the extent that this presumption is seen as more problematic in the legal context, automated methods look a comparatively less promising avenue for the study of judicial discourse. Another factor, more practical in character, pertains to validation. In party manifesto research, researchers can easily compare the estimates generated by the application of computerised methods to the results derived from expert surveys or from projects relying on human coding (see e.g. Klemmensen et al., 2007). In that sense, projects such as the Comparative Manifestos Project—whether or not one treats their

measurements as the political science equivalent of the gold standard—provide a useful benchmark, which has no direct equivalent in judicial politics. Validating judicial position estimates generated by computer-based approaches thus looks a more hazardous enterprise. Finally, the mixed results obtained by McGuire and Vanberg (2005) using Wordscores to measure the ideological content of US Supreme Court opinions in establishment of religion, free exercise of religion, and search and seizure cases may have contributed to chill initial excitement for text-scaling approaches; as did the fact that their paper never appeared in a professional journal.

The present paper suggests a more optimistic prospect for automated text-scaling applications to judicial reasoning. I assess the performances of Wordscores and Wordfish using a well-documented body of case-law: the 16 opinions rendered by the German Federal Constitutional Court on European integration. The unsupervised text-scaling approach—that is, Wordfish—turns out to produce convincing results, especially for the first 14 opinions of the German Court. Jurisprudential position estimates on the assumed pro/anti-integration dimension appear to be consistent with standard accounts encountered in EU law textbooks and law review articles. The analysis further demonstrates that supervised text-scaling—Wordscores—can also generate plausible estimates but that the performance and accuracy of the Wordscores algorithm is highly sensitive to the choice of reference texts. These results strongly suggest that the basic assumption underpinning latent traits scaling models for political texts—namely, that word choice is driven by rhetorical considerations—also holds for judicial discourse. Opinion-writing on a constitutional court, too, is a form of political communication, albeit a special one. While legalese are there to maintain the perception that judges are doing “law” and not politics, judges face the same incentive to present their policy positions in the best light and to choose frames that will appeal to their audiences. As the analysis shows, there are, of course, limitations in what text-scaling methods predicated on simplifying assumptions about human language can achieve. But these limitations result from characteristics inherent to the models rather than from their application to the judicial context.

The paper is organized as follows. Section 2 discusses the basic assumptions of latent traits models for textual data along with the general rationale motivating their application to legal reasoning. Section 3 then describes the substantive theoretical considerations that guided the choice of opinions and the text-scaling methodology employed to analyse them. Section 4

sets out the results. I conclude with some brief reflections on the future of computer-assisted text-scaling methods for the study of legal argumentation and judicial policy-making.

## **2 Judicial Reasoning as a Specialized Form of Political Communication**

### **2.1 Basic Assumptions of Computerised Text-Scaling Methods: Rhetorical Dimension of Political Communication**

Human language is an eminently complex phenomenon. While words represent its most conspicuous manifestation, they are, in fact, only one component of linguistic communication. Grammar and syntactic rules entail that two sentences containing the exact same words can enunciate radically distinct propositions. Consider, as a way of example, the two sentences “everyone loves someone” and “someone loves everyone”. Though the words they contain are perfectly identical, the two sentences have distinct truth conditions. That one is true does not entail that the other is as well. Adding to the complexity is the fact that linguistic meaning is heavily context-dependent. Depending on where and when it is uttered, the same sentence may be ascribed widely different meanings. The sentence “I would like a table”, for instance, will normally be understood differently when uttered in a restaurant and when uttered in a furniture shop. However, a speaker may defeat what would be, under given circumstances, the normal interpretation of such a sentence by producing an additional speech act. For example, upon entering a restaurant a speaker may specify that he or she would like “a table not to eat, but to buy it”. So the meaning actually ascribed to the uttered sentence will be determined not only by implicit background assumptions about the context but also by the sentence’s “co-text”, that is the other words and sentences uttered during the same conversation.

These are examples of simple, basic speech acts. So one can only begin to imagine the extraordinary difficulty involved in developing computer models able to parse the sentences of complex documents such as party manifestos and judicial opinions in a way that mimics the human brain. An achievement that, despite recent advances in natural language processing, continues to

elude computer linguists. The techniques popularized by the text-as-data movement, though, do not seek, nor claim to emulate all these aspects of linguistic meaning. Rather than attempting to model all the intricacies of human communication, they go the opposite path, reducing complexity in dramatic fashion. In short, texts are turned into sequences of word counts—what linguists call “bags of words”. These are then processed as numeric input in a statistical model, which in turn yields low-dimensional summaries of the texts as output. As a consequence, the procedure inevitably discards the syntactical and contextual information present in the original texts.

Similar to the other computer-based techniques that are on their way to become part of the standard toolkit of empirical political science, models of latent traits scaling for texts such as Wordscores and Wordfish do not only reduce complex textual structures to lexical frequencies but also assume conditional and positional independence in the word generation process. Conditional independence entails that, given a document  $D_i$ , the occurrence of word  $W_j$  has no effect on the occurrence of any other word in  $D_i$ . Positional independence means that the position of  $W_j$  in  $D_i$  is indifferent—whether it occurs at the beginning, the middle, or the end, has no influence, whatsoever, on the analysis. From a linguistic standpoint, there is no doubt that these assumptions are incorrect. Even in a relatively small corpus, the use of words tends to be highly correlated—use of “health”, for example, makes the occurrence of “care” more likely. Likewise, it is not hard to see why the position in which a word appears will often affect its meaning. Hence, inasmuch as they fail to provide an accurate account of the data-generating process, text-scaling models are justifiably said to be “wrong” (see Grimmer and Stewart, 2013). And yet both Wordfish and Wordscores have been showed to produce reliably accurate estimates of quantities of interest to students of politics from small as well as large text collections.

How come that text-scaling models work at all considering that the linguistic data so clearly violate their basic assumptions? The answer lies in the rhetorical nature of political communication. In a world in which politicians used the language of formal logic to articulate and contrast their policy positions, there would be little hope of recovering these unobserved positions from observed patterns of word usage. This, however, is not our world. Parties engaged in electoral competition do not oppose each other on an issue by issue basis using logical operators to differentiate their stances. Rather, they choose words with distinct connotations that emphasise opposing frames (Budge, 2001). Thus party manifestos on the left of the political spectrum

emphasise “equality” and “redistribution” while party manifestos on the right emphasise “freedom”, “entrepreneurship” and “taxes” (against).<sup>1</sup> The concept of “opposite frames” bears some stressing here. Right-wing parties do not counter the left’s emphasis on “redistribution” and “equality” by declaring themselves “against equality” and “contra redistribution”. Instead, party communication will usually emphasise a distinct frame of “lower taxes”, “small government” and “freedom”. Obviously, party strategists are aware that in order to attract voters they must be attentive to their choice of words. Computer-assisted text-scaling works precisely by exploiting differences in the words used to compose and identify these opposing frames. The political actors’ positions on the dimension of interest to the researcher are thus estimated on the basis of the relative patterns of emphasis detected in the texts (Lowe, 2013).

## **2.2 Motivation for Latent Traits Scaling Applications to Judicial Texts**

What motivates the application of text-scaling to judicial opinions is the recognition that judicial discourse, too, is a form of political communication. Judicial opinion-writing fulfils multiple functions. The most obvious one is to give reasons for the court’s decision on the merits. The opinion is there to explain and justify why one litigant lost and the other won. But opinion-writing, particularly on peak courts, also performs a policy making function. The opinion accompanying the case disposition typically articulates (often elaborating from previous opinions) a “doctrine”—which is the name by which lawyers commonly refer to judicially enacted policies—indicating how the judges intend to treat similar cases in the future. This doctrine then serves as a policy signal to courts lower down the judicial hierarchy as well as to future litigants and other policy makers, including sometimes international tribunals. Opinions, in other words, are what enables judges to make policies in the course of settling individual disputes. Save when the case disposition simultaneously decides the outcome of a hotly contested presidential election, the “opinions themselves, not who won or lost, are the crucial form of political behaviour by the appellate courts, since it is the opinions which provide the constraining directions to the public and

---

<sup>1</sup>Consistent with that picture of political language is the finding in Laver et al. (2003) that “tax” correlates with more right-leaning parties.

private decision-makers who determine 99 per cent of the conduct that never reaches the courts” (Shapiro, 1968, 39). Now, just as other public decision makers, judges must communicate their policy determinations in a way that is conducive for social cooperation. While certain constraints arise from public expectations about what constitutes legitimate judicial behaviour, judges thus face the same incentive to present their policy determinations in the best possible light. In other words, the third function of judicial opinion-writing is rhetorical. Composing, a judicial opinion, like composing a party manifesto or drafting an executive or legislative press release, is about conveying policy information. But as is the case with other political texts, judicial opinion-writing is an exercise in persuasion, too. This makes it reasonable to assume that variations in word usage will not merely reflect topical structures but also changes in judicial positioning over the same latent dimension.

As some will be quick to object, the bag-of-words approach on which text-scaling models rely cannot distinguish a judicial opinion asserting “we stand for the rule of law and against democracy” from an opinion asserting “we stand for democracy and against the rule of law”. Yet we can reasonably assume that few judges will want to articulate their jurisprudential positions in such manner, which in turn warrants the expectation that cases where changes in jurisprudential policy will be solely reflected in changes in the order of words will be rare. Rather than inverting the order of words, judges articulating distinct positions will generally go for words associated with distinct frames. They will use negative frames to characterize positions they reject and positive ones to present their own policies.

Another objection against the application to the judicial domain of methods primarily designed for the analysis of partisan texts arises from the judges’ imperative need to sound “legal”. Judicial discourse differs from other forms of political communication—i.e. communication about the exercise of public authority—in that judges are expected to justify their policies in legal terms, employing legal reasoning. The peculiarly judicial necessity to appear “legal” means judges have an incentive to lard their opinions with references to precedents, legislative materials, legal instruments and other legalese to fend off the charge of political usurpation. This particular feature of judicial discourse may represent a challenge for text analysis applications that seek to recover latent policy positions from the judges’ words, as we might expect the legalistic noise to obscure the policy signal. We have reasons to believe, though, that the supervised and unsupervised text-scaling techniques applied in the present research can handle this problem relatively



well. To the extent that legalese are context-invariant, they will be treated as other uninformative words, such as grammatical words and conjunctions (e.g. “the”, “a”, “but” in English, or “der”, “die”, “das”, “darum” in German). On the other hand, when legalese vary significantly in their frequency across judicial texts, we may assume this to reflect changes in the latent policy position (McGuire and Vanberg, 2005). As empirical studies have shown (Benesh and Czarnezki, 2009, Corley et al., 2005), judges often invoke methods and authorities in ad hoc fashion, picking and choosing those which best suit the outcome they have already reached. In such circumstances, far from constituting noise, we should expect legalese to have directional association with policy positions. Having said this, the proof of the methodological pudding must lie, ultimately, in the eating. Consequently, the next two sections move on to assess the empirical performances of the two most popular text-scaling models, Wordscores and Wordfish, for the analysis of a well-known body of case-law.

### **3 Data and Methodology**

#### **3.1 Textual Data: The German Federal Constitutional and European Integration, 1967-2012**

Given the simplifying assumptions on which latent traits scaling models rest and the need for validation, the set of opinions used to assess the performances of text-scaling models should meet the following criteria. First, to control for the influence of context, which, as seen above, is ignored by the models, the dimension of interest should be dominant in the texts. To the extent that the objective of the researcher is to situate the texts on a policy space, this entails that the texts should all focus on this policy issue. Second, to facilitate validation and to assess the overall plausibility of the results, the policy positions expressed by these texts should be well-known. Ideally, there should be scholarly consensus over the location of these texts on the policy dimension of interest. Third, in order to assess the ability of the models to track changes in judicial positioning, the texts should exhibit a substantial amount of variability in the policy positions they express. Finally, because automated content analysis methods are known to perform poorly when the texts are few and short, the opinions should be relatively long and numerous.

Finding a set of judicial opinions meeting all these criteria is all but easy.

One reason is that, owing to variations in writing and publication practices, not to mention language, we cannot easily compare opinions across separate judicial institutions. The opinions must come from the same tribunal. This means we must look for a policy area where the same tribunal has rendered many salient decisions articulating diverging policy positions over time. The nearest we could find to a corpus meeting these properties are the 16 opinions issued by the German Federal Constitutional Court on European integration. The Court’s jurisprudence on the relationship between EU law and German law spans five decades, from the late 1960s to the present. Its rulings on the issue have received considerable attention from legal scholars both in and outside Germany. It is also widely recognised that the German Court has oscillated between integration-friendly stances and a more defiant, Eurosceptic posture (Davies, 2012, Möllers, 2011, Mayer and Walter, 2011, Stein, 2011, Alter, 2001). While the rulings have differed in length (see Figure 1) as well as in saliency, it seems reasonably safe to assume that European integration and its implications for German constitutional law constitutes the dominant policy dimension in the opinions issued by the Court.

**Figure 1 about here**

I used scholarly accounts to identify the first 14 rulings. The 14 rulings are those legal scholars cite when they chart the evolution of the GFCC’s position (see e.g. Stein, 2011, Meessen, 1994). The first opinion goes as far back as 1967 (*EWG-Verordnung*), while the last one is as recent as 2010 (*Honeywell*).<sup>2</sup> Later opinions explicitly cite earlier ones, suggesting that the Court itself regards them as belonging to the same line of precedents.<sup>3</sup> Court watchers see the Court of Justice of the European Union (CJEU) as the primary target audience of this particular set of GFCC opinions. The German rulings are central to a whole strand of (normatively oriented) EU law scholarship that stresses the importance of judicial dialogue and constitutional pluralism in determining the limits, pace and direction of legal integration in Europe (Bogdandy, 2008, Baquero Cruz, 2008). Eurosceptic rulings are typically described as “warning shots”, whereas more integration-friendly are characterised as peace offerings or, somewhat less charitably, as surrender to

---

<sup>2</sup>GFCC decisions are commonly referred to by (non-official) terms meant to describe the main issue at stake or the decision’s central finding rather than by the names of the parties to the case.

<sup>3</sup>Note, though, that the German legal system does not have an explicit doctrine of *stare decisis*.

the European Court (Dyevre, 2011, Schmid, 2001, Stein, 2011).<sup>4</sup>

Together with these 14 opinions, I also included the two more recent rulings rendered by the GFCC on the Euro crisis: the 2011 opinion on the European Financial Stability Facility (*EFSSF*) and the 2012 decision on the European Stability Mechanism (*ESM*). Note, however, that, although there is no doubt that these two decisions also address the issue of European integration, they seem to reflect an appreciable agenda shift. In contrast to the first 14 opinions, which seemed primarily addressed to the CJEU, it did not look as if judicial dialogue was really at stake in the decisions over the Euro crisis (Wendel, 2013).

Although concurring and dissenting opinions are allowed,<sup>5</sup> GFCC opinions are rendered *per curiam* “In the name of the people” (*Im Namen des Volkes*) so that authorship is not formally linked to a specific judge. Yet in practice the Court’s rules of procedure entrust opinion-writing to a rapporteur-judge (*Berichterstatter*), who is primarily designated on the basis of criteria of substantive specialization. Relatively little is known about his or her influence on the final draft (see Kranenpohl, 2009). GFCC opinions divide into two parts. The first summarizes the facts, prior proceedings and arguments of the parties, whereas the second (beginning after the letter B) considers the justiciability and merits of the case. Since my focus is on the policy position of the Court, I elected to restrict the scope of my content analysis to the latter.

### 3.2 Supervised and Unsupervised Approaches to Text-Scaling

As for the methodology, Wordfish and Wordscores are based on the same basic principle of low dimensional reconstruction of relative patterns of emphasis (Lowe, 2013). Aside from this conceptual similarity, however, the two methods diverge significantly in their mathematical structure and practical implementation. Wordfish is based on Item Response Theory (IRT), which underpins a family of statistical models that originated in the field of educa-

---

<sup>4</sup>A formal theoretic account of legal integration conceptualizes the rulings as a means for the GFCC to signal its policy preferences to the CJEU in a dynamic issue-trading equilibrium (Dyevre, 2014).

<sup>5</sup>Several of the decisions included in our analysis came with a concurring opinion. Because these opinions have attracted less scholarly attention, thus making validation more difficult, I elected to exclude them from the estimation.

tional studies but are now commonly used in the analysis of voting behaviour in legislative assemblies and judicial bodies (see e.g. Hanretty, 2012). Also, contrary to Wordscores, Wordfish works without human supervision. This means that the researcher does not need to identify reference texts, nor to assign them specific scores. Instead, given a minimal constraint, the model is designed to identify the dominant, latent dimension in the textual data. More specifically, the count  $C$  of word type  $W_j$  in document  $D_i$  is modelled as a Poisson process:

$$\begin{aligned} C_{ij} &\sim \text{Poisson}(\lambda_{ij}) \\ \log\lambda_{ij} &= \alpha_i + \psi_j + \theta_i\beta_j \end{aligned} \tag{1}$$

where  $\alpha$  is a document-level parameter for loquaciousness, basically ensuring that results are not merely driven by differences in length among the documents.  $\psi$  is a measure of word fixed effect, reflecting the extent to which a word type occurs in the same proportion across the documents under analysis.  $\beta$  and  $\theta$  are the main quantities of interest, with  $\beta$  capturing the degree to which a word type discriminates among the documents and  $\theta$  representing the document’s latent position on the identified dimension.<sup>6</sup>

Unlike Wordfish, Wordscores operates through human supervision, as a machine learning algorithm. Wordscores assumes no distributional form and represents an approximation of correspondence analysis (Lowe, 2013, 2008). As supervised method, Wordscores presupposes that the researcher has a priori knowledge about the position of at least two reference texts on the dimension of interest. The Wordscores algorithm then generates word scores for all the word types occurring in the reference texts. Finally, these word scores are used to estimate the positions of the remaining texts—the “test set” in computer-linguistic parlance or “virgin texts” in the initial introduction of the Wordscores approach (Laver et al., 2003). Formally, word score for word type  $W_i$  from reference document  $D_r$  with known position  $\theta_r$  is computed as:

$$W_i^{\text{Wordscore}} = \sum_r^R \theta_r Pr(D_r|W_i) \tag{2}$$

Where:

$$Pr(D_r|W_i) = \frac{Pr(W_i|D_r)}{\sum_r^R Pr(W_i|D_r)} \tag{3}$$

---

<sup>6</sup>Parameter values are jointly estimated by alternating conditional maximum likelihood.

The score of test document  $D_t$  is computed as the average of the scored word types it contains weighted by their relative count  $C$ :

$$D_t^{Textscore} = \frac{1}{C_t W} \sum_i^N W_i^{Wordscore} * C_{it} \quad (4)$$

Similar to the  $\alpha$  parameter in the Wordfish model, weighting the average of the scored word types by their relative count in  $D_t$  ensures that textscores are not influenced by document length. Over Wordfish Wordscores has the disadvantage that it is very sensitive to the choice of reference documents, which may in part explain the disappointing results obtained by McGuire and Vanberg (2005). To compute the textscores of the documents in the test set, Wordscores relies exclusively on the word types occurring in the reference texts. For that reason, words that do not appear in the reference texts, but do in the test documents, are all treated as noise.<sup>7</sup> Consequently, in the analysis that follows I conduct separate estimations using different reference texts and compare the results with those generated by Wordfish.<sup>8</sup>

### 3.3 Validation

Because there is no absolute guarantee that the estimates generated by the automated method really captures the quantity of interest to the researcher, validation is required. Applications of text-scaling techniques to political texts have variously relied on manually-coded measures, expert surveys (Klemmensen et al., 2007) and human benchmarking (Lowe and Benoit, 2013) to validate their estimates. My validation procedure here is exceedingly simple: I simply compare the estimates generated by the models to what legal commentators said about the decisions.

---

<sup>7</sup>In some contexts, though, this weakness can turn into a strength. Precisely because the reference texts are critical to the results of the Wordscores approach, it is less vulnerable to the curse of dimensionality. Wordfish, by contrast, is designed to pick the dimension that explains the largest amount of variance in the data as a whole. This, however, may or may not be the dimension of interest to the researcher.

<sup>8</sup>Wordfish and Wordscores as implemented in the Austin package for R.

## 4 Results

### 4.1 Unsupervised Text-Scaling

After downloading the opinions,<sup>9</sup> I manually removed the arguments of the parties, converted the documents into plain text files, eliminated numbers and converted all characters to lower case before turning the words into a term frequency matrix. I first ran the Wordfish model with a minimum amount of pre-processing.

#### Figure 2 about here

Plotted in Figure 2 are the resulting estimates for the first 14 opinions. These turn out to be broadly consistent with scholarly accounts of the GFCC’s fluctuating jurisprudential stance. The first *Solange* opinion was the GFCC’s first warning shot to the Court of Justice. It stipulated that, as long as the European Communities lacked a judicially enforced catalogue of fundamental rights, German citizens could challenge the acts emanating from the supranational institutions before the GFCC and the GFCC would accept to consider the merits of their claim (Meessen, 1994, 512, Stein, 2011, 220, Stone Sweet, 2004, 88, Davies, 2012). In the late 1970s and early 1980s, the German Court issued a string of equivocal rulings, including the aptly named “*Vielleicht*” (German for “maybe”) decision, thereby suggesting that it might be willing to reconsider the *Solange* holding (Meessen, 1994, 513, Stein, 2011, 221). Then, twelve years after *Solange I* came the second *Solange* decision. To the extent that the GFCC announced that it would no longer consider the merits of cases brought against Community acts for their compatibility with basic German fundamental rights standards, the opinion was viewed as more integration-friendly (Meessen, 1994, 513). Yet acceptance of integration came with an important qualification. Indeed, while acknowledging the efforts made by the Court of Justice to develop a human rights jurisprudence, the GFCC made it clear that it would stop reviewing EU acts only as long as the European judges provided a level of human protection comparable to that of Germany. In so doing, the *Solange II* opinion preserved much of the language of the first, notably the “*solange*” (as long as) formula (Meessen, 1994, Stein, 2011, Stone Sweet, 2004, Alter, 2001). The promixity of the two *Solange* opinions in terms of both language and doctrine would seem to explain why their positional estimates are close to one

---

<sup>9</sup>From <http://www.servat.unibe.ch>.

another. Also consistent with the perception of the legal commentariat is the position assigned by the model to the *Kloppenburg* opinion. Legal scholars described *Kloppenburg* as the “high water mark” of the pro-European trend initiated by *Solange II* (Meessen, 1994, 513).

Equally, the sharp upturn that follows *Kloppenburg* in Figure 2 as a result of the markedly Eurosceptic position attributed to the *Maastricht* opinion is in line with the comments found in EU studies. Looking back at the ruling a decade later, Alec Stone Sweet, for one, notes that it “sent shock waves through the Community” (Stone Sweet, 2004, 92). At the time, pro-integration scholars vehemently criticized the opinion for its threatening language and nationalistic overtones (Meessen, 1994, Weiler, 1995). Even so, if *Maastricht* seemed to mark the commencement of a cold war with the European Court, the *Banana* decision at the turn of the millenium appeared to announce a new period of détente (Stein, 2011, Alter, 2001). Such was the contrast with the belligerent tone of *Maastricht* that legal academics were beginning to see the case law of the German Court as incoherent (Schmid, 2001, Stein, 2011). Judging, again, by what the legal literature says about the ruling, the Wordfish approach appears to make an equally good job at estimating the position of last three rulings in this set of cases. When tensions with the CJEU resurfaced in the mid-2000s, the language employed by the GFCC in its ruling on the European Arrest Warrant hinted that it was going to toughen its line once again (Satzger and Pohl, 2006). Tensions with the CJEU culminated in the GFCC *Lisbon* decision. The much-awaited opinion was received—as much if not more than *Maastricht*—as a sign that the GFCC had definitely turned against integration (Tomuschat, 2010, Dyevre, 2011, Payandeh, 2011, Stein, 2011). Pro-European academics lamented the opinion’s emphasis on German statehood and sovereignty and the limits it placed on the delegation of authority to EU institutions (Grosser, 2009, Möllers, 2011, Thym, 2009). Barely a year later, though, the GFCC made yet another about-face with its *Honeywell* ruling, whose EU-friendly tone took many a commentator by suprise (Mayer and Walter, 2011, Stein, 2011).

As it turns out, the only estimate that really looks out of tune with the standard doctrinal narrative of the Court’s case law is the one corresponding to the first opinion in the series, the 1967 *EWG Verordnung* opinion. This is usually described as a more integration-friendly ruling than the first *Solange* decision (Stein, 2011, Alter, 2001, 220). A careful examination of the 1967 opinion suggests that the problem may stem from the peculiar manner in which the GFCC cast the issue at stake. Indeed, whereas subsequent opinions

would cast the issue as one about the relationship between EU law and domestic (constitutional) law, the 1967 opinion dismissed the claim brought against the EEC regulation on the grounds that EEC regulations were not acts of a “German public authority” and as such fell outside the scope of the GFCC’s review powers.

**Figure 3 about here**

To help spot the words that drive the outcome, Figure 3 plots the 9693 word types occurring in the corpus against their  $\psi$  (word fixed effect) and  $\beta$  (word weight) parameter values. The resulting word cloud exhibits the expected Eiffel-tower shape. At the top, we find the words that tend to appear in the same proportion across all opinions. Among these are declensions of the definite articles “*der*”, “*die*”, “*das*”, along with basic connectors and prepositions such as “*aber*”, “*durch*”, “*wegen*”. Near the apex of the word cloud are also legalesees such as “*Verfassung*”, “*Art*” (short for “*Artikel*”) and “*GG*” (short for “*Grundgesetz*”). As expected, GFCC opinions are peppered with legalese, which contribute to make them sound “legal”. Yet these terms turn out to have little impact on the estimation. The model effectively filter them out as noise.<sup>10</sup> As for the words that really determine the positional estimates depicted in Figure 2, we find those at the the feet of the Eiffel tower. These words have high absolute  $\beta$  values—or weight. Words with negative  $\beta$  value make decisions more integration-friendly, those with positive  $\beta$  value more Eurosceptic.

**Figure 3 about here**

Table 1 provides examples of words associated with more Eurosceptic opinions and words correlating with more integration-friendly ones.

Broadly speaking, more Eurosceptic opinions coincide with a treaty-state frame. Words such as “*Volkes*” (people), “*Herrschaft*” (authority), or “*Streitkräfte*” (armed forces) are part of the vocabulary of the nation-state. This is precisely the sort of language in the *Maastricht* and *Lisbon* decisions that drew heavy criticism from pro-integration scholars (Möllers, 2011, Weiler, 1995). More startling, at first sight, is that many Eurosceptic words appear to borrow from the language of European Treaties: “*EUV Lissabon*” (EU Treaty of Lisbon), “*Unionsbürger*” (Union citizen), “*Brückenklauseln*” (passerelle clauses), etc. This is less surprising than it seems, though. In-

---

<sup>10</sup>Noise, though, is in the eye of the beholder. Identifying the constants of judicial discourse—i.e. the context-invariant markers a court uses to sound “legal”—may in itself be a valuable contribution to the study of law and courts.



Table 1: European Integration and Judicial Frames

Top 20 Eurosceptic Words	Top 20 Integration-Friendly Words
<i>EUV Lissabon</i> (Lisbon Treaty)	<i>Intervention</i> (intervention)
<i>Volkes</i> (people)	<i>Rückwirkung</i> (retroactive effect)
<i>Lissabon</i> (Lisbon)	<i>Durchführung</i> (application)
<i>Unionsbürger</i> (Union citizen)	<i>Marktstörungen</i> (market distortions)
<i>ordentlichen</i> (ordinary)	<i>Vorratstelle</i> (warehouse)
<i>Einsatz</i> (use)	<i>Getreide</i> (cereals)
<i>Streitkräfte</i> (armed forces)	<i>Interventionspflicht</i> (intervention duty)
<i>Brückenklauseln</i> (passerelle clauses)	<i>Bundesgerichtshof</i> (Federal Supreme Court)
<i>Begleitgesetzgebung</i> (implementing legislation)	<i>Interventionskäufe</i> (price support)
<i>Organstreit</i> (abstract review)	<i>Uhr</i> (hour)
<i>Einfluss</i> (influence)	<i>Einfuhr</i> (import)
<i>Verlust</i> (loss)	<i>bzw.</i> (respectively)
<i>Teilhabe</i> (participation)	<i>Rechtsstellungen</i> (legal status)
<i>repräsentativen</i> (representative)	<i>Handelsplätze</i> (trade centres)
<i>Brückenverfahren</i> (passerelle procedure)	<i>Interventionsbeschränkung</i> (intervention limit)
<i>speziellen</i> (special)	<i>Marktorganisationen</i> (market structures)
<i>Herrschaft</i> (authority)	<i>Rechtsverordnungen</i> (regulations)
<i>Gesetzgebungsorgane</i> (legislative organs)	<i>Beschwerdeführerinnen</i> (litigants)
<i>Änderung</i> (revision)	<i>Rechtsansprüche</i> (legal claims)

deed, the more Eurosceptic opinions seek to reassert German statehood in the face of deepening integration represented by the Treaties. The integration-friendly opinions, meanwhile, exhibit a distinct, internal market frame. This is reflected in words like “*Vorratstelle*” (warehouse), “*Marktstörungen*” (market distortions), “*Getreide*” (cereals), “*Intervention*” (intervention). This being said, Table 1 suggests that the model does pick some noise. Among the top 20 integration-friendly words appear “*Bundesgerichtshof*” (Supreme Federal Court), “*Uhr*” (hour) and “*bzw.*” (short for “respectively”). These make little sense as integration-friendly words. More worrying, though, is the fact that the model gives high weight to words that must, for *structural* reasons, be absent from some opinions. This applies for “*Lissabon*”, for example. Obviously, the city of Lisbon had little relevance for European integration when the GFCC rendered its first *Solange* ruling in 1974, as Portugal had not even entered the supranational club. So the word’s absence cannot be attributed to its author’s choice.

Applications of computerised text-scaling techniques address this common problem at the pre-processing stage of the textual analysis. Standards practices include: removing uninformative word types, also known as “stop-words”; reducing variants of the same lexeme to one and the same stem (“constitutional”, “constitutionality” and “constitutions” thus become “constitution”)<sup>11</sup>; and eliminating words with extremely low frequency. These procedures all proceed by discarding data. But they are supposed to render the estimation more robust by making it less dependent on rare words or

<sup>11</sup>Stemming approximates what linguists call lemmatization, the reduction of variants of a lexeme to its base form or dictionary entry.

rare variants of more common words.

#### Figure 4 about here

Figure 4 illustrates the effect of various pre-processing specifications on the estimation. Stemming (Panel A) alone, or when combined with stop-word removal (Panel B), turns out to have no noticeable impact on the positional estimates. When we add trimming, though, the story is somewhat different, as can be seen in Panel C. Eliminating words occurring less than three times or appearing in only one document shrinks the number of word types to a little over 2,500. When the estimation is run with this configuration of the textual data, the pattern of anti- and pro-integration opinions remains fundamentally the same. The 1998 opinion on the Economic and Monetary Union (EMU), though, becomes the most Eurosceptic opinion, while *Maastricht* scores higher than *Lisbon*. Interestingly, this configuration hardly changes the kind of word types associated with integration-friendly opinions. Words with high, negative weight still relate to the internal market.<sup>12</sup> Yet the word types with high positive weight now define a distinct monetary union frame with lexemes such as “*Währung*”, “*Geld*”, etc. This suggests that the model does, after all, pick some topical structure. But the explanation could be that we are dealing with two overlapping dimensions which differ only in the way they express disagreement with European integration. The first emphasises statehood and is dominant in the untrimmed word frequency matrix. The second emphasises the implications of the currency union for Germany and becomes dominant only in the trimmed dataset. This hypothesis is consistent with the positional estimates generated by the analysis of the 16 opinions.

#### Figure 5 about here

Depicted in Figure 5 are the position estimates for all 16 opinions. As far as the first 14 opinions are concerned, the evolution of the GFCC’s case law appears largely unchanged. More noticeable is that *Lisbon* no longer scores as the most Eurosceptic opinion and is superseded by the *EFSSF* and *ESM* decisions. In light of the comments that these decisions have attracted, though, these positional estimates seem to exaggerate the degree of Euroscepticism of these two rulings. Whether *EFSSF* was more Eurosceptic than *Lisbon* or less so is a question that it is difficult to settle on the basis of the legal literature (Schmidt, 2013, Wendel, 2013). But we should expect *ESM* to score as less Eurosceptic than *Lisbon*. On the other hand, the estimation seems to have the move right: after *Honeywell* the GFCC shifted to a more Eurosceptic

---

<sup>12</sup>We find “*Einfuhr*”, “*Ernährung*”, “*Erzeugnisse*” ...

position with *EFSF* before reverting to a more integration-friendly stance in *ESM* (Wendel, 2013, Schmidt, 2013). As both *ESM* and *EFSF* dealt with the common currency, this suggests that it is the second of the two dimensions of Euroscepticism that Wordfish captures in this analysis. A look at the words' parameter values supports this hypothesis, with words related to the currency union scoring highest on the  $\beta$  parameter.

## 4.2 Supervised Text-Scaling

I now turn to Wordscores. As reference texts for the Wordscores estimation I picked out the *EWG* and *Maastricht* opinions. Consistent with the doctrinal accounts discussed above, I rated *EWG* -1 and *Maastricht* +1 on Euroscepticism. These two opinions served to generate the word scores that I then used to estimate the positions of the remaining 14 opinions. Remember that the procedure ignores the word types that do not appear in the reference texts. With *EWG* and *Maastricht* as reference texts, this results in only 2557 (or 23.05 per cent) of the 11,095 word types appearing in the 14 opinions being scorable. The corresponding positional estimates are shown in Figure 6.

### Figure 6 about here

Compared to Figure 5, Figure 6 is certainly harder to reconcile with what legal scholars and others have said about the GFCC's case law. *Solange I* scores as considerably less Eurosceptic than *Solange II*, and even *Kloppenburg*. Likewise, though Wordscores appears to get the sequence of jurisprudential moves mostly right for the decisions that came after *Maastricht*, it appears to underestimate the shock that *Lisbon* and *Maastricht* represented in the evolution of the Court's stance on European integration. Owing to the way in which Wordscores computes test document scores, it is harder to determine what particular words drive the results. But we may doubt that the estimation truly captures Euroscepticism. However, using a different set of reference texts, in that case *Lisbon* and *Honeywell*, we obtain estimates that more closely resemble those generated by Wordfish.

### Figure 7 about here

Rating *Lisbon* +1 and *Honeywell* -1, a higher share of scorable words in the test set (49.34 per cent) suggests that the Wordfish algorithm comes closer to capture the dominant dimension in the corpus. As can be seen in Figure 7, the resulting estimates closely track those produced by Wordfish. In keeping with doctrinal accounts, the estimation clearly ascribes a more integration-friendly stance to *Solange II* and *Kloppenburg* than to *Solange I*.

Only in minor details does the estimation appear to underperform Wordfish: *Maastricht* comes out as markedly more Eurosceptic than *Lisbon* and the *EFSF-ESM* sequence is wrong. For the rest, the estimation repeats the mistake of classifying the *EWG* opinion as Eurosceptic.<sup>13</sup>

## 5 Conclusion

The foregoing analysis suggests that judicial scholars are wrong to shun the text-scaling techniques made popular by the text-as-data movement, even though these were originally designed for the analysis of explicitly political texts such as party manifestos and legislative speeches rather than for the study of judicial opinions. As I have argued, judicial opinion-writing, too, constitutes a form of political discourse. Also, I hope to have demonstrated the potential contribution that latent traits scaling models can make to judicial politics research. Wordfish and Wordscores can generate judicial position estimates that are surprisingly reliable when compared with the accounts appearing in legal scholarship. Besides helping researchers in reconstructing the evolution of judicial doctrines over time, text-scaling approaches may also shed light on how judicial communication operates in practice. Treating judicial discourse as a specialised form of political communication, researchers may begin to investigate the context-dependent aspects as well as the constants of judicial reasoning. But this assessment also highlights the limitations of these popular text-scaling tools. This does not only apply for the choice of reference texts in Wordscores. Obviously, the models pick too much noise to obviate the need for external validation. Nor are they designed in a way that makes it easy to discern topical information from positional signal. Moreover, we are dealing with a methodological approach that is still in its infancy. Popular as Wordfish and Wordscores may have been with political scientists, better techniques may already exist to perform the same task. Thus, before turning to Wordfish and Wordscores, students

---

<sup>13</sup>I also compared the results of the two models rating the reference texts based on the Wordfish estimates. So I took the two opinions identified by the untrimmed Wordfish model as occupying the two extremes of the pro/anti integration scale, *Lisbon* and *Rückwirkende Richtlinie*. I then plugged their Wordfish value into the Wordscores algorithm to generate word scores. Finally, based on these word scores, I estimated the positions of the remaining fourteen opinions. The correlation between the two methods proved high, with Pearson  $\rho = .89$  and (as a measure of the monotonicity of the correlation) Spearman  $\rho = .88$ .

of law and courts might want to consider methods more explicitly designed to handle multi-dimensional data, such as multiple correspondence analysis or canonical correlation analysis (Lowe, 2013).

## References

Alter, Karen J. *Establishing the supremacy of European law: The making of an international rule of law in Europe*. Oxford University Press Oxford, 2001.

Baquero Cruz, Julio. “The Legacy of the Maastricht-Urteil and the Pluralist Movement.” *European Law Journal* 14, 4: (2008) 389–422. <http://onlinelibrary.wiley.com/doi/10.1111/j.1468-0386.2008.00419.x/abstract>.

Benesh, Sara C., and Jason J. Czarnecki. “Ideology of Legal Interpretation, The.” *Wash. U. J. L. & Pol’y* 29: (2009) 113. <http://heinonline.org/HOL/Page?handle=hein.journals/wajlp29id=115div=collection>

Benoit, Kenneth, and Michael Laver. “Estimating Irish party policy positions using computer wordscoring: the 2002 election – a research note.” *Irish Political Studies* 18, 1: (2003) 97–107. <http://dx.doi.org/10.1080/07907180312331293249>.

Benoit, Kenneth, Michael Laver, Christine Arnold, Paul Pennings, and Madeleine O. Hosli. “Measuring National Delegate Positions at the Convention on the Future of Europe Using Computerized Word Scoring.” *European Union Politics* 6, 3: (2005) 291–313. <http://eup.sagepub.com/content/6/3/291>.

Bogdandy, Armin von. “Pluralism, Direct Effect, and the Ultimate Say: On the Relationship Between International and Domestic Constitutional Law.” *Int’l J. Const. L.* 6: (2008) 397–413.

Budge, Ian. “Validating Party Policy Placements.” *British Journal of Political Science* 31, 01: (2001) 179–223.

Carrubba, Cliff, Barry Friedman, Andrew D. Martin, and Georg Vanberg. “Who Controls the Content of Supreme Court Opinions?” *American Journal of Political Science* 56, 2: (2012) 400–412. <http://onlinelibrary.wiley.com/doi/10.1111/j.1540-5907.2011.00557.x/abstract>.

- Clark, Tom S., and Benjamin Lauderdale. “Locating Supreme Court Opinions in Doctrine Space.” *American Journal of Political Science* 54, 4: (2010) 871–890. <http://onlinelibrary.wiley.com/doi/10.1111/j.1540-5907.2010.00470.x/abstract>.
- Corley, Pamela C. “The Supreme Court and Opinion Content The Influence of Parties’ Briefs.” *Political Research Quarterly* 61, 3: (2008) 468–478. <http://prq.sagepub.com/content/61/3/468>.
- Corley, Pamela C., Paul M. Jr. Collins, and Bryan Calvin. “Lower Court Influence on U.S. Supreme Court Opinion Content.” *The Journal of Politics* 73, 01: (2011) 31–44.
- Corley, Pamela C., J. D. Robert M. Howard, and David C. Nixon. “The Supreme Court and Opinion Content: The Use of the Federalist Papers.” *Political Research Quarterly* 58, 2: (2005) 329–340. <http://prq.sagepub.com/content/58/2/329>.
- Corley, Pamela C., and Justin Wedeking. “The (Dis)Advantage of Certainty: The Importance of Certainty in Language.” *Law & Soc’y Rev* 48, 1: (2014) 35–62. <http://onlinelibrary.wiley.com/doi/10.1111/lasr.12058/abstract>.
- Davies, Bill. *Resisting the European Court of Justice*. Cambridge: Cambridge University Press, 2012. <http://ebooks.cambridge.org/ebook.jsf?bid=CB09781139170024>.
- Dyevre, Arthur. “The German Federal Constitutional Court and European Judicial Politics.” *W. Eur. Politics* 34, 2: (2011) 346–361.
- . “Domestic Judicial Defiance in the European Union: A Geostrategic Map of the EU Judiciary.” Washington D.C.: Social Science Research Network, 2014. <http://papers.ssrn.com/abstract=2455301>.
- Epstein, Lee, and Jack Knight. “Reconsidering Judicial Preferences.” *Annual Review of Political Science* 16, 1: (2013) 11–31. <http://www.annualreviews.org/eprint/SUGKGEWM7zFDgXIqBx/full/10.1146/annurev-po>
- Evans, Michael, Wayne McIntosh, Jimmy Lin, and Cynthia Cates. “Recounting the Courts? Applying Automated Content Analysis to Enhance Empirical Legal Research.” *Jour-*

- nal of Empirical Legal Studies* 4, 4: (2007) 1007–1039.  
<http://onlinelibrary.wiley.com/doi/10.1111/j.1740-1461.2007.00113.x/abstract>.
- Friedman, Barry. “Taking Law Seriously.” *Perspectives on Politics* 4, 02: (2006) 261–276.
- Gibson, James L., Gregory A. Caldeira, and Vanessa A. Baird. “On the Legitimacy of National High Courts.” *Am. Pol. Sci. Rev.* 92, 2: (1998) 343.  
<http://www.jstor.org/discover/10.2307/2585668?uid=3738016uid=2uid=4sid=211012111>
- Grimmer, Justin, and Brandon M. Stewart. “Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts.” *Political Analysis*  
<http://pan.oxfordjournals.org/content/early/2013/01/21/pan.mps028>.
- Grosser, Alfred. “Federal Constitutional Court’s Lisbon Case: Germany’s Sonderweg - An Outsider’s Perspective, The.” *German L.J.* 10: (2009) 1263.  
<http://heinonline.org/HOL/Page?handle=hein.journals/germlajo2009id=1273div=colle>
- Hanretty, Chris. “The Decisions and Ideal Points of British Law Lords.” *British Journal of Political Science* FirstView: (2012) 1–3.
- Hinkle, Rachael K., Andrew D. Martin, Jonathan David Shaub, and Emerson H. Tiller. “A Positive Theory And Empirical Analysis Of Strategic Word Choice In District Court Opinions.” *Journal of Legal Analysis* 4, 2: (2012) 407–444. <http://jla.oxfordjournals.org/content/4/2/407>.
- Kidd, Quentin. “The Real (Lack of) Difference between Republicans and Democrats: A Computer Word Score Analysis of Party Platforms, 1996–2004.” *PS: Political Science & Politics* 41, 03: (2008) 519–525.
- Klemmensen, Robert, Sara Binzer Hobolt, and Martin Ejnar Hansen. “Estimating policy positions using political texts: An evaluation of the Wordscores approach.” *Electoral Studies* 26, 4: (2007) 746–755.  
<http://www.sciencedirect.com/science/article/pii/S0261379407000601>.
- Klüver, Heike. “Measuring Interest Group Influence Using Quantitative Text Analysis.” *European Union Politics* 10, 4: (2009) 535–549.  
<http://eup.sagepub.com/content/10/4/535>.

- Kranenpohl, Uwe. “Herr des Verfahrens oder nur Einer unter Acht? Der Einfluss des Berichterstatters in der Rechtsprechungspraxis des Bundesverfassungsgerichts.” *Zeitschrift für Rechtssoziologie* 30, 2: (2009) 135–163. <http://cat.inist.fr/?aModele=afficheNcpsidt=22411752>.
- Lauderdale, Benjamin E., and Tom S. Clark. “Scaling Politically Meaningful Dimensions Using Texts and Votes.” *American Journal of Political Science* 58, 3: (2014) 754–771. <http://onlinelibrary.wiley.com/doi/10.1111/ajps.12085/abstract>.
- Laver, Michael, Kenneth Benoit, and John Garry. “Extracting Policy Positions from Political Texts Using Words as Data.” *American Political Science Review* 97, 02: (2003) 311–331.
- Laver, Michael, Kenneth Benoit, and Nicolas Sauger. “Policy competition in the 2002 French legislative and presidential elections.” *European Journal of Political Research* 45, 4: (2006) 667–697. <http://onlinelibrary.wiley.com/doi/10.1111/j.1475-6765.2006.00313.x/abstract>.
- Lax, Jeffrey R. “The New Judicial Politics of Legal Doctrine.” *Annual Review of Political Science* 14, 1: (2011) 131–157. <http://www.annualreviews.org/doi/abs/10.1146/annurev.polisci.042108.134842>.
- Lax, Jeffrey R., and Charles M. Cameron. “Bargaining and opinion assignment on the US Supreme Court.” *Journal of Law, Economics, and Organization* 23, 2: (2007) 276–302.
- Lowe, Will. “Understanding Wordscores.” *Political Analysis* 16, 4: (2008) 356–371. <http://pan.oxfordjournals.org/content/16/4/356>.
- . “There’s (Basically) Only One Way to Do it.”, 2013. <http://papers.ssrn.com/abstract=2318543>.
- Lowe, Will, and Kenneth Benoit. “Validating Estimates of Latent Traits from Textual Data Using Human Judgment as a Benchmark.” *Political Analysis* 21: (2013) 298–313. <http://pan.oxfordjournals.org/content/21/3/298.abstract>.
- Mayer, Franz C., and Maja Walter. “Die Europarechtsfreundlichkeit des BVerfG nach dem Honeywell-Beschluss.” *jura* 33, 7: (2011) 532–542. <http://www.degruyter.com/view/j/jura.2011.33.issue-7/jura.2011.078/jura.2011.078>



- McGuire, Kevin T., and Georg Vanberg. "Mapping the policies of the US Supreme court: Data, opinions, and constitutional law." In *Annual Meeting of the American Political Science Association*. Washington, DC. 2005.
- Meessen, Karl M. "Hedging European Integration: The Maastricht Judgment of the Federal Constitutional Court of Germany." *Fordham Int'l L.J.* 17: (1994) 511–530. <http://heinonline.org/HOL/Page?handle=hein.journals/frdint17id=525div=collection>
- Möllers, Christoph. "German Federal Constitutional Court: Constitutional Ultra Vires Review of European Acts Only Under Exceptional Circumstances; Decision of 6 July 2010, 2 BvR 2661/06, Honeywell." *European Constitutional Law Review (EuConst)* 7, 01: (2011) 161–167.
- Myerson, Roger B. "Theoretical comparisons of electoral systems." *European Economic Review* 43, 4–6: (1999) 671–697. <http://www.sciencedirect.com/science/article/pii/S0014292198000890>.
- Owens, Ryan J., and Justin P. Wedeking. "Justices and Legal Clarity: Analyzing the Complexity of U.S. Supreme Court Opinions." *Law & Society Review* 45, 4: (2011) 1027–1061. <http://onlinelibrary.wiley.com/doi/10.1111/j.1540-5893.2011.00464.x/abstract>.
- Payandeh, Mehrdad. "Constitutional review of EU law after Honeywell; Contextualizing the relationship between the German Constitutional Court and the EU Court of Justice." *Common Market Law Review* 48, 1: (2011) 9–38. <http://www.kluwerlawonline.com/abstract.php?area=Journalsid=COLA2011002>.
- Proksch, Sven-Oliver, and Jonathan B. Slapin. "How to Avoid Pitfalls in Statistical Analysis of Political Texts: The Case of Germany." *German Politics* 18, 3: (2009) 323–344. <http://www.tandfonline.com/doi/abs/10.1080/09644000903055799>.
- . "Position Taking in European Parliament Speeches." *British Journal of Political Science* 40, 03: (2010) 587–611.
- Proksch, Sven-Oliver, Jonathan B. Slapin, and Michael F. Thies. "Party system dynamics in post-war Japan: A quantitative content analysis of electoral pledges." *Electoral Studies* 30, 1: (2011) 114–124. <http://www.sciencedirect.com/science/article/pii/S0261379410001046>.

- Satzger, Helmut, and Tobias Pohl. "The German Constitutional Court and the European Arrest Warrant 'Cryptic Signals' from Karlsruhe." *J Int Criminal Justice* 4, 4: (2006) 686–701. <http://jicj.oxfordjournals.org/content/4/4/686>.
- Schmid, Christoph U. "All Bark no Bite: Notes on the Federal Constitutional Court's "Banana Decision"." *Eur. L. J.* 7: (2001) 95.
- Schmidt, Susanne K. "Sense of Deja Vu: The FCC's Preliminary European Stability Mechanism Verdict, A." *German L.J.* 14: (2013) 1. <http://heinonline.org/HOL/Page?handle=hein.journals/germlajo14id=5div=collection>
- Shapiro, Martin M. *The Supreme Court and Administrative Agencies*. New York, NY, US: Free Press, 1968.
- Slapin, Jonathan B., and Sven-Oliver Proksch. "A Scaling Model for Estimating Time-Series Party Positions from Texts." *Am. J. Pol. Sci.* 52, 3: (2008) 705–722. <http://onlinelibrary.wiley.com/doi/10.1111/j.1540-5907.2008.00338.x/abstract>.
- Stein, Professor Dr Torsten. "Always Steering a Straight Course? The German Federal Constitutional Court and European Integration." *ERA Forum* 12, 1: (2011) 219–228. <http://link.springer.com/article/10.1007/s12027-011-0200-5>.
- Stone Sweet, Alec. *The Judicial Construction of Europe*. Oxford [u.a.]: Oxford University Press, 2004.
- Thym, Daniel. "In the Name of Sovereign Statehood: A Critical Introduction to the Lisbon judgment of the German Constitutional Court." *Common Market Law Review* 46, 6: (2009) 1795–1822. <http://www.kluwerlawonline.com/abstract.php?area=Journalsid=COLA2009074>.
- Tomuschat, Christian. "Lisbon: Terminal of the European Integration Process?: The Judgment of the German Constitutional Court of 30 June 2009." *Zeitschrift für ausländisches öffentliches Recht und Völkerrecht* 70, 2: (2010) 251–282. <http://dialnet.unirioja.es/servlet/articulo?codigo=3366107>.

Weiler, J.h.h. “Does Europe Need a Constitution? Demos, Telos and the German Maastricht Decision.” *European Law Journal* 1, 3: (1995) 219–258. <http://onlinelibrary.wiley.com/doi/10.1111/j.1468-0386.1995.tb00030.x/abstract>.

Wendel, Mattias. “Judicial Restraint and the Return to Openness: The Decision of the German Federal Constitutional Court on the ESM and the Fiscal Treaty of 12 September 2012.” *German L.J.* 14: (2013) 21–52. <http://heinonline.org/HOL/Page?handle=hein.journals/germlajo14id=25div=collectio>

Figure 1: GFCC Rulings on European Integration 1967-2012, Opinion Length

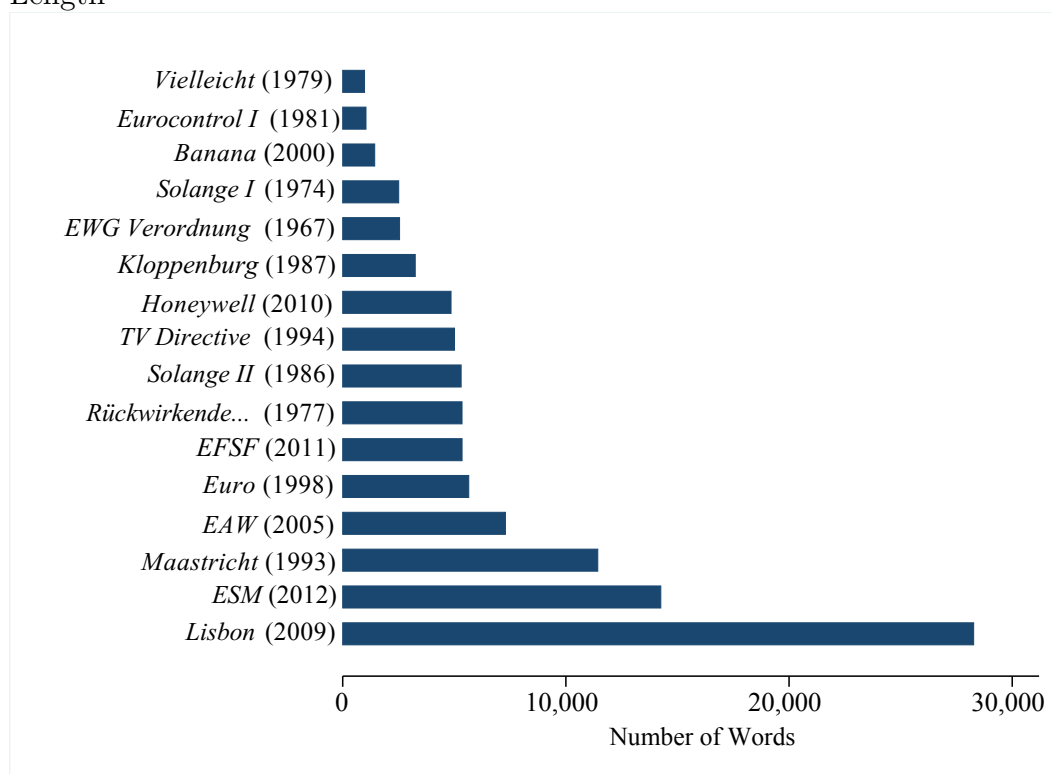
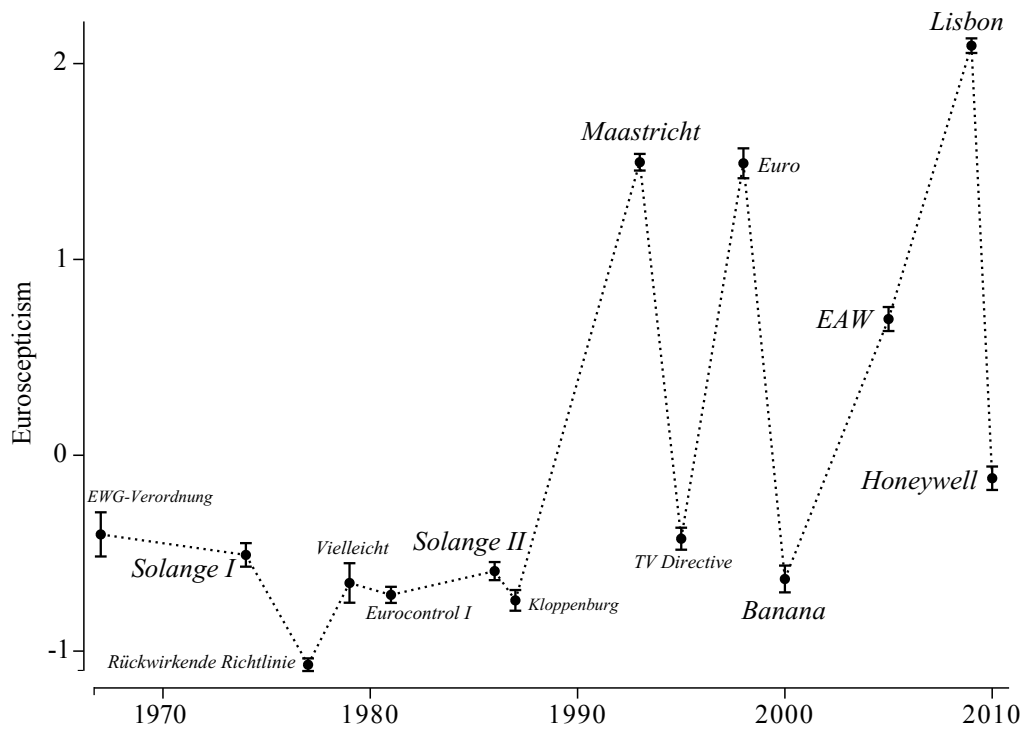
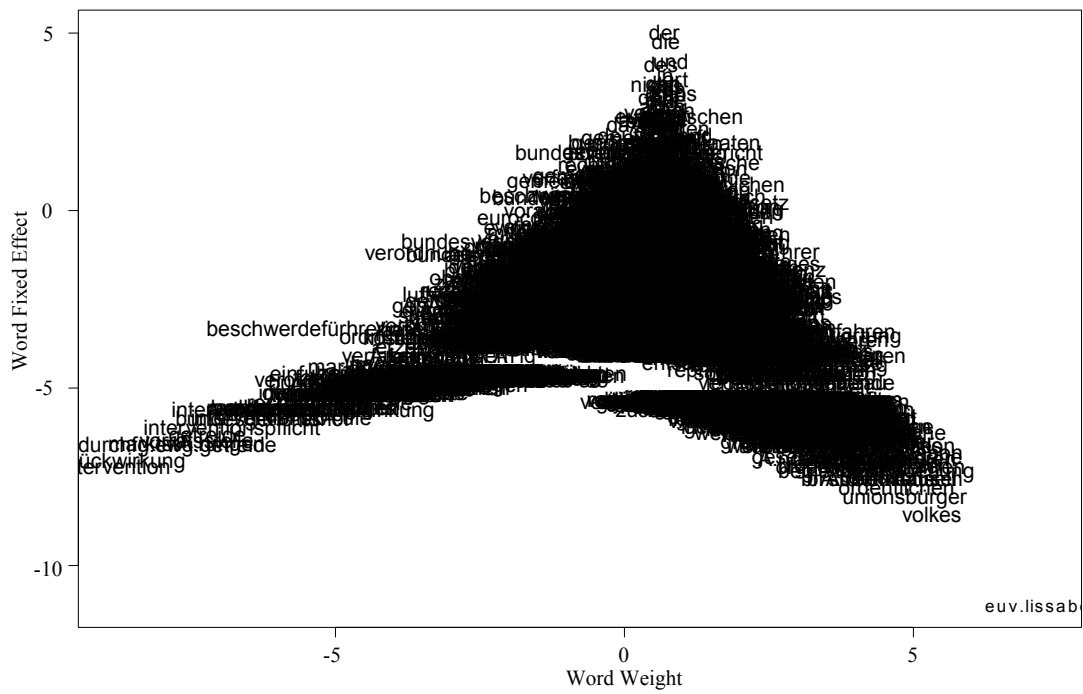


Figure 2: Text-Scaling (Wordfish) of GFCC Opinions on European Integration, 1967-2010



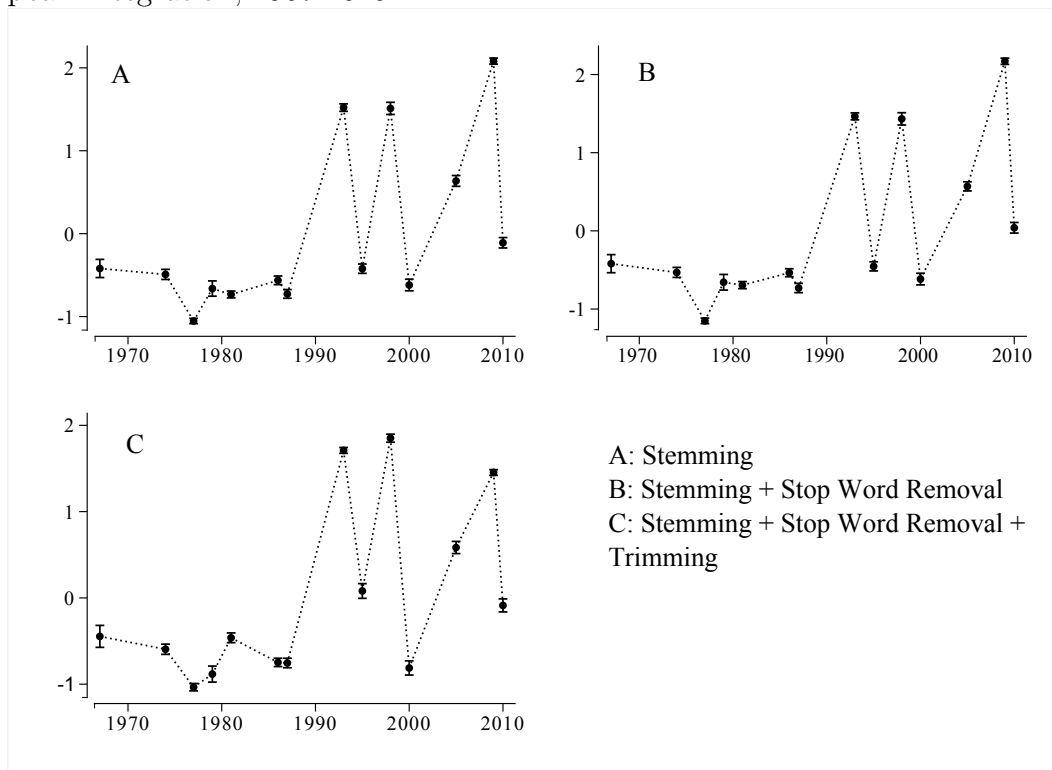
Notes: Model is identified via constraint *EWG Verordnung* < *Maastricht*. Vertical bars denote 95 per cent confidence interval from parametric bootstrapped standard errors (500 iterations).

Figure 3: Word Weight and Word-Fixed Effect, GFCC Opinions on European Integration, 1967-2010



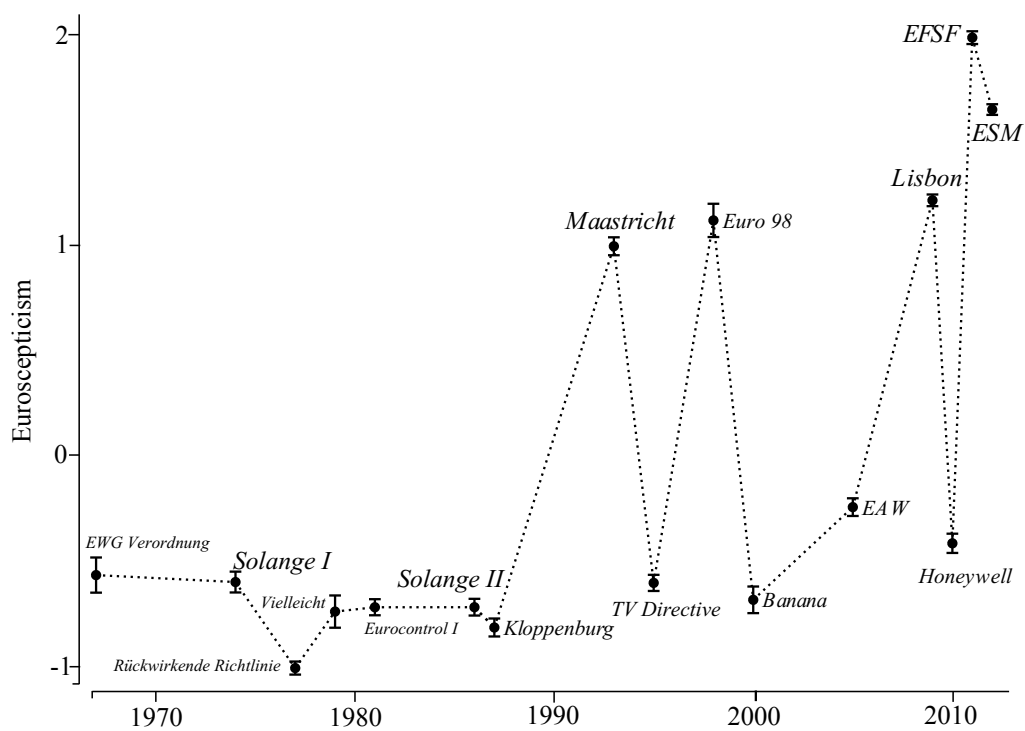
Notes: Word cloud shows the  $\psi$  (word-fixed effect) and  $\beta$  (word weight) parameter values for the 9,693 word types occurring in the fourteen GFCC opinions.

Figure 4: Effect of Pre-Processing on Positional Estimates, GFCC on European Integration, 1967-2010



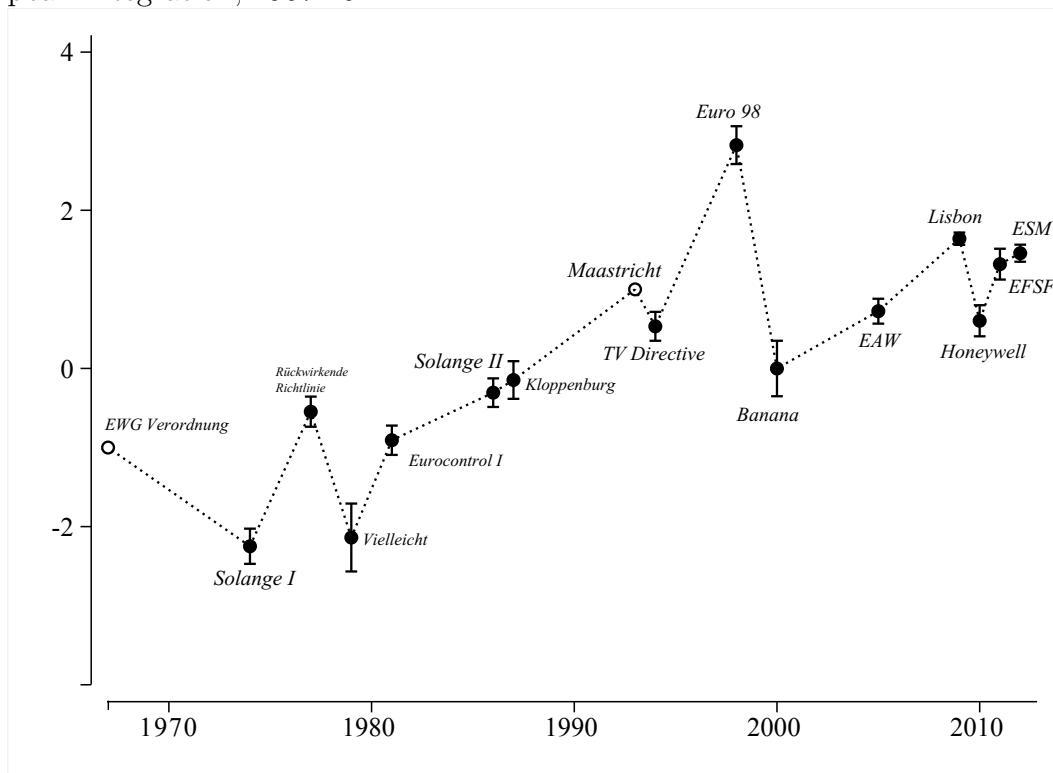
Notes: Trimming (Panel C) involved removing words occurring less than three times as well as words appearing in less than two opinions.

Figure 5: Text-Scaling (Wordfish), GFCC Opinions on European Integration, 1967-2012



Notes: Model is identified via constraint *EWG Verordnung* < *Maastricht*. Vertical bars denote 95 per cent confidence interval from parametric bootstrapped standard errors (500 iterations).

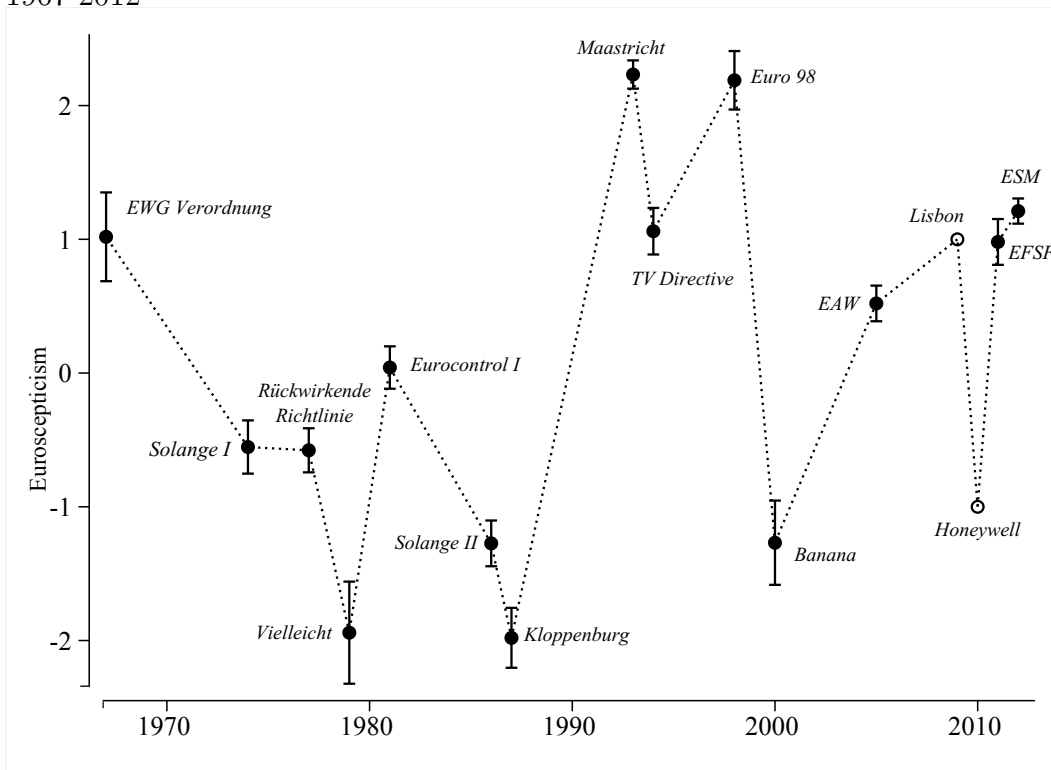
Figure 6: Supervised Text-Scaling (Wordscores), GFCC Opinions on European Integration, 1967-2012



Notes: *EWG Verordnung* and *Maastricht* used as train set. Vertical bars denote 95 per cent confidence interval computed from the variance of the test document's scored word types.



Figure 7: Supervised Text-Scaling, GFCC Opinions on European Integration, 1967-2012



Notes: *Lisbon* and *Honeywell* used as train set. Vertical bars denote 95 per cent confidence interval computed from the variance of the test document's scored word types.