Legitimation Discourses in Four Western Democracies:
Methodological Annotations

This appendix provides readers with basic methodological information pertaining to our empirical study of legitimation discourses in four Western democracies: Switzerland, Germany, Britain, and the United States. We first describe the procedure used to establish our monitoring and case-study text corpora, and then discuss the reliability of our text selection and coding routines. A more detailed presentation of these topics – as well as the full codebook of our study and other pieces of supplementary information – will be made available shortly.

Text selection

Our comparison of national legitimation discourses in Switzerland, Germany, Britain, and the United States is based on the analysis of pertinent articles and statements from two opinion-leading quality newspapers with nationwide coverage and readership per country: *Neue Zürcher Zeitung* and *Tagesanzeiger* (Switzerland), *Süddeutsche Zeitung* and *Frankfurter Allgemeine Zeitung* (Germany), *Guardian* and *Times* (Britain), *New York Times* and *Washington Post* (US). Only the Monday through Saturday issues were considered even for those dailies that have Sunday editions.

In a first step, the electronic versions of these papers – available through the media database Factiva (http://global.factiva.com) – were searched; relevant articles were downloaded and stored in files generated by MaxQDA, a qualitative text analysis software that allows researchers to link texts with a rectangular data matrix, with a row for each text. In a second step, we identified and coded the legitimation statements contained in these articles. Finally, our empirical analysis of discursive structures draws on two *different* text corpora and related datasets; these were, however, established in a similar fashion:

- a 'monitoring' corpus of 2,374 articles and 3,924 statements published in 2004;
- a 'case-study' corpus of 660 articles and 1,205 statements published during the 1990s and 2000s in the context of three sets of policy-related debates.

In the following, we explain how the texts and propositions for the two corpora and datasets were selected. In both instances, we relied on our notion of a legitimation grammar (Schneider/Nullmeier/Hurrelmann 2007) to classify articles and statements as pertinent or not. Hence

only a very specific kind of propositions – legitimation statements – was considered relevant for the purposes of this study. These statements – which convey a positive or negative evaluation of specific legitimation *objects* on the basis of explicit normative standards, or legitimation *patterns* – served as our basic units of research. Thus articles were included in one of the two corpora if they contained one or more of these legitimation statements.

A number of further precisions are in order. First, propositions of a clearly evaluative kind that do not state the underlying normative standard ('America is great', 'the political system of Germany is illegitimate', and the like) were treated as borderline cases of communicative (de-)legitimation and therefore included as well. Secondly, the stylized – and semantic rather than syntactical – nature of this grammar has to be underlined. Even a quick glance at real-world political communication suggests that legitimation statements may be formulated in a virtually unlimited number of ways, and hence they are not necessarily phrased along the *grammatical* lines of ideal-typical compound sentences taking the form 'LO x is (il-)legitimate because of LP y'. Moreover, it is readily apparent that legitimation statements – many of which are clad in figurative or metaphorical language – may draw on an essentially unlimited *vocabulary*.

For instance, the three core elements of these propositions are frequently spread over more than one sentence – and 'elliptical' formulations are also widespread. Only a minority of statements employ the actual word 'legitimate' and its derivations or make the 'is (il-) legitimate because' portion of the grammar explicit. More often the affirmative or critical thrust of the evaluation is merely implied by the value-laden character of the LP vocabulary itself (words like [in-]efficient, [un-]fair, and so on). In the most extreme case, the whole statement with its LO, LP, and evaluation components may be condensed in a single expression – as in a passing reference to the German *Bundestag* (federal parliament) as 'the talkshop in Berlin'. Of course, examples of this kind also illustrate the futility of attempts to develop closed and exhaustive *dictionaries* of legitimacy-related vocabulary, a point to which we return below.

Thirdly, statements taken from newspaper articles may reflect the author's own legitimacy assessments, or they may come in the form of (in-)direct quotations and even of propositions merely *ascribed* to other individuals or groups by the author ('many citizens believe that…'). However, our study deliberately focused on the textual level – that is, the propositional content of individual legitimation statements – and on the broad structures or trajectories of legitimation discourses. We therefore refrained from coding and examining speaker categories. By the same token, verification of the authenticity of claims and beliefs attributed to others by journalists appeared unnecessary. What counted for our purposes was the greater

or lesser prevalence of different types of statements at the 'surface' of legitimation discourses and public spheres, regardless of their author or accuracy. However, an extension of our grammar to speaker categories could be easily implemented and would indeed be crucial for more elaborate research into the dynamics and mechanisms of legitimacy-related communicative (inter-)action, the composition and strategies of discourse coalitions and discursive elites, or related topics.

Finally, our searches were not restricted to any particular section of the examined newspapers. Most legitimation statements are, of course, embedded in political news coverage, commentaries, and editorials. But relevant propositions may also be contained in various other feature pages, and even in rather unexpected types of articles, including movie reviews, sports reporting, and so on. There is no obvious substantive reason for excluding legitimation statements put forward in unlikely places, and hence it is one of the key advantages of the routine described here that it enabled us to search whole newspaper issues for legitimation statements without restricting our attention in an a priori fashion to a selection of article types (editorials, and so on), or to newspaper segments like the front page or the political news section.[1]

The broad scope of our searches was made possible by the partially automated nature of our routine, as described below. A great number of articles could be discarded in the automated (pre-)selection step of the routine, and hence without actually *reading* them. Yet it is important to underline that the ultimate identification of pertinent articles and statements, as well as their coding, entailed a 'reconstruction' of manifest content and its 'translation' into the framework of the legitimation grammar and its variables. The essentially interpretive nature of article selection and coding, however, proved demanding enough in pre-tests to have them performed by the research team itself rather than delegating them to (student) coders, as is usually done (see, for instance, Budge et al. 2001; the reliability of CMP data – which are nevertheless widely used – is, for instance, discussed – and criticized – in Volkens 2001; Laver, Benoit, and Garry 2003).

Thus articles for the two corpora were chosen, and all legitimation statements were coded, by Zuzana Krell-Laluhová (Switzerland), Frank Nullmeier and Achim Wiesner (Germany), Achim Hurrelmann (Britain), and Steffen Schneider (US). Moreover, articles and statements whose relevance or correct coding appeared doubtful to the individual researchers were discussed on a regular basis, and decided upon in a consensual fashion by the whole team. This approach enabled us to develop a common understanding of relevant articles and

variable categories, to fine-tune the rules for text selection and coding as we went along, and to perform these tasks with increasing speed and accuracy.

In light of our team approach to text selection and coding, inter-coder reliability tests might, then, appear dispensable; a number of informal pre-tests and formal reliability tests were nevertheless performed, albeit taking the partially 'discursive' and consensual nature of the text selection and coding procedure into account. Further details of this procedure and the findings of our test for the reliability of the selection procedure are outlined in the next sub-section, against the backdrop of the monitoring corpus, before we explain which adaptations were made to establish the case-study corpus.

Monitoring corpus

Our declared goal in compiling the monitoring corpus was to identify *all* legitimation statements published by the examined newspapers in a single year (2004), or at least as many as possible. On the one hand, given the novelty of our text analytical perspective, an especially solid basis for descriptive and explanatory inferences on the structures of Swiss, German, British, and US legitimation discourses (that is, a large N of legitimation statements) seemed appropriate. On the other hand, reading each and every article published in that year was beyond our means. We therefore devised a partially automated search routine, making heavy use of the powerful and sophisticated search engine offered by Factiva. Its first step – which still yielded a considerable number of articles to be read – nevertheless helped us weed out clearly irrelevant texts while, at the same time, ensuring a minimal amount of 'false negatives' (articles that are not recognized as pertinent by the automated part of the routine, and hence discarded, even though they contain relevant propositions).

The first step of the procedure was dictionary-based and closely tied to the two key variables of our legitimation grammar, objects (LO) and patterns (LP). Thus we began by creating small LO and LP dictionaries in English and in German, with the requisite adaptations for national idiosyncracies regarding political terminology and institutional arrangements. For instance, the four versions of the LO dictionary all contain the word 'parliament' or 'Parlament', as well as the proper names of the legislature and its chambers in each of the four examined political systems – *Nationalrat* and *Ständerat* (Switzerland), *Bundestag* and *Bundesrat* (Germany), House of Commons and House of Lords (Britain), as well as Congress, House of Representatives and Senate (US). Conversely, only the British LO dictionary contains an entry related to the monarchy, and so on. In a similar vein, the different versions of the LP dictionary are made up of value-laden words referring to the kinds of normative crite-

ria that we expected to be most prominent in the context of legitimacy-related propositions – words like (il-)legitimate, (in-)acceptable or (un-)justifiable, (un-)democratic, (il-)legal, (in-) effective, and so on.

Finally, the four versions of the dictionaries are not *exactly* parallel because there may a greater or lesser number of synonyms for each dictionary item in English and German, because different truncations were used as necessary and appropriate in the two languages, and because a greater or lesser number of problematic homonyms – whose inclusion would have deteriorated the accuracy of searches – had to be dropped from the word lists or otherwise dealt with in each case (for instance, 'Staat' is unproblematic in German, whereas even 'the state' proved highly ineffective but nevertheless had to be included in the English-language searches). The complex word lists employed to retrieve articles from Factiva, then, coupled the LO and the LP dictionaries together as follows:

(LO_1 *or* LO_2 *or* … LO_n) *same* (LP_1 *or* LP_2 *or* … LP_n)

Factiva's 'same' operator functions just like a logical 'and' but restricts searches to one and the same paragraph, thus making it more likely that LO and the LP vocabulary identified in an article indeed belongs to one and the same proposition rather than being spread across the entire text, and hence being unrelated. So each of the articles retrieved for further inspection contained at least one *paragraph* (including the title) with at least one object-related and one pattern-related *term* from the LO and LP dictionaries. For instance, a paragraph containing the term 'democracy' (LO) along with words like '(in-)effective' or '(un-) accountable' would be identified by our routines. Finally, because of the maximum length of a search request in Factiva, the LP portion of the routine had to be split in half (word list 1, word list 2), and each request had to be run twice:

(1) LO_i *same* word list 1 (= the first half of the LP vocabulary)
(2) LO_i *same* word list 2 (= the second half of the LP vocabulary)

Hence we actually used eight different search routines to retrieve articles from Factiva – and articles might have been found by part 1 or 2 of the routines (or both). Again, a couple of further precisions are in order. First, only the paragraphs actually containing two or more of our search words (highlighted in bold by Factiva) were considered – that is, read – in the sec-

ond part of our procedure (the final selection of relevant articles), unless the disambiguation of their meaning forced us to consider surrounding paragraphs as well.

Secondly, it proved fairly unproblematic to establish a more or less closed-ended dictionary of object-related search words linked to our hierarchy of legitimation objects. The odd article or paragraph may, of course, refer to a political system or institution in an entirely idiosyncratic fashion, and hence be overlooked by our routines because of the closed-ended nature of the LO vocabulary used. Yet on the basis of our pre-tests, we have every reason to believe that this is extremely infrequent: A commentator may, for instance, call a national parliament a 'talk shop', as in our example above, but it is most unlikely that she will never use its proper name or generic terms like 'parliament' or 'legislature' throughout her article, or even throughout an individual paragraph.

Readers might be more suspicious about our LP dictionary, and understandably so. There is obviously no way to construct a similarly closed-ended (and at the same time short) dictionary of value-laden terms – including any number of idiosyncratic ones – that might be used in the context of legitimation statements; journalists or the speakers they cite obviously do not do us the favour of using the signal word '(il-)legitimate' whenever they formulate such evaluative propositions. Yet a range of terms suggest themselves, whether against the backdrop of democratic theory (words and expressions like 'accountable', 'responsive', and 'popular sovereignty') or in light of diagnoses that postulate a transformation of legitimacy away from democratic criteria ('efficiency', 'effectiveness', and so on). Sometimes a small number of rather obvious signal words (as with 'efficiency', etc.) are used in many or most references to the underlying evaluation standard while other normative criteria are less intimately linked with specific terms. The search words ultimately retained are thus linked to normative criteria prominent in academic debates, or their usefulness was confirmed by way of informal pre-tests, but they do not by any means constitute exhaustive dictionaries of the potentially relevant vocabulary.

The unavoidable limitations of our LP vocabulary, of course, raise the issue of 'false negatives'. Ideally, the automated step of our text selection procedure should have minimized both the number of 'false negatives' and 'false positives' (articles identified by the Factiva searches that prove irrelevant upon further inspection). As it turns out, however, whereas the search routines were remarkably successful with regard to the first benchmark, they were much less effective in weeding out irrelevant material – that is, articles with no political and legitimacy-related content whatsoever, or merely factual political news reports.

Table 1 Accuracy of automated search routines ('false negatives')

	CH	DE	GB	US	Total/Overall
(1) number of articles	42	61	36	63	202
(2) number of paragraphs	907	1,285	851	2,185	5,228
(3) number of paragraphs with search words	172	171	93	277	713
(4) … of which: number of paragraphs with legitimation statements	82	84	47	85	298
Ratio (4) / (3)	*47.7*	*49.1*	*50.5*	*30.7*	*41.8*
(5) number of paragraphs without search words	735	1,114	758	1,908	4,515
(6) … of which: number of paragraphs with 'false negatives'	9	28	7	39	83
Ratio (6) / (5)	*1.2*	*2.5*	*0.9*	*2.0*	*1.8*

To test for the extent to which the automated step of our procedure made us overlook pertinent statements (and articles), a random sample of roughly 200 texts (16.3 % of the articles in the monitoring corpus – each text containing exactly one previously identified and coded legitimation statement – was read *in full* (remember that only the paragraphs containing search words had initially been read, and articles had been classified as relevant or discarded on that basis). Table 1 gives the number of articles considered for each of the four countries (1), the overall number of paragraphs in these articles (2), and the number of paragraphs containing search words (3). As further indicated by the table, the share of paragraphs found to be relevant (4) among the automatically pre-selected ones may be estimated to be roughly between 30 and 50 per cent for the examined countries, indicating a relatively high – if by no means satisfactory – accuracy of the automated search routines. By contrast, the lower portion of the table shows that the number and share (roughly two per cent) of paragraphs containing 'false negatives' is low. We are therefore confident that our monitoring corpus indeed documents the vast bulk of legitimation statements made in the examined newspapers over the year 2004.

On the other hand, Table 1 also indicates the fact that the automated routines still cast a much too wide net, and hence we could not dispense with the close reading of pre-selected texts. The final selection was made by each of the individual coders alone whenever the (ir-) relevance of articles seemed obvious; doubtful cases were discussed and decided upon jointly, and in a consensual fashion, during regular sessions of the research team. How reliable was this text selection procedure? A formal test of inter-coder reliability was performed on the basis of a random sample of eight days – one for each newspaper – in 2004; the use of our

automated search routines yielded 124 articles for those dates (Table 2). These articles were read individually by the four team members and coded as (ir-)relevant or doubtful cases.

Table 2 Reliability test, text selection – sample

Newspaper	**Date**	**Number of hits, automated routines**
NZZ (word list 1)	09/16/2004	14
Tagesanzeiger (word list 2)	05/22/2004	10
FAZ (word list 1)	03/26/2004	15
SZ (word list 2)	02/09/2004	17
Guardian (word list 1)	05/05/2004	21
Times (word list 2)	08/09/2004	9
New York Times (word list 1)	02/28/2004	26
Washington Post (word list 2)	06/16/2004	12
Sum	-	124

Each of the four researchers considered between 14 and 16 articles – that is, between 11.3 per cent and 12.9 per cent of the sample – to be relevant. However, eleven articles – 8.9 per cent of the sample – were also marked as doubtful cases. As shown in Table 3, the levels of pair-wise inter-coder agreement (in per cent) are very satisfactory; all pairs achieve results close to the average value of 91.1 per cent, and no problematic coders emerge.[2]

Table 3 Reliability test, text selection – findings

Pairwise intercoder agreement (%)				
	Coder 1	Coder 2	Coder 3	Coder 4
Coder 1	(100.0)	92.7	91.1	90.3
Coder 2		(100.0)	90.3	91.1
Coder 3			(100.0)	91.1
Coder 4				(100.0)

Case study corpus

The case study corpus was established in a comparable fashion but with the objective to identify the volume and nature of legitimation statements legitimacy-related communication in the context of debates and conflicts related to specific policy issues. The text searches were in each case restricted to three-month time windows 'anchored' by important events. The specific debates to be examined in the areas of institutional reform, social and foreign policy – as well as the precise time windows – were selected in the following fashion:

- in a first step, the four country experts of the team scanned policy developments in the three areas for particularly salient issues that were the subject of intense political conflicts and heightened public attention at least once in the (early) 1990s and once in the late 1990s or 2000s according to the literature;
- in a second step, this qualitative picture of attention cycles thus established was confirmed with the help of Factiva, by examining the monthly up and down of issue-related search words and expressions ('healthcare reform', etc.) for the 1990 to 2005 period;
- the final selection of issues was made on the basis of the following criteria: substantive importance of the chosen issue or debate; comparability of debates across countries and time periods; existence of a clearly recognizable 'focusing' event (parliamentary votes, elections, and so on). The time windows usually begin two months before that event and end a month thereafter (only a few start and end dates were chosen differently for substantive reasons).

Table 4 Case studies, number of statements and raw values

	Time window 1				**Time window 2**			
	# statements	**Intensity level**	**Legitimacy level**	**Democratic quality**	**# statements**	**Intensity level**	**Legitimacy level**	**Democratic quality**
CH								
Inst. reform	54	63.6	35.2	60.8	38	13.2	57.9	56.6
Foreign policy	143	51.4	50.3	51.0	46	33.0	43.5	39.4
Social policy	15	25.0	53.3	69.2	47	15.9	31.9	39.5
DE								
Inst. reform	26	57.9	42.3	40.0	76	11.9	17.1	54.8
Foreign policy	21	23.3	33.3	22.2	29	10.7	41.4	10.3
Social policy	40	5.8	15.0	40.6	20	3.1	20.0	44.4
GB								
Inst. reform	78	20.9	16.0	55.0	142	14.4	44.0	50.6
Foreign policy	70	10.8	46.0	63.3	67	9.9	61.0	70.7
Social policy	16	4.3	50.0	30.8	43	3.9	28.0	29.4
US								
Inst. reform	28	11.7	19.2	76.9	37	23.9	54.1	42.1
Foreign policy	39	1.3	79.5	80.6	88	2.4	71.6	66.2
Social policy	27	5.6	29.6	30.8	15	7.4	13.3	85.7

Once again, a two-step procedure was employed to identify relevant articles. For the pre-selection of texts, we now used issue-specific dictionaries and search routines. The final selection was, again, made on the basis of a thorough reading of the articles yielded by the Factiva searches; articles containing one or more legitimation statements were retained in the case study corpus. Table 4 indicates the number of legitimation statements identified and analyzed for each time window and country, together with the raw values used to depict intensity and legitimacy levels, as well as the democratic quality of debates.

Identification and coding of legitimation statements

The identification and coding of legitimation statements in the paragraphs containing two or more words from our LO and LP dictionaries followed the construction of our text corpora. A number of subsidiary rules were employed to identify legitimation statements and to 'translate' them into the variables of the legitimation grammar. The general rule for coders was to avoid 'over-interpretations'. The object, evaluative character and positive or negative thrust of would-be statements had to be explicitly stated. Wherever any of these elements could not be disambiguated, propositions were not included. Cases where the ironic or non-ironic character of propositions, and hence the affirmative or critical thrust of evaluations, remained unclear are apposite here – and so are 'legitimation statements' that are presumably *implied* in different other types of propositions. One could, for instance, argue that conditional sentences ('if social programs were cut any further, the welfare state would no longer ensure distributive justice') or political demands ('we need a more democratic judiciary') presuppose legitimacy assessments ('the current welfare state ensures distributive justice'; 'the judiciary in its current form is undemocratic') – yet coders were instructed to stop short of making far-reaching interpretive steps of this kind. Finally, historical statements ('the German welfare state of the 1970s was bloated and inefficient') as well as prognostic ones ('the German welfare state regime will no longer be viable in a few decades from now') were ignored.

Two final precisions need to be made. First, legitimation statements may, of course, be more complex than suggested by our grammar in the sense that they include more than one legitimation object or pattern, or weigh positive and negative assessments against each other; consider the following examples:

- 'Congress and the Supreme Court are equally unaccountable' (two LO);
- 'the members of the House of Lords have no democratic legitimacy and are totally ignorant' (two LP);
- 'German federalism may well be inefficient – but it guarantees adequate representation of regional interests) (a weighing of positive and negative assessments).

In each of these scenarios (or where they were combined), our procedure was to break complex statements into simple ones – one for each LO (Congress, Supreme Court) in case 1, one for each LP (and the House of Lords as LO) in case 2, a delegitimating and a legitimating statement on German federalism in case 3. This, of course, entails a certain loss of information, especially with regard to the third kind of scenario. The procedure, however, greatly facilitated the organization of our datasets.

Finally, a maximum of ten (simple) legitimation statements per article was included in the datasets. This limitation – implemented for entirely pragmatic reasons – hardly led to a great loss of pertinent information, given that the average number of legitimation statements per article in our material is only 1.7. Moreover, articles focusing very much on legitimacy-related issues also tend to *repeat* specific assessments; the instruction was to code (and count) such repeated statements – that is, statements with the same LO, LP and evaluation (positive or negative) – only once per article.

For each statement, then, we coded the three variables of our legitimation grammar – legitimation object, whether the evaluation is positive or negative, and legitimation pattern – as well as the policy or issue context in which the proposition was made. As with the text selection procedure, the team's 'country experts' individually coded statements from 'their' portion of the corpora (the Swiss, German, British and US papers, respectively) but doubtful cases were discussed and decided upon jointly. In the following, we restrict ourselves to a few cursory remarks on the categories of these variables and a glance at the results of reliability tests performed for each of them. The reliability tests used the sample of 202 articles (and statements) already used above, for the estimation of 'false negatives'.

Legitimation objects (LO, xLO): A list of 21 legitimation objects, arranged in a hierarchy of four *tiers* of objects (regime principles of the democratic nation state; regime or political community as a whole; core institutions; key actor groups), was considered (Table 5). Overall, the coding of this variable proved unproblematic because coders could usually rely on the manifest content of texts and the 'official' names of institutions (Congress, *Nationalrat*, and so

on), or on unequivocal generic words and expressions such as 'direct democracy' and 'federalism'. The distinction between references to the political community v. the regime as a whole (tier 2 of our hierarchy) and the correct identification of references to the interest-group system were, perhaps, most challenging. But as illustrated by Table 6, the coding of this variable proved rather unproblematic at the level of individual objects (with an average agreement of 84.1 per cent between the four coders), and even more so at the aggregated (xLO) level (89.6 per cent).

Table 5 Objects of legitimation: Aggregate and individual categories

(1) Regime principles	Territorial state, national sovereignty Constitutionalism Democratic government - Direct, representative democracy - Westminster, consensus democracy - Parliamentary, presidential democracy Interventionist/welfare state
(2) Political system or political community as a whole	Political system Political community/nation/*demos*
(3) Political institutions	Monarchy, republic Legislative branch Executive branch Judicial branch Electoral system Unitary, federal system
(4) Groups of actors	Political class, elites Party system Interest group system

Table 6 Reliability test, object variable – findings

Pairwise intercoder agreement (%)				
	Coder 1	Coder 2	Coder 3	Coder 4
Coder 1	(100.0)	88.9	81.0	76.2
Coder 2		(100.0)	92.6	85.7
Coder 3			(100.0)	79.9
Coder 4				(100.0)
(xLO = object group)				
	Coder 1	Coder 2	Coder 3	Coder 4
Coder 1	(100.0)	91.0	86.2	84.7
Coder 2		(100.0)	95.2	92.1
Coder 3			(100.0)	88.4
Coder 4				(100.0)

Positive or negative evaluation (DERELEG): As indicated above, the coding of this dichotomous variable should not cause problems – perhaps with the (rare) exception of cases where propositions might be read as ironic, and the like. And indeed a look at the results of our reliability test (Table 7), does not suggest problems (the average agreement is 95.7 per cent).

Table 7 Reliability test, (de-)legitimation variable – findings

Pairwise intercoder agreement (%)				
N = 193	Coder 1	Coder 2	Coder 3	Coder 4
Coder 1	1	97.4	95.3	93.2
Coder 2		1	97.9	95.8
Coder 3			1	94.8
Coder 4				1

Legitimation patterns (LP, xLP): A list of 23 legitimation patterns and a typology of four *groups* of patterns (democratic v. non-democratic, input v. output), with the residual categories of 'general' statements – not mentioning of a specific pattern – and 'other' patterns, were used (Table 8). We note in passing that there are, in principle, four rather than two ways to combine each LP with the two categories of the DERELEG variable. A speaker could, for instance, *legitimate* a political system or institution because it is assessed as being democratic but might also *delegitimate* it for the same reason. These cases would certainly merit a sustained qualitative analysis, and they are not necessarily as implausible as it might seem. A constitutional court, for instance, might well be evaluated *negatively* because it succumbs to democratic and populist pressures (the 'tyranny of the majority'), and hence because it appears to be 'too democratic'. Still, the present analysis ignores such unusual cases for the simple reason that they are exceedingly rare – that is, in the vast balk of statements, the positive and

negative version of the used evaluation standard (for instance, effective v. ineffective, fair v. unfair) is coupled with the expected (positive v. negative) thrust of the assessment.

This is undoubtedly a demanding variable, if alone for the sheer number of its categories. Moreover, it forces coders to grasp the meaning of argumentations that are frequently put forward in rather idiosyncratic terms. Therefore it does not come as a surprise that the average inter-coder agreement at the level of individual and aggregate patterns (xLP) (80.0 and 82.1 per cent, respectively) is somewhat lower than the corresponding values for the (x)LO and DERELEG variables. Still, the values seem entirely adequate for our exploratory study (Table 9).

Table 8 Patterns of legitimation: Aggregate and individual categories

	Democratic	**Non-democratic**
Input	*Popular sovereignty* – power resides in the people	*Charisma* – leadership by a strong personality
	Participation – citizens may actively contribute to decision making	*Expertise* – leadership by experts
	Deliberation – political processes are based on reason(ing)	*Religion* – political processes are grounded in religious authority
	Transparency – political processes are public and accessible	*Tradition* – political processes follow traditional rules
	Accountability – rulers may be controlled and removed	*Moderation* – political style is conciliatory
	Legality – domestic or international legal rules are respected	
	Credibility – political processes conform to stated objectives, no hidden agenda	

Output	*Human rights* – individual rights and civil liberties are guaranteed	*Effectiveness and efficiency* – common problems are addressed successfully, or solutions are cost-effective
	Empowerment – material and cognitive preconditions of citizenship are guaranteed	*Distributive justice* – resources and burdens are distributed equally
	Common good – political results serve the population as a whole, not just privileged elites	*Stability* – political results are durable and lasting
	Reversibility – political results are not irrevocable	*Integration* – outputs reflect or enhance the polity's cohesion and identity
		Morality – outputs conform to moral standards
		Sovereignty – enhancement of a polity's autonomy, power, or interest
		International standing – enhancement of a polity's status in the world

Table 9 Reliability test, pattern variable – findings

Pairwise intercoder agreement (%)				
	Coder 1	Coder 2	Coder 3	Coder 4
Coder 1	(100.0)	88.0	78.7	74.9
Coder 2		(100.0)	90.7	77.6
Coder 3			(100.0)	69.9
Coder 4				(100.0)
(xLP = pattern group)				
	Coder 1	Coder 2	Coder 3	Coder 4
Coder 1	(100.0)	89.6	83.1	72.1
Coder 2		(100.0)	93.4	80.3
Coder 3			(100.0)	74.3
Coder 4				(100.0)

Issue: Finally, for each statement, we identified the policy or issue context in which it was made – is the paragraph that contains the statement, for instance, dealing with issues of foreign or social policy, or with institutional reform (Table 10)? Where necessary, coders could

infer the issue context from surrounding text or the entire article where a reading of the paragraph itself did make the issue explicit enough. Here, too, a glance at the results of our reliability tests indicates somewhat greater inconsistency (with an average pair-wise inter-coder agreement of 77.6 per cent) than for the other coding decisions, although the overall level of agreement, again, seems adequate (Table 11).

Table 10 Issue areas

Issue area
Institutional processes and reforms
Social and economic policy
Foreign policy
Domestic security and migration
Culture and education
Other issues

Table 11 Reliability test, issue variable – findings

Pairwise intercoder agreement (%)				
	Coder 1	Coder 2	Coder 3	Coder 4
Coder 1	(100.0)	85.6	72.2	69.0
Coder 2		(100.0)	86.1	82.9
Coder 3			(100.0)	69.5
Coder 4				(100.0)

Finally, as suggested above, the complete text corpora have been stored, and hence remain accessible for qualitative analyses, as MaxQDA files. The legitimation statements were initially coded in MaxQDA as well, and the resulting data matrices were exported to SPSS for our quantitative analyses (two versions of this data matrix exist: the original spreadsheet format – as imported from MaxQDA – with case rows referring to *articles*, and one with case rows referring to the individual legitimation statements; this latter version was used for our univariate analyses and cross-tabulations).

References

Budge et al. 2001. *Mapping Policy Preferences. Parties, Electors and Governments, 1945-1998*. Oxford: Oxford University Press.

Lacy, S./Riffe, D. 1996. Sampling Error and Selecting Intercoder Reliability Samples for Nominal Content Categories. *Journalism & Mass Communication Quarterly* 73, 963-73.

Laver, Michael/Benoit, Kenneth/Garry, John. 2003. Extracting Policy Positions from Political Texts Using Words as Data. *American Political Science Review* 97, 2, 311-31.

Neuendorf, Kimberly A. 2002. *The Content Analysis Guidebook*. Thousand Oaks: Sage.

Riffe, D./Lacy, S./Fico, F.Y. 1998. *Analyzing Media Messages. Using Quantitative Content Analysis in Research*. Mahwah: Lawrence Erlbaum.

Schneider, Steffen/Nullmeier, Frank/Hurrelmann, Achim. 2007. Exploring the Communicative Dimension of Legitimacy: Text Analytical Approaches, in: Achim Hurrelmann/Steffen Schneider/Jens Steffek (eds), *Legitimacy in an Age of Global Politics*. Houndmills, Basingstoke: Palgrave Macmillan.

Volkens, Andrea. 2001. Manifesto Research Since 1979: From Reliability to Validity, in: Michael Laver (ed.), *Estimating the Policy Positions of Political Actors*. London: Routledge.

Notes

[1] Factiva offers a 'region' filter, as well as the option to exclude a small number of prima facie irrelevant text types (event schedules, listings of stock prices, obituaries, weather reports, and so on). Both options were used but proved relatively ineffective (for instance, irrelevant text types occasionally slipped through the net).

[2] We are well aware of the fact that inter-coder agreement in per cent has drawbacks as a measure of reliability: It does not account for chance agreement and may therefore be considered as too 'liberal'. Yet measures like Scott's pi, Cohen's kappa or Krippendorff's alpha 'have been criticized as being overly conservative, giving credit only to agreement beyond chance, a tough challenge in the case of extreme distributions' (Neuendorf 2002, 151). Many

of the coding decisions whose reliability we tested were, in fact, made challenging by such extreme distributions (final selection of texts) or a complex and highly differentiated coding scheme; also note that achieving high reliability scores is more challenging where latent content – as opposed to manifest – content is coded (Neuendorf 2002, 146; Riffe et al. 1998, 107). The reported scores should be assessed – and seem quite adequate – in light of these considerations, especially if it is kept in mind that a substantial number of articles and legitimation statements was selected and coded by the entire team (a feature that is not captured by the scores – which are 'conservative' in that sense – but obviously reduces the level of inconsistency in our data). Finally, we follow Lacy and Riffe (1996; Riffe et al. 1998) in compensating a lower assumed level of reliability in the 'population' (the entire dataset) with sample sizes which tend towards the higher end of the 'industry standard' of as little as 50 cases (Neuendorf 2002, 159).