

Skandalisierung korpuslinguistisch. Ein empirisch-linguistischer Blick auf die Berichterstattung zur "Wulff-Affäre"

Noah Bubenhofer (Mannheim und Dresden)

Abstract

What kind of language is typical of the mass media coverage of a scandal? Which linguistic means have the effect of scandalizing a public person like a politician? The study presented in this paper uses corpus linguistic methods to investigate a specific scandal in recent German politics: Federal President Christian Wulff was accused by the media of not having informed transparently about a private loan he had taken out. Further allegations of corruption presented by the media resulted in his resignation after only two months. The paper shows how automatic language pattern analyses of a corpus of articles about Wulff in two leading German newspapers reveal the typical patterns used in a discourse of scandalization. Additionally, also characteristic grammar features like tempus forms are incorporated in the analysis. The texts scandalizing a person use more patterned language than other news texts on politics. The patterns reflect the characteristic speech acts of scandalizing like accusing, informing, or speculating on the political future of the scandalized person, but also speech acts like defending and regretting. The use of tempus shows that scandalizing texts are more dynamic and vivid compared to other political coverage.

1 Die Wulff-Affäre: Ein Skandal?

Der Diskurs zur sogenannten "Wulff-Affäre" ist terminologisch durch die Begriffe "Affäre" und "Skandal" geprägt. Die Skandalforschung kennt eine Vielzahl von Definitionen von "Skandal", die sich jedoch zumindest bezüglich der involvierten Akteure im Kern ähneln und auf der "Skandaltriade" von Skandalisiertem, Skandalisierer¹ und dem Publikum beruhen (Neckel 1989: 58; Beckmann 2006: 62). Auch Holly (2003: 47–68) zielt auf diese Triade ab und nennt zugleich die jeweiligen Funktionen der Akteure:

1. Es muss ein "abweichendes Verhalten" des Skandalisierten vorliegen.
2. Dieses abweichende Verhalten muss durch die Skandalisierer "enthüllt" werden.
3. Das Skandalpublikum muss sich "empören", wobei Korrektivschritte der "Aufarbeitung" und der "Folgen" eingeleitet werden.

Beim ehemaligen deutschen Bundespräsidenten Christian Wulff liegen in der Summe mehrere abweichende Verhaltensformen vor, begonnen mit einer kritischen Antwort Wulffs im Rahmen einer Befragung im niedersächsischen Landtag vom 18. Februar 2012 auf die Frage

¹ In der Literatur sind sowohl die Schreibweisen Skandalisierung, Skandalisierer etc. (z. B. Liedtke 2003; Holly 2003) als auch Skandalierung, Skandalierer etc. (z. B. Beckmann 2006; Hitzler 1989) gebräuchlich.

nach geschäftlichen Beziehungen zwischen ihm und dem Unternehmer Egon Geerkens, die er verneinte (Wikimedia Foundation 2012: Lemma "Wulff-Affäre"). Streng formal gesehen war diese Antwort korrekt, doch Wulff verschleierte damit, dass er von Geerkens Frau einen Privatkredit über eine halbe Million Euro zur Finanzierung seines Eigenheims erhalten hatte. Kritisch an diesem Vorgang ist dabei nicht nur die Antwort Wulffs auf die Frage im Landtag, sondern auch die Entgegennahme des zinsgünstigen Kredits, woraus der Vorwurf der Begünstigung erwächst.

Der Wulff-Skandal umfasst viele weitere Momente abweichenden Verhaltens, die jedoch erst im Laufe des Skandals neu skandalisiert wurden. Wichtigstes Element ist dabei wahrscheinlich die "Medienaffäre", die im Vorwurf besteht, dass Wulff verfassungswidrig versuchte, die Berichterstattung über die Kreditaffäre zu verhindern. Wulff versuchte am 12. Dezember 2012 bei Kai Diekmann, dem Chefredakteur der Bild-Zeitung, die am darauffolgenden Tag geplanten Artikel zur Aufdeckung der Einzelheiten zum Privatkredit zu verhindern. Die Berichterstattung ist demnach auch das zweite Element der obigen Definition von "Skandal", nämlich die Enthüllung des abweichenden Verhaltens durch die Skandalisierer. Die Komplexität des Wulff-Skandals zeigt sich darin, dass diese Enthüllung nur ein neuer Ausgangspunkt für abweichendes Verhalten und weitere Enthüllungen war. Denn es bedurfte weiterer solcher Enthüllungen (anderer Zeitungen), um den Beeinflussungsversuch Wulffs bei der Bild-Zeitung publik zu machen.

Im Laufe des Skandals gab es mehrere Momente der "Empörung" des Skandalpublikums, darunter Aussagen von Politikern, Umfragen in der Bevölkerung, Leserbriefe sowie auch Demonstrationen.² Zu den Versuchen Wulffs, den Skandal aufzuarbeiten oder zu korrigieren, zählt ein Doppelinterview der beiden Sender des öffentlich-rechtlichen Fernsehens am 4. Januar 2012 mit Wulff, in dem er Fehler eingestand. Der Rücktritt Wulffs erfolgte am 17. Februar 2012. Doch auch danach war der Skandal nicht abgeschlossen, da sich eine Diskussion darüber entfachte, ob Wulff Anspruch auf den Ehrensold habe.

In Anlehnung an Beckmann (2006) kann die Wulff-Affäre auch als "komplexes Handlungsspiel im Bereich öffentlicher Moralisierungskommunikation" (gestützt auf Bergmann/Luckmann 1999) gelesen werden, bei dem die Skandalisierung aus den Komponenten des Vorwurfs, der Entrüstung und der Aufforderung besteht (Beckmann 2006: 69) und die Medienakteure ein Entrüstungsklima schaffen, in dem Emotionen manifestiert werden können. Der dynamische Aspekt von Skandal, der mit "Skandalisierung" oder "skandalisieren" betont wird, macht zudem deutlich, dass ein Skandal nicht per se ein Skandal ist, sondern durch die entsprechende Etikettierung zum Skandal gemacht wird. Es geht darum, "eine (woraus auch immer resultierende) Definitionsmacht selektiv einzusetzen, um eine Person im Zusammenhang mit einem "empörungswürdigen" Ereignis oder Sachverhalt sozial zu diskriminieren" (Hitzler 1989: 337). "Ein Sachverhalt wird dadurch zum Skandal, daß er bekannt gemacht und erfolgreich als Skandal *definiert* worden ist" (ebd.: 334).

Die diskurseigene Bezeichnung für die Geschehnisse um Wulff lautet oft "Affäre", nicht unbedingt "Skandal" (vgl. z. B. den gleichnamigen Wikipedia-Artikel, s.o.). Der Affären-Begriff dient hierbei jedoch in erster Linie als Referenz auf die zeitlich und sachlich nicht

² Vgl. "Wulff-Affäre: Rücktrittsforderungen und Rücktritt" in Wikimedia Foundation (2012).

genau umrissenen Geschehnisse der Vorgeschichte und des Rücktritts von Wulff. Typologisch soll im Folgenden diese Affäre als Skandal bezeichnet werden.

Der Wulff-Skandal weist alle Elemente eines Skandals auf, zeigt aber auch auf, dass ein Skandal eine komplexe Mikrostruktur in Form von darin enthaltenen weiteren Teilskandalen haben kann. Da das Interesse bei dieser Untersuchung jedoch in erster Linie bei der Kontrastierung von Skandal und Nicht-Skandal liegt, wird im Folgenden die Makrostruktur des Skandals in den Fokus genommen und die Skandal-Zeit mit der Berichterstattung davor verglichen. So lässt sich aus den außersprachlichen Begebenheiten folgendes Zeitschema ableiten:

- Ministerpräsident Niedersachsen, 1. Kabinett: 4. März 2003 bis 25. Februar 2008;
- Ministerpräsident Niedersachsen, 2. Kabinett: 26. Februar 2008 bis 30. Juni 2010;
- Bundespräsident: ab 1. Juli 2010;
- Beginn der Berichterstattung über die Kredit-Affäre: 13. Dezember 2011;
- Rücktritt: 17. Februar 2012.

Mit korpuslinguistischen Methoden soll untersucht werden, ob sich die Zeit des Skandals, die mit der Berichterstattung über die Kredit-Affäre beginnt und mit dem Rücktritt endet, linguistisch von der Zeit davor unterscheidet. Da die vorliegende Präsentation der Studie das Augenmerk auf die Darstellung der korpuslinguistischen Methodik legt und aufzeigen möchte, welche Analysemethoden fruchtbringend für Fragestellungen dieser Art genutzt werden können, müssen an dieser Stelle die Ausführungen zur (linguistischen) Skandalforschung beendet werden.³

1.1 Datengrundlage und Korpusdesign

Journalisten der Bild-Zeitung und der Süddeutschen Zeitung wurden im Mai 2012 für den Henri-Nannen-Journalistenpreis im Bereich "Investigation" für die Berichterstattung zur Wulff-Affäre nominiert und ausgezeichnet, wobei die Journalisten der Süddeutschen Zeitung die Annahme des Preises verweigerten, um gegen die Auszeichnung einer Boulevard-Zeitung zu protestieren.⁴ Der Preis und der Eklat sind ein Hinweis dafür, dass die beiden Zeitungen einen wichtigen Beitrag zur Berichterstattung leisteten und gleichzeitig zwei sehr unterschiedliche Pole von Journalismus abdecken. Aus diesem Grund wurden für die vorliegende Studie alle online verfügbaren Artikel dieser beiden Zeitungen zu Wulff verwendet.

Für die meisten korpuslinguistischen Untersuchungen ist es sinnvoll, mit einem Referenzkorpus zu arbeiten, das einer ähnlichen Domäne entstammt wie das Untersuchungskorpus, aber innerhalb dieser Domäne einen allgemeinen Sprachgebrauch repräsentiert. Mögliche Referenzkorpora wären z. B. das umfangreiche Deutsche Referenzkorpus (DeReKo) des IDS (Institut für Deutsche Sprache o.J.) oder das DWDS-Korpus (Berlin-Brandenburgische Akademie der Wissenschaften 2011), die beide die

³ Vgl. die Sammelbände von Burkhardt/Pape (2003) und Bulkow/Petersen (2011) für einen guten Überblick über die linguistische Skandalforschung.

⁴ Vgl. www.henri-nannen-preis.de/preistraeger_2012.php, Stand: 29. Juni 2012, und Schneider (2012). Alle im Folgenden genannten URLs wurden am 29. Juni 2012 das letzte Mal geprüft.

geschriebene deutsche Gegenwartssprache repräsentieren können.⁵ Der Nachteil dieser Korpora ist deren beschränkte Nutzungsmöglichkeit, da primär die online angebotenen Recherchewerkzeuge verwendet werden müssen und die Analysemöglichkeiten dadurch beschränkt sind. Daher wurde für die vorliegende Untersuchung folgendes Referenzkorpus gewählt: Alle online verfügbaren Artikel der Bild-Zeitung und der Süddeutschen Zeitung zu Angela Merkel, der gegenwärtigen Bundeskanzlerin Deutschlands. Es wäre besser, sämtliche (eventuell inländische) politische Berichterstattung der beiden Zeitungen als Referenzkorpus verwenden zu können, der technische Aufwand wäre jedoch deutlich größer gewesen. Texte zu Merkel stellen einen vertretbaren Kompromiss dar, denn:

1. Die Texte handeln genauso von einer politischen Persönlichkeit wie die Wulff-Texte.
2. Sie entstammen ebenfalls der Domäne der politischen Berichterstattung.
3. Die Prominenz Merkels garantiert eine große Anzahl an Artikeln.

Kritisch hingegen ist, dass es eine Korrelation zwischen Merkel und Wulff gibt: Die Kandidatur Wulffs als Bundespräsident geht auf Merkels Initiative zurück und während der Wulff-Affäre stand Merkel deswegen ebenso im medialen Rampenlicht. Es gibt deshalb eine Reihe von Artikeln, die in beiden Korpora vorkommen, da sie sowohl von Wulff als auch von Merkel handeln.

1.2 Korpusgewinnung und -aufbereitung

Die Gewinnung der Artikel erfolgte über die Recherchemöglichkeiten auf den Websites von Bild.de und Süddeutsche.de. Die Archivfunktionen erlauben eine Volltextsuche. Deshalb wurden die beiden Korpora über die Eingabe der Stichwörter "Merkel" und "Wulff" erstellt, was natürlich nur eine Annäherung an die gewünschten Korpora darstellt, da es namensgleiche Personen geben könnte. Die Prominenz der Personen rechtfertigt jedoch dieses Vorgehen.

Am Stichtag der Korpuszusammenstellung, dem 23. März 2012, wurden für das Untersuchungskorpus alle bis dahin in den Archiven verfügbaren Texte zum Stichwort "Wulff" und eine annähernd gleich große Anzahl Texte zum Stichwort "Merkel" heruntergeladen. Die Archive enthalten weit mehr Artikel zu Merkel als zu Wulff, weshalb das Herunterladen bei den Merkel-Texten rückwärts chronologisch nach etwa 1800 Artikeln pro Zeitung gestoppt wurde. Die Allgemeinen Geschäftsbedingungen der beiden Zeitungen erlauben die private Nutzung der Artikel und das vorübergehende Speichern auf dem eigenen Rechner, solange die Daten nicht gewerblich genutzt und nicht an Dritte weitergegeben werden.⁶ Das Herunterladen erfolgte automatisiert über die Unix-Programme "curl" und "wget" in Kombination mit "grep", "sort" und "uniq":

⁵ Wobei je nach Forschungsvorhaben allerdings genau geprüft werden muss, welcher Ausschnitt geschriebener Gegenwartssprache tatsächlich repräsentiert werden kann. Das DeReKo ist beispielsweise sehr presselastig und hat seinen Schwerpunkt im Zeitraum der 1980er-Jahre bis heute, ist mit etwa 5 Milliarden Wortformen jedoch das größte öffentlich verfügbare deutschsprachige Korpus. Die DWDS-Korpora umfassen neben Zeitungskorpora auch ein sogenanntes Kernkorpus, das mit 100 Millionen Wörtern alle Dekaden von 1900 bis 2000, gliedert nach unterschiedlichen Textsorten, abdeckt.

⁶ Vgl. www.sueddeutsche.de/verlag/agb-allgemeine-geschaeftsbedingungen-1.526836 und www.bild.de/corporate-site/agb/bild-de/artikel-agb-2952414.bild.html.

Der Nachteil der Privatnutzung gemäß den gültigen AGB ist, dass das Korpus nicht innerhalb der Forschergemeinschaft weitergegeben werden kann. Jedoch können die URLs der heruntergeladenen Artikel

1. Mit "curl" werden alle Suchresultateseiten einer Suche nach "Wulff" bzw. "Merkel" heruntergeladen.
2. Die Resultateseiten enthalten die Links zu den eigentlichen Artikeln. Diese sind mit speziellen HTML-Tags gekennzeichnet, sodass mit einer Suche mit "grep" alle Zeilen ausgegeben werden können, die einen Artikellink enthalten. In Verbindung mit einem Perl-Befehl (s.u.) werden die eigentlichen Links extrahiert und die URL wird so modifiziert, dass jeweils die Druckversion des Artikels erreicht werden kann, die keine unerwünschten Seitenelemente wie Navigation und Anreißer anderer Artikel enthält.
3. Die Liste der Links wird mit "sort" und "uniq" alphabetisch geordnet und vereinheitlicht, sodass jeder Link nur einmal vorkommt.
4. Mit "wget" werden nun die Artikel über die URL-Liste heruntergeladen, wobei die Abarbeitung der URL-Liste in zufälliger Reihenfolge und mit Wartepausen zwischen dem Laden der einzelnen Artikel geschieht, damit der Server nicht überlastet wird.

Zur weiteren Aufbereitung der Artikel (Extraktion des eigentlichen Artikeltextes, Extraktion der Metadaten) wurden die Texte mit TagSoup⁷, einem in Java programmierten XML-Parser, der auch mit nicht validem HTML⁸ zurechtkommt, verarbeitet und dann mit XSL-Stylesheets in eine valide XML-Struktur überführt, wie das folgende Beispiel zeigt:

```
<?xml version="1.0" encoding="UTF-8"?>
<text xmlns:xhtml="http://www.w3.org/1999/xhtml">
  <title>
    <kicker>Weltgrösste IT-Messe in Hannover</kicker>
    <headline>Die Cebit 2010 öffnet ihre Tore</headline>
  </title>
  <timestamp>
    <date>01.03.2010</date>
    <time>16:03 Uhr</time>
  </timestamp>
  <content>
    <p>In Hannover öffnet heute die weltgrößte Computermesse Cebit ihre
Pforten. Bundeskanzlerin Angela Merkel gibt heute Abend im Congress Centrum
um 18 Uhr den Startschuss für die Eröffnungsfeier – zusammen mit dem
spanischen Ministerpräsidenten José Luis Rodriguez Zapatero: Spanien ist in
diesem Jahr das Partnerland der Messe.</p>
    [...]
  </content>
</text>
```

Somit sind die Texte für die korpuslinguistische Bearbeitung vorbereitet. Die Texte wurden mittels zweier automatischer Wortarten-Tagger annotiert, mit dem TreeTagger (Schmid 1994,

weitergegeben werden: Sie finden sich auf der Begleitwebsite
www.bubenhofer.com/bubenhofer2012skandalisierung/.

⁷Vgl. ccil.org/~cowan/XML/tagsoup/

⁸ Der HTML-Code von Webseiten entspricht meistens nicht exakt den Vorgaben des W3C (World Wide Web Consortium, www.w3.org) und ist deshalb nicht valide (wird aber von den Browsern trotzdem korrekt dargestellt). Um einen solchen Code in eine valide XML-Struktur konvertieren zu können, muss ein Parser deshalb "robust" programmiert sein, damit er auch nicht valide Codes verarbeiten kann.

1995) unter Verwendung der Standard-Trainingsdatei und des Stuttgart-Tübingen-Tagsets⁹ (Schiller/Teufel/Thielen 1995) sowie dem RFTagger (Schmid/Laws 2008). Beide Tagger liefern pro Token Wortartinformationen. Der TreeTagger differenziert 55 unterschiedliche Wortarten und Interpunktionszeichentypen, der RFTagger jedoch rund 700 unterschiedliche Kombinationen von Wortart, Genus, Kasus und Numerus. Zudem wurden die Texte mit einem von der Forschergruppe *semtracks*¹⁰ entwickelten Programm auf Tempusformen hin annotiert. Die Annotation beruht auf der TreeTagger-Annotation und dem STTS-Tagset.

Die Korpora wurden nach dem Tagging in die Corpus Workbench (Evert/The OCWB Development Team 2010) importiert, um sie über die Corpus Query Language CQP abfragbar zu machen. Die Corpus Workbench bietet darüber hinaus Schnittstellen zu anderen Programmiersprachen an, so zu Perl (Wall/Christiansen/Orwant 2000). Dies erlaubt es, automatisiert Abfragen zu tätigen, die dann im vorliegenden Fall in der Statistikumgebung R (Adler 2010) weiterverarbeitet werden können.¹¹

Nach der korpuslinguistischen Aufbereitung umfasst das Wulff-Korpus gut 2,3 und das Merkel-Korpus gut 2,5 Millionen Wörter. Die Details sind den Tabellen 1 (Untersuchungskorpus) und 2 (Referenzkorpus) zu entnehmen, wobei Tabelle 1 die Daten nach den politisch wichtigen Abschnitten gliedert.

Periode	(unbekannt)	Bild	Süddeutsche	Total
Wulff Abgeordneter im Landtag (ab 1994; Daten ab 3. Dez. 2001)			2056	2056
Wulff Ministerpräsident Kabinett I (4. März 2003 bis 25. Februar 2008)		91464	386764	478228
Wulff Ministerpräsident Kabinett II (26. Februar 2008 bis 30. Juni 2010)		246027	351533	597560
Wulff Bundespräsidentschaft (1. Juli 2010 bis 17. Februar 2012)		537476	530670	1068146
Gauck Kandidatur (18. Februar bis 18. März 2012)		80639	98344	178983
Gauck Bundespräsidentschaft (seit 19. März 2012; Daten bis 23. März 2012)		16587	9994	26581
(ohne Datum)	2908			2908
Total	2908	972193 (1615)	1379361 (1724)	2354462 (3339)

Tabelle 1: Untersuchungskorpus – Anzahl Wörter (Anzahl Artikel in Klammern)

⁹ Vgl. www.ims.uni-stuttgart.de/projekte/corplex/TagSets/stts-table.html.

¹⁰ Vgl. www.semtracks.org.

¹¹ Inzwischen ist auch die R-Bibliothek "rcqp" von Bernard Desgraupes und Sylvain Loiseau verfügbar, mit der direkt aus R Abfragen in der Corpus Workbench gemacht werden können: cran.r-project.org/web/packages/rcqp/index.html; vgl. auch: www.bubenhofer.com/sprechtakel/2012/06/05/statistische-analysen-von-korpora-mit-r-direkt-auf-die-cwb-zugreifen/.

Jahr	(unbekannt)	Bild	Süddeutsche	Total
2007			701	701
2008			107498	107498
2009			165554	165554
2010		26626	174206	200832
2011		900904	620055	1520959
2012		210390	326555	536945
(ohne Datum)	2451			2451
Total	2451	1137920	1394569	2534940
		(1675)	(1801)	(3476)

Tabelle 2: Referenzkorpus – Anzahl Wörter (Anzahl Artikel in Klammern)

Wie die Tabellen 1 und 2 zeigen, sind Untersuchungs- und Referenzkorpus ähnlich groß, was den Vergleich einfacher macht. Die Zahlen zeigen bereits Unterschiede zwischen den Zeitungen: Der durchschnittliche Wulff-Artikel bei der Bild-Zeitung umfasst 602 Wörter, bei der Süddeutschen Zeitung hingegen 800 Wörter. Ein Artikel zu Merkel umfasst bei der Bild-Zeitung im Schnitt 679 und bei der Süddeutschen Zeitung 774 Wörter.

Die beiden Tabellen zeigen auch, dass es eine Reihe von unklassifizierten Artikeln gibt, die kein Datum und keine Zuordnung zu einer Zeitung aufweisen. Eine manuelle Kontrolle zeigte, dass es sich um Bild-Artikel aus Regionalausgaben handelt, die im Format leicht von den restlichen Artikeln abweichen und deshalb nicht korrekt aufbereitet wurden. In allen weiteren Berechnungen bleiben diese Artikel unberücksichtigt.

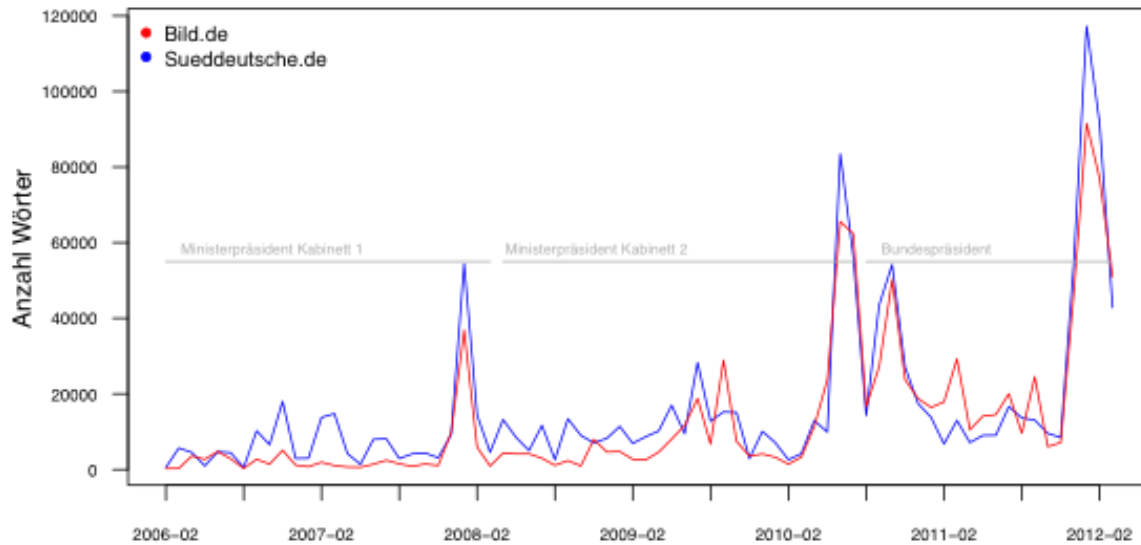
Aus rechtlichen Gründen (siehe Fußnote 7) kann das Korpus nicht weitergegeben oder veröffentlicht werden. Auf der diesen Artikel begleitenden Website¹² können jedoch die URLs der heruntergeladenen Artikel eingesehen werden, sodass das Korpus davon ausgehend selber erstellt werden kann.

2 Analysen

2.1 Überblick

Zunächst soll ein Überblick über die Publikationsdynamik zur Wulff-Affäre geschaffen werden. Abbildung 1 zeigt im zeitlichen Verlauf nach Perioden gegliedert die Anzahl publizierter Wörter zu Wulff in den beiden Zeitungen.

¹² Vgl. www.bubenhofer.com/bubenhofer2012skandalisierung/



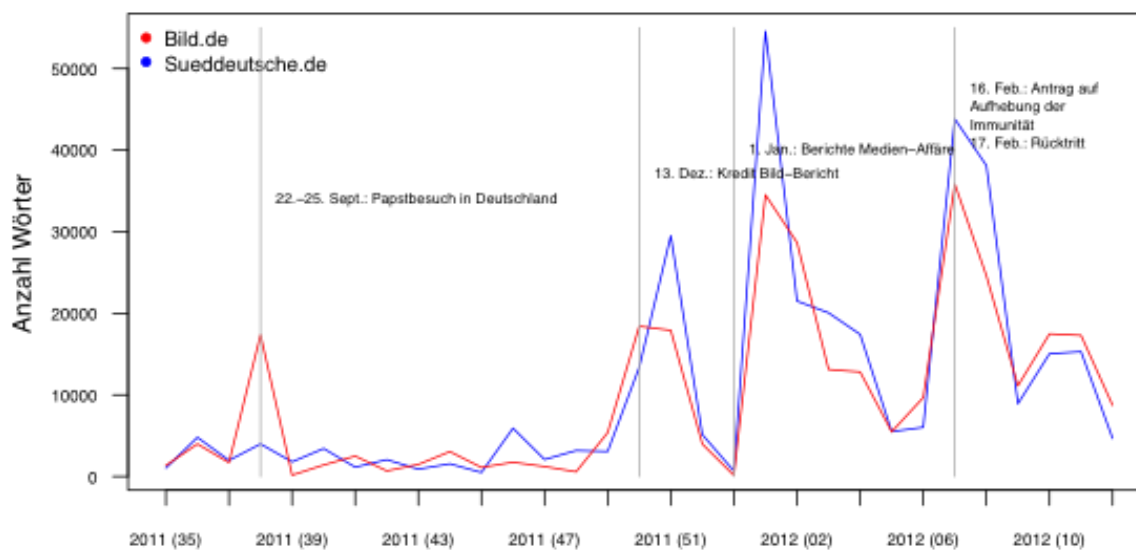
Werte monatsweise aggregiert; Bild.de: n=972.193, Süddeutsche.de: n=1.379.361

Abbildung 1: Anzahl Wörter in Artikeln zu Wulff auf Bild.de und Süddeutsche.de

Die Verlaufskurven ähneln sich: In Zeiten des Wahlkampfs (vor der zweiten Ministerpräsidentenschaft und vor der Bundespräsidentschaft) schnellen die Werte deutlich hoch, ebenso kurz vor dem Ende des Bundespräsidiums. Die Zeit der Bundespräsidentschaft ist ebenso durch besonders rege Berichterstattung geprägt, wobei gegen Ende, während der eigentlichen Wulff-Affäre, besonders hohe Werte erreicht werden. Damit ist die am 13. Dezember 2012 beginnende Skandalisierung sichtbar, als die Bild-Zeitung berichtete, Wulff habe am 25. Oktober 2008 von Edith Geerkens einen Privatkredit entgegengenommen und damit den Landtag im Februar 2010 in die Irre geführt.

Die Süddeutsche Zeitung publizierte meistens mehr und/oder längere Artikel als die Bild-Zeitung, mit Ausnahme in der Zeit des Bundespräsidiums bis zum eigentlichen Beginn der Affäre. Dies ist bemerkenswert, sind Bild-Artikel im Wulff-Korpus im Schnitt doch deutlich kürzer als die Artikel der Süddeutschen Zeitung (vgl. Tabelle 1 oben). Im Januar 2012, dem Beginn der "Medienaffäre", erreichen beide Zeitungen die höchsten Werte.

Abbildung 2 zeigt eine detailliertere Ansicht der letzten Wochen von Wulffs Präsidentschaft. Deutlich sind die Ausschläge ab Woche 50 im Jahr 2011 zu sehen. Der erste Ausschlag findet ab dem 13. Dezember 2011 statt, dem Publikationsdatum des Bild-Artikels zum Privatkredit. Der zweite Ausschlag erfolgt ab dem 1. Januar 2012, als die Berichte zu Wulffs Versuch, die Berichterstattung im Dezember zu verhindern, erschienen. Ebenfalls in der ersten Januarwoche, am 4. Januar 2012, fand das viel beachtete Doppelinterview der beiden öffentlich-rechtlichen Sender mit Wulff statt. Diese Woche scheint damit ein Kulminationspunkt zu sein, der die bisherige Spitze im Dezember 2011 übertrifft. Ein weiterer Höhepunkt der Berichterstattung erfolgt dann in Woche 7 ab dem 16. Februar, als die Staatsanwaltschaft Hannover den Antrag stellt, die Immunität Wulffs aufzuheben. Einen Tag danach kündigt Wulff seinen Rücktritt an.



Werte wochenweise aggregiert; Bild.de: n=303.421, Süddeutsche.de: n=336.980

Abbildung 2: Anzahl Wörter in Artikeln zu Wulff auf Bild.de und Süddeutsche.de gegen Ende der
Präsidentschaft

Interessant ist ein Ausschlag in der Bild-Berichterstattung in Kalenderwoche 37: Die meisten Wulff-Artikel in dieser Zeit handeln vom Papstbesuch in Deutschland, in dessen Rahmen der Bundespräsident ebenso immer wieder erwähnt wird. Auffällig ist, dass dies für die Süddeutsche Zeitung kein Thema ist. Die "Skandalisierungsspitzen" der Berichterstattung in den beiden Zeitungen sehen im Gegensatz dazu immer so aus, dass die Süddeutsche Zeitung ebenso breit (zudem etwas wortreicher) berichtet wie die Bild-Zeitung. Andere Wulff-Themen davor scheinen jedoch weitgehend dem Boulevard vorbehalten zu sein.

2.2 Sprachgebrauchsmuster

Sprachgebrauchsmuster sind Einheiten bestehend aus mehreren linguistischen Entitäten wie Wörtern, Wortarten, Lemmata, Tempus- oder Modusinformatoren etc., die in einem Korpus musterhaft, also rekurrent, verwendet werden und pragmatisch gedeutet werden können (Bubenhofer 2009: 43–44.). Eine Operationalisierungsmöglichkeit dieser Sprachgebrauchsmuster ist die Berechnung von komplexen n-Grammen, also Ketten von Wörtern: Komplexe n-Gramme sind eine erweiterte Form von n-Grammen (ebd.: 118; Manning/Schütze 2002: 192), die nicht nur aus einer Folge von Wortformen bestehen, sondern auch aus einer Kombination von Wortformen und Wortartinformationen bestehen können.¹³ Während ein n-Gramm beispielsweise als Wortformenkette *so verbringen wir* definiert ist, werden bei der Berechnung von komplexen n-Grammen die Wortarten miteinbezogen, sodass eine Reihe von ähnlichen Wortformen-n-Grammen abstrakter als "so – finites Verb – Personalpronomen" gefasst wird.¹⁴

¹³ Vgl. das ähnliche Konzept der Collocations von Stefanowitsch und Gries (2003).

¹⁴ Zur Berechnung der komplexen n-Gramme wurde das von der Forschergruppe semtracks (www.semtracks.org) entwickelte Programm *cwb-n-grams* verwendet, das eine Erweiterung zur Open Corpus Workbench darstellt (Evert/The OCWB Development Team 2010).

Bei der hohen Zahl an kombinatorisch möglichen komplexen n-Grammen¹⁵ ist es nicht mehr sinnvoll, die n-Gramme nach ihrer absoluten Frequenz zu ordnen und sich bloß auf die häufigsten n-Gramme zu beschränken.¹⁶ Denn die häufigsten n-Gramme sind interpretatorisch kaum interessant, da sie bloß häufige syntaktische Strukturen widerspiegeln. Stattdessen müssen die n-Gramme eines Korpus mit den n-Grammen eines Referenzkorpus verglichen werden: Mittels eines statistischen Signifikanztests kann ermittelt werden, welche n-Gramme im Vergleich zum Referenzkorpus besonders typisch für das jeweilige Untersuchungskorpus sind (Bubenhofner 2009: 149–152; vgl. für eine Übersicht Culpeper 2009). Dadurch kann die Menge der n-Gramme massiv reduziert werden. Durch die Wahl des Referenzkorpus wird darüber hinaus kontrolliert, bezüglich welcher Analyseebene die Spezifik der n-Gramme berechnet werden soll: So ist auf diachroner Ebene denkbar, verschiedene Zeitabschnitte miteinander zu vergleichen, auf synchroner Ebene beispielsweise Textsorten, Autoren, Institutionen, kommunikative Praktiken etc.

Im Fall der vorliegenden Studie sind mehrere Vergleiche zwischen Untersuchungs- und Referenzkorpus denkbar. Einerseits kann das Wulff-Korpus mit dem Referenzkorpus der Merkel-Texte kontrastiert werden, um die typischen Sprachgebrauchsmuster der Wulff-Berichterstattung zu eruieren. Interessanter ist jedoch die Frage, welche Sprachgebrauchsmuster typisch für die Skandalisierung sind. Dann umfasst das Untersuchungskorpus nur die Wulff-Texte aus der Skandalzeit und das Referenzkorpus beinhaltet nicht nur die Merkel-Texte, sondern auch alle Wulff-Texte aus der Vor-Skandalzeit. Denn so entsteht ein Referenzkorpus, das sich nur auf der Ebene Skandal/nicht Skandal unterscheidet, nicht aber bezüglich Personen (Wulff/Merkel), Textsorte (Zeitungstexte) und Domäne (Politik). So sollten die gefundenen Differenzen tatsächlich einen Zusammenhang mit der Skandalisierung bei Wulff zu tun haben. Zusätzlich ließen sich auch Vergleiche zwischen den Zeitungen anstellen, um zu sehen, ob die beiden Zeitungen unterschiedliche sprachliche Mittel der Skandalisierung anwenden. In dieser Studie soll der Fokus aber nicht darauf liegen.

Im Folgenden werden nun die beiden unten stehenden Korpora miteinander verglichen:

1. Skandalkorpus: Alle Texte aus Bild-Zeitung und Süddeutscher Zeitung zu Wulff in der Zeit vom 13. Dezember 2011 bis 17. Februar 2012.
558.440 Wörter
2. Referenzkorpus: Alle Texte aus Bild-Zeitung und Süddeutscher Zeitung zu Wulff oder Merkel ausschließlich der Skandaltex-te.
4.250.285 Wörter

¹⁵ Die Anzahl der kombinatorisch möglichen n-Gramme hängt von der Anzahl der unterschiedlichen Wortformen (also der Types) im Korpus ab. Bei der Berechnung von 4-Grammen und angenommenen 50.000 Types gibt es $50.000^4 = 6.25E + 18$ mögliche 4-Gramme. Bei 4-Grammen, die aus einer beliebigen Kombination von Wortform und Wortart bestehen können, ergeben sich elf unterschiedliche Kombinationen von Wortform und Wortart pro 4-Gramm, also elfmal mehr 4-Gramme als bei der Berechnung von konventionellen 4-Grammen, die auf der Wortform basieren.

¹⁶ Bereits bei der Berechnung von "konventionellen" n-Grammen, die auf Wortformen basieren, ist dieses Verfahren nur bedingt sinnvoll.

Für die Berechnung der komplexen n-Gramme wurden alle Wortformen und Wortarten des STTS-Tagsets berücksichtigt; n lag zwischen 3 und 5, die n-Gramme haben also eine Länge von drei bis fünf Wörtern, wobei keine Lücken zugelassen sind.

n-Gramm	Freq. Skandal	Freq. Referenzkorpus	p
NE (/\$(52/CARD)/\$(¹⁷	116	107	< 0
NE NE (/\$(52/CARD	112	110	< 0
Kredit-/TRUNC und NN	89	40	< 0
(/\$(52/CARD)/\$(119	115	< 0
NE (/\$(52/CARD	117	117	< 0
der/ART Kredit-/TRUNC und NN	40	16	< 0
500/CARD 000/CARD NN	59	33	< 0
der/ART Kredit-/TRUNC und	41	16	< 0
NE)/\$(:/\$. "/\$(55	67	< 1e-08
das/ART NN des/ART NN	198	594	< 1e-08
wegen/APPR des/ART NN der/ART	39	35	< 1e-07

Tabelle 3: Liste der häufigsten 11 von 652 komplexen n-Grammen, die signifikant häufiger im Skandalkorpus als im Nicht-Skandalkorpus vorkommen (absolute Frequenzen und Wahrscheinlichkeit p für einen zufälligen Frequenzunterschied)

Bei einem Signifikanzniveau von $p < 0,5$ ergaben sich 652 für das Skandalkorpus signifikante komplexe n-Gramme (vgl. Tabelle 3). Für das Nicht-Skandalkorpus ergaben sich 4633 komplexe n-Gramme; diese werden bei der weiteren Analyse jedoch nicht weiter beachtet.

Die Liste der 652 komplexen n-Gramme muss nun manuell gesichtet und linguistisch interpretiert werden. Welche n-Gramme interessant sind, hängt vom Untersuchungsinteresse ab. Im vorliegenden Fall stehen n-Gramme im Vordergrund, die sich pragmatisch hinsichtlich einer besonderen Funktion im Rahmen eines Skandals deuten lassen. Um die teilweise abstrakten n-Gramme besser verstehen zu können, wurden für jedes n-Gramm zehn zufällig ausgewählte Beispiele für Realisierungen im Korpus ausgegeben. Zudem wurde berechnet, in welchem Verhältnis die Anzahl der (unterschiedlichen) Realisierungen zum n-Gramm (Muster) steht:

VAFIN ADJD VAPP ,/\$,

ist klar geworden ,

wäre besser gewesen ,

wäre gut gewesen ,

wäre aufrichtig gewesen ,

sei besorgt gewesen ,

sei falsch gewesen ,

ist klar geworden ,

war bekannt geworden ,

TTR 2 (Muster/Realisierungen): 3.23

TTR 3 (Muster/unterschiedliche Realisierungen): 6.67

17 Die Kürzel der Wortarten folgen dem STTS-Tagset (Schiller/Teufel/Thielen 1995; vgl. linglit193.linglit.tu-darmstadt.de/CQP/Bundestag/help-tagset.html). Die Einheiten mit Schrägstrich bedeuten, dass beide Elemente gegeben sein müssen: "der/ART" bedeutet, dass das Element aus der Wortform *der*, getaggt als bestimmter Artikel, besteht.

Das n-Gramm (fett gedruckt) besteht also aus einem Hilfsverb, einem adverbialen oder prädikativen Adjektiv, einem auxiliären Partizip Perfekt und einem Komma. Realisiert wird das n-Gramm im Korpus mit Ausdrücken wie *ist klar geworden [Komma]*, *wäre besser gewesen [Komma]* etc. Ferner werden zwei Type-Token-Verhältnisse berechnet, wobei diese mit TTR 2 und TTR 3 benannt werden, um sie nicht mit dem ansonsten üblichen Type-Token-Maß zu verwechseln, das auf Wörtern beruht. Hier ist der Type das komplexe n-Gramm (Muster) und die Realisierungen sind die Tokens, wobei unterschieden wird zwischen der Anzahl der Realisierungen generell (TTR 2) und den unterschiedlichen Realisierungen (TTR 3). Je tiefer die Werte, desto produktiver ist das Muster, also desto mehr unterschiedliche Realisierungen gibt es. Je höher die Werte, desto weniger (unterschiedliche) Realisierungen gibt es.

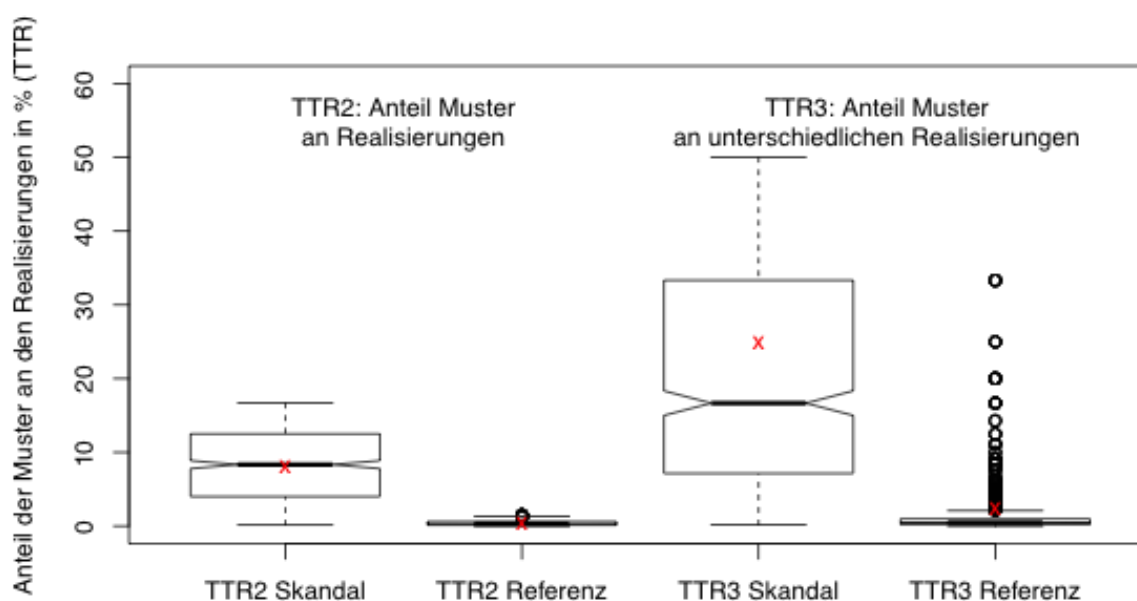


Abbildung 3: Verhältnis der Muster zu den Realisierungen im Skandal- und im Referenzkorpus

Wie Abbildung 3 zeigt, unterscheiden sich die TTR-Maße sehr stark zwischen dem Skandal- und dem Referenzkorpus. Dies bedeutet, dass die n-Gramme im Skandalkorpus sehr viel weniger produktiv sind, also viel weniger (unterschiedliche) Realisierungen pro Muster gemessen werden. Im Skandalkorpus ist die Sprache also formelhafter als im Referenzkorpus. Doch welche n-Gramme sind nun typisch für das Skandal-Korpus?

2.2.1 Bezeichnung des Skandals

1. **der/ART Kredit-/TRUNC und NN** (TTR 2: 2,5; TTR 3: 20)
 der Kredit- und Urlaubsaffäre
 der Kredit- und Medienaffäre
 der Kredit- und Medien-Affäre
2. **500/CARD 000/CARD NN** (TTR 2: 1,69; TTR 3: 33,3)
 500 000 Euro

3. **NN bei/APPR NN NN** (TTR 2: 4; TTR 3: 6,67)
Anruf bei Bild-Chefredakteur Kai
Drohanruf bei Bild-Chefredakteur Kai
Bundespräsidenten bei Bild-Chefredakteur Kai
Mailbox-Nachricht bei Bild-Chefredakteur Kai
Razzia bei Ex-Wulff-Sprecher Glaeseker

2.2.2 Enthüllung und Aufbereitung

4. **in/APPR PPOSAT NN als/KOKOM** (TTR 2: 2.0; TTR 3: 11,11)
in seiner Zeit als
in seine Rolle als
5. **ADJA NN zu/APPR PPOSAT** (TTR 2: 4; TTR 3: 5,3)
kritischen Fragen zu seinem
engen Bezug zu seinem
nähere Details zu meiner
undurchsichtigen Verbindungen zu seinem

Typisch für das Skandalkorpus sind n-Gramme, die Elemente des Enthüllens und der späteren Aufbereitung des Skandals benennen. So steht z. B. die Wendung *in seiner Zeit als [(niedersächsischer) Ministerpräsident]* im Zusammenhang mit vergangenen Handlungen Wulffs, die Skandalgegenstand sind:

6. Wulff **hatte in seiner Zeit als** Ministerpräsident Niedersachsens zur Ablösung des umstrittenen Privatkredits für sein Wohnhaus, den ihm die Unternehmergattin Edith Geerkens gewährt hatte, von der BW-Bank ein besonders zinsgünstiges kurzfristiges Darlehen über etwa 500 000 Euro erhalten. (Bild.de, Stand: 29.12.2011: "Kredit-Affäre um Bundespräsident Wulff: Anzeige gegen BW-Bank")

2.2.3 Bewertung

7. **VVINF ,/\$, ob/KOUS NE** (TTR 2: 6,25; TTR 3: 6,67)
urteilen , ob Wulff
nachgehen , ob Wulff
prüfen , ob Wulff
8. **Ein/ART NN ,/\$, PRELS** (TTR 2: 2,4; TTR 3: 4,6)
Ein Mann , der
Ein Präsident , der

Weitere typische n-Gramme bewerten die Vorgänge des Skandals, also die Sachverhalte (abweichendes Verhalten), den Skandalisierten und sein Handeln, wie beispielsweise der folgende Beleg zeigt:

9. "Wenn es keine Sicherheiten gab – zum Beispiel einen Grundbucheintrag – und Herr Wulff also zu den genannten Konditionen gar keinen Kredit bei einer Bank bekommen hätte, muss man einen Verstoß bejahen", sagte Ulrich Battis von der Humboldt-Universität zu Berlin. Allerdings sei für ihn noch der Amtsbezug unklar. "Man muss sehr genau **prüfen, ob** das eine rein private Angelegenheit war – oder eine, die einen

engen Bezug zu seinem Amt als Ministerpräsident hatte." (Süddeutsche.de, Stand: 17.12.2011: "Affäre um Privatkredit: Verfassungsrechtler werfen Wulff Rechtsbruch vor")

Dieser Beleg zeigt zudem den Einbezug eines Experten, der ein Urteil über das Verhalten Wulffs abgeben soll, sich allerdings nicht festlegen möchte und auf einer genauen Prüfung beharrt. Dies ist kein Einzelfall: Statt eines Experten können es der Ältestenrat des niedersächsischen Landtags (Süddeutsche.de, Stand: 20.12.2011: "Bundespräsident unter Druck: Maschmeyer finanzierte Werbung für Wulff-Buch") oder die Generalstaatsanwaltschaft (Süddeutsche.de, Stand: 31.01.2012: "Bundespräsident in Bedrängnis: SPD verhöhnt "Pinocchio" Wulff") sein, die in eine ähnliche Rolle gebracht werden.

Auf die besondere pragmatische Funktion von *Ein Mann/Präsident, der* wird weiter unten noch eingegangen.

2.2.4 Entschuldigen, bedauern und rechtfertigen

10. **dass/KOUS ADV ein/ART ADJA** (TTR 2: 7,7; TTR 3: 25)
dass hier ein falscher [Eindruck entstehen konnte]
11. **Das/PDS VAFIN PPER VVINF** (TTR 2: 14,3; TTR 3: 50)
Das hätte ich sagen
Das hätte ich vermeiden
12. **VAFIN ADJD VAPP ,/\$,** (TTR 2: 3,22; TTR 3: 6,67)
ist klar geworden ,
wäre besser gewesen ,
wäre aufrichtig gewesen ,
13. **Das/PDS VMFIN man/PIS VVINF** (TTR 2: 11,11; TTR 3: 25)
Das kann man schlucken
Das muss man verantworten
Das kann man erklären

Auffällig sind Floskeln der Entschuldigung und der Rechtfertigung. Sie sind auch deshalb so typisch (und darüber hinaus weisen sie hohe TTR-Werte auf), weil es sich um Aussagen von Wulff handelt, die er während des TV-Interviews am 4. Januar 2012 machte und die mehrmals zitiert worden sind.

In den Wendungen *wäre besser gewesen* und *wäre aufrichtig gewesen* zeigen sich Teile eines Sprechakts des Bedauerns (vgl. für die Abgrenzung zur Entschuldigung Liedtke 2003: 76–77), wobei im ersten Fall die Mehrzahl der Belege das TV-Interview zitiert:

14. "Ich erkenne an, dass hier ein falscher Eindruck entstehen konnte. Ich bedauere das", erklärte der Präsident. "**Es wäre besser gewesen**, wenn ich auf die Anfrage der niedersächsischen Abgeordneten im Landtag über die konkreten Fragen hinaus auch diesen privaten Vertrag mit Frau Geerkens erwähnt hätte, denn in der Sache hatte und habe ich nichts zu verbergen." (Bild.de, Stand: 15.12.2011: "Nach dem Wirbel um das Privat-Darlehen für den Bundespräsidenten: Wulff äußert sich: "Hätte Kredit erwähnen sollen"")

2.2.5 Folgen

15. VAFIN das/ART NN des/ART (TTR 2: 3,33; TTR 3: 5,3)
hat das Amt des
hat das Ansehen des
ist das Amt des
wird das Amt des
wird das Antlitz des
16. **im/APPRART NN VVINF VMFIN** (TTR 2: 2,78; TTR 3: 4,76)
im Amt halten könnte
im Amt halten kann
im Amt bleiben will
im Amt halten wollen

Schließlich finden sich Floskeln, die die Folgen des Skandals beschreiben oder darüber spekulieren. Eine erste Gruppe von Wendungen dreht sich darum, dass das Amt "Schaden genommen" habe:

17. Jetzt gibt es erste Rücktrittsforderungen gegen das Staatsoberhaupt: Ulrich Maurer, stellvertretender Vorsitzender der Linken im Bundestag, fordert Wulff auf, die Konsequenzen aus der Kredit- und Medienaffäre zu ziehen: "Christian Wulff **hat das Amt des Bundespräsidenten** und damit die Bundesrepublik massiv **beschädigt**. Das Maß ist voll." (Süddeutsche.de, Stand: 03.01.2012: "Bundespräsident in der Kritik: Linke fordert Rücktritt von Wulff")

Belege dieser Art zeigen, wie die Medien ein "Entrüstungsklima" schaffen, indem die sie Akteure zitieren, die sich entrüsten (Beckmann 2006: 72). Ähnlich werden Einschätzungen Dritter über die zukünftige Entwicklung wiedergegeben:

Nach dem Antrag der Staatsanwaltschaft auf Aufhebung der Immunität von Christian Wulff rechnen mehrere CSU-Vorstandsmitglieder mit einem Rücktritt des Bundespräsidenten. Wulff werde sich angesichts dieser neuen Entwicklung wohl kaum noch **im Amt halten** können, hieß es übereinstimmend von mehreren CSU-Vorständen. (Bild.de [dpa], Stand: 17.02.2012: "CSU-Vorstandsmitglieder rechnen mit Rücktritt Wulffs")

2.2.6 Weitergehende Analysen

Sind die typischen n-Gramme identifiziert, können weitere interpretatorische Schritte im Anschluss daran prüfen, ob sie den Status eines Sprachgebrauchsmusters haben. Exemplarisch soll auf das n-Gramm "Ein NN, der" (Beispiel 8) hingewiesen werden.¹⁸ Die relativ niedrigen TTR-Werte verweisen auf eine erstaunliche Produktivität des Musters. Diese geht in erster Linie auf den Platzhalter des Nomens im n-Gramm zurück. Tabelle 4 zeigt, wie und mit welchen Frequenzen der Platzhalter im Korpus gefüllt wird.

¹⁸ Die Analysen zu diesem Muster entstanden in Kooperation mit Katrin Hein und werden in Hein/Bubenhofer (im Druck) ausführlich dargestellt.

Realisierung von NN	#	Weitere Realisierung mit Frequenz 1
Mann	9	Abend, Abschied, Affront, Anblick, Argument,
Präsident	7	Außenminister, Beispiel, Bild-Reporter, CDU-
Bundespräsident	7	Wähler, CSUler, Eintrag, Erlebnis,
Land	3	Geschäftsmodell, Gesetz,
Staatsoberhaupt	3	Gewinnabführungsvertrag, Halbfinale, Haushalt,
Termin	2	Katholik, Konsortium, Kraftakt, Männerbündnis,
Augenzeuge	2	Medium, Ministerpräsident, Name, Plan, Radler,
Vorgang	2	Satz, Schmunzeln, Schritt, SPD-
Umstand	2	Präsidiumsmitglied, Spiel, Sportsmann, Staat,
Thema	2	Tiefschlag, Überblick, Unternehmensmodell, Verhalten, Vorwurf, Waisenkind, Wunsch

Tabelle 4: Realisierungen von NN (Lemmata) im komplexen n-Gramm "Ein/Art NN ,/\$, PRELS"

Die häufigsten Realisierungen verweisen (dies zeigt sich bei einer Durchsicht der Belege) auf Wulff selber, der hier *Mann*, *Präsident* oder *Bundespräsident* genannt wird. Die Realisierungen lassen sich weiter differenzieren. Die wichtigste Differenzierung ist wohl, ob NN durch eine Personenbezeichnung oder eine Sache besetzt wird. Verengt man das Muster auf NN_{Personenbezeichnungen}, lässt sich die Hypothese aufstellen, dass dieses Muster eine spezielle Skandalisierungsfunktion aufweist. Im Korpus manifestieren sich zwei unterschiedliche Möglichkeiten zur Einbettung des Musters. Zum einen kommt es als selbstständiger Aussage- oder Fragesatz vor, der keiner weiteren Ergänzung bedarf:

18. Auch für den Kommentator des Bonner "General-Anzeigers" ist es schwer vorstellbar, dass Wulff im Amt bleibt. Im Fall staatsanwaltlicher Ermittlungen könne er zwar formal im Berliner Schloss Bellevue bleiben. "**Ein Bundespräsident, der** darauf wartet, ob gegen ihn Anklage erhoben wird?", fragt das Blatt. (Bild.de, Stand: 17.02.2012: "Das schreiben die Zeitungen: "Den Zeitpunkt, sich in Würde zu verabschieden, hat Wulff verpasst"")

Andererseits kommt das Muster in Verbindung mit einem Nebensatz vor, der in inhaltlicher Hinsicht die Funktion übernimmt, eine direkte Schlussfolgerung aus den im Muster konstatierten negativen Vorfällen zu ziehen:

19. **Ein Mann, der** sich in dieser Weise verdächtig gemacht hat, kann nicht mehr Bundespräsident sein. Wulff fehlte die Einsicht, dies rechtzeitig zu erkennen. (Süddeutsche.de, Stand: 18.02.2012: "Wulff-Rücktritt: Merkels dritter Versuch muss sitzen")

Der indefinite Artikel scheint die Funktion zu haben, die Aussage zu generalisieren und formuliert somit eine Art Schlussregel, die die Verletzung eines Verhaltensmaßstabes thematisiert (in Anlehnung an Toulmin 1958; vgl. auch ausführlich Egger 2006):

Allgemeine Schlussregel: Wer X tut, der kann eine bestimmte Funktion nicht mehr ausführen.

Schlussregel im speziellen Fall: Wer einer kriminellen Tat beschuldigt wird, kann nicht Bundespräsident sein.

Datum: Wulff wird einer kriminellen Tat beschuldigt.

Schlussfolgerung: Wulff kann nicht mehr Bundespräsident sein.

Dieses Muster findet sich auch in anderen Kontexten, wo es eine ähnliche Funktion aufweist:

20. Der Direktor des Simon Wiesenthal Centers in Jerusalem, Efraim Zuroff, forderte Oettinger zum Rücktritt auf. Die Äußerungen Oettingers seien "absolut unfassbar", erklärte Zuroff. **Ein Ministerpräsident, der** die Nazi-Vergangenheit Filbingers leugne und reinwasche, sei "untragbar". (Süddeutsche.de, Stand: 13.04.2007: "Filbinger-Verteidigung: Merkel rückt von Oettinger ab")

Allerdings finden sich auch positiv konnotierte Kontexte, wie der folgende Beleg aus dem DeReKo exemplarisch zeigt:

21. Von Thabo Mbeki wird allgemein eine härtere Gangart erwartet: [...] **"Ein Präsident, der** den Leuten in den Hintern tritt" – so nannte ihn eine Zeitung kürzlich lobend. (E99/JUN.14844, Zürcher Tagesanzeiger, 02.06.1999, S. 5, Ressort Ausland: "Dem ANC ist der Sieg sicher")

Daher erfüllt das Muster nicht nur eine negative Skandalisierungsfunktion, sondern kann auch eine positiv verstärkende Wirkung haben.

Diese Ausführungen müssen an dieser Stelle genügen; in Hein/Bubenhofer (im Druck) wird dieses Muster ausführlich in einem konstruktionsgrammatischen Rahmen diskutiert, um den pragmatischen Mehrwert aufzuzeigen.

2.3 Wortarten- und Tempusformenverteilung

Die Sprachgebrauchsmuster lassen bereits vermuten, dass während der Skandalisierung bestimmte Wortarten generell häufiger vorkommen als andere. Ebenso ist es plausibel, dass bestimmte Tempusformen, z. B. Passivformen, seltener im Skandalkorpus vorkommen als in den anderen Daten. Dies lässt sich statistisch prüfen: Die zu prüfende Hypothese behauptet, dass eine Reihe von Wortarten und Tempusformen in den Texten zwischen dem 13. Dezember 2011 und dem Rücktritt Wulffs am 17. Februar 2012 in Relation zur jeweiligen Textlänge signifikant häufiger vorkommen als in der Zeit davor.

Text-ID	Datum	Periode	Korpus	Quelle	Wörter	ADJA	ADJA_rel	ADJD	ADJD_rel
http://...	15.09.10	Skandal	Merkel	Süddeutsche	808	31	0,0384	17	0,0210
http://...	29.01.08	Referenz	Wulff	Bild	235	13	0,0553	14	0,0595
...									

Tabelle 5: Die Datengrundlage für die Analysen zur Wortartenverteilung

Die zu analysierenden Daten liegen im Format wie in Tabelle 5 ausschnittsweise dargestellt vor.¹⁹ Pro Text wurden für alle Wortarten (gemäß STTS-Tagset) und für alle Tempusformen die relativen Anteile berechnet. Zusätzlich wurde jeder Text einer Periode zugeordnet: Texte

¹⁹ Die Tabelle wurde mittels eines Perl-Skripts und der Corpus Workbench Perl-API erstellt. Mit dem CQP-Befehl "[]" wurde nach allen Wörtern im Korpus gesucht und dann mit "tabulate" eine Tabelle erstellt, in der pro Wort die gewünschten Metadaten (insbesondere Datum und Quelle), die Wortart und die Tempusform stehen. Das Perl-Skript aggregierte diese Daten zu einer Tabelle wie oben in Tabelle 5 dargestellt.

des Wulff-Korpus wurden der Skandalperiode bzw. der Vor-Skandalperiode zugeordnet, die Texte des Referenzkorpus der Referenzperiode.

Die unabhängige Variable "Periode" ist demnach kategorial (vor Skandal, Skandal, Referenz), die abhängigen Variablen der Wortarten- und Tempusanteile sind metrisch skaliert (0–1). Die einfachste Möglichkeit ist der Vergleich der Mittelwerte mit sogenannten Boxplots, die auch die Streuung und die Ausreißer visualisieren. Anschließend kann bei unklaren Fällen noch eine einfaktorielle ANOVA (analysis of variance, Gries 2008: 266–275) durchgeführt werden.

2.3.1 Wortarten

Bei den Wortarten ergeben sich praktisch keine signifikanten Unterschiede, wenn man alle Texte beider Zeitungen berücksichtigt. So könnte man z. B. annehmen, dass Personalpronomen in Skandaltexten eine größere Rolle spielen als in den anderen Daten. Der Boxplot in Abbildung 4 lässt bereits vermuten, dass ein signifikanter Unterschied zwischen den Skandal- und den anderen Texten vorhanden ist, was eine einfaktorielle ANOVA bestätigt ($F_{2, 7223} = 16,9$; $p < 0,001^{***}$). Doch die Varianzen in den drei Korpora unterscheiden sich bereits signifikant, wie ein durchgeführter Bartletts-Test zeigt (Bartletts $K^2 = 26,8639$; $df = 2$; $p = 1,467e-06$), sodass die Anwendung des Tests kritisch ist (Gries 2008: 268). Und die Variable der Korpuszugehörigkeit erklärt nur gerade 0,4% der Varianz (multiples $R_2: 0,004657$; adjustiertes $R_2: 0,004382$).

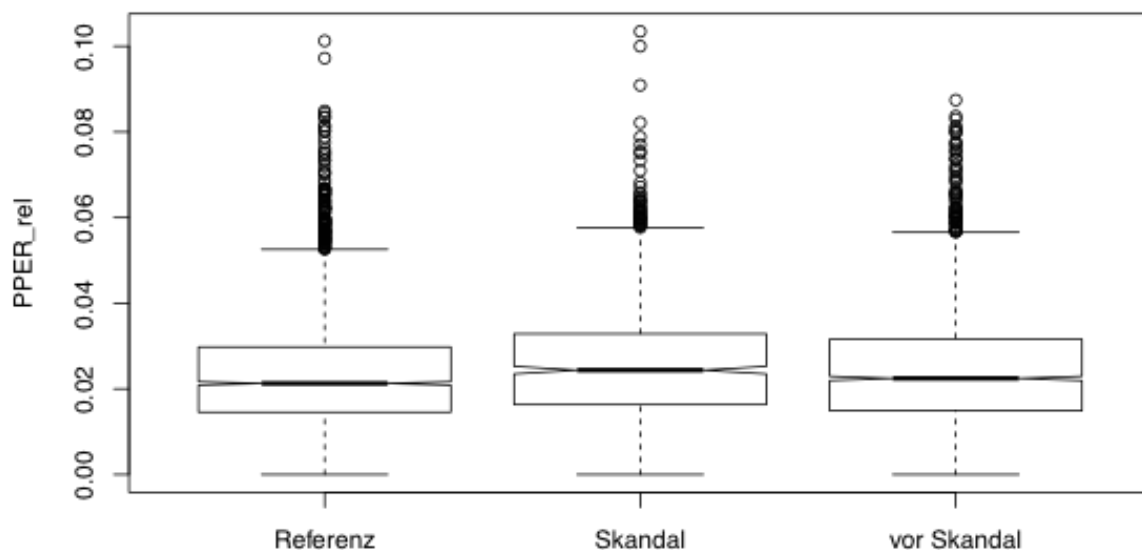


Abbildung 4: Irreflexive Personalpronomen in den verschiedenen Korpora: Es sind nur sehr minimale Unterschiede sichtbar

Warum dies so ist, zeigt sich, wenn die Verwendung der Wortarten der beiden Zeitungen verglichen wird. Bei einigen Wortarten unterscheidet sich die Verteilung signifikant, wie z. B. der Gebrauch von Adjektiven zeigt (vgl. Abbildung 5). Doch diese Fragestellung ist in der vorliegenden Studie nicht von Interesse.

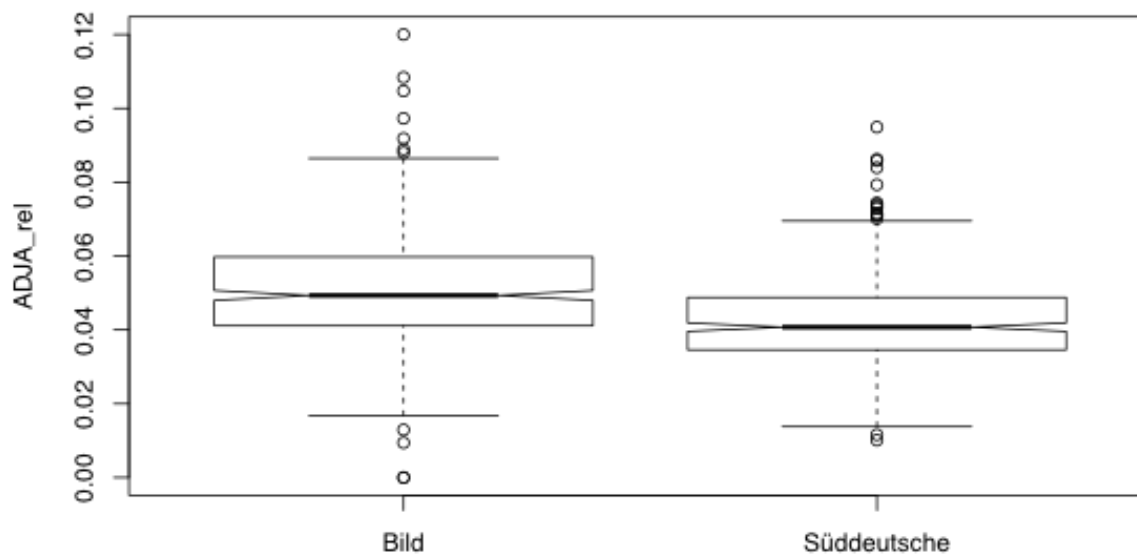


Abbildung 5: Adjektiv-Anteile in der Bild-Zeitung und der Süddeutschen Zeitung über alle Korpora hinweg

2.3.2 Tempusformen

Im Gegensatz zu den Wortarten ergeben sich bei der Verwendung der unterschiedlichen Tempusformen teilweise signifikante Unterschiede in den Texten. Allerdings lassen sich praktisch keine signifikanten Unterschiede zwischen dem Skandalkorpus und den anderen Daten finden, dafür aber signifikante Unterschiede des Wulff-Korpus insgesamt im Vergleich zum Referenzkorpus. Die deutlichsten Differenzen betreffen Aktiv und Passiv.

Niedrige Werte im Wulff-Korpus:

- **Perfekt Indikativ Passiv** ($F_{2, 7223} = 3824$; $p < 0,001^{***}$; mult. $R^2 = 0,5143$; adj. $R^2 = 0,5142$)
 - *Ex-Bundespräsident Christian Wulff (52) **ist** am Donnerstagabend mit einem Großen Zapfenstreich aus dem Amt **verabschiedet worden**.* (Bild.de, Stand: 08.03.2012: "Zapfenstreich mit Vuvuzela-Lärm: Wulffs Abschied aus Bellevue")
 - *Beim Angriff einer US-Drohne im pakistanischen Grenzgebiet zu Afghanistan **sind** am Dienstag mindestens 20 mutmaßliche Aufständische **getötet worden**.* (Süddeutsche.de, Stand: 16.11.2010, Politik kompakt: "Cameron preist Britannien als Weltmacht")
- **Präsens Indikativ Passiv** (vgl. Abbildung 6) ($F_{2, 7223} = 8071$; $p < 0,001^{***}$; mult. $R^2 = 0,6909$; adj. $R^2 = 0,6908$)
 - *Die Bürger urteilen differenzierter, als dies oftmals in den Medien **unterstellt wird**.* (Süddeutsche.de, Stand: 04.01.2012: "Bundespräsident in der Kritik: "Ich kann Wulffs Verhalten nicht mehr verstehen"")
 - *Derzeit **werden** weitere fünf Marder für die deutschen Kampftruppen an den Hindukusch **verlegt**.* (Süddeutsche.de, Stand: 19.11.2010, Politik kompakt: "Fidel Castro: Bin nicht mehr Parteichef")

- **Futur I Indikativ Aktiv mit Modalverb** ($F_{2, 7223} = 2752$; $p < 0,001^{***}$; mult. $R^2 = 0,4324$; adj. $R^2 = 0,4323$)
 - *"Wir werden uns mächtig strecken müssen", weiß Gino Leonhard (39), der Verlegenheits-Spitzenkandidat der FDP. (Bild.de, Stand: 29.08.2011: "Landtagswahl Mecklenburg-Vorpommern: Kehrt Rot-Rot im Norden zurück?")*

Hohe Werte im Wulff-Korpus:

- **Präsens Indikativ Aktiv** (vgl. Abbildung 7) ($F_{2, 7223} = 3135$; $p < 0,001^{***}$; mult. $R^2 = 0,4647$; adj. $R^2 = 0,4646$)
 - *Die "Berliner Zeitung" berichtet von Ungereimtheiten beim Leasen eines neuen Audi Q3. (Bild.de, 18.01.2012: "Kredit- und Medien-Affäre: Wulff veröffentlicht Journalisten-Anfragen und Antworten")*
 - *"Die Sache stinkt zum Himmel." (Bild.de, Stand: 18.01.2012: "Kredit- und Medien-Affäre: Wulff veröffentlicht Journalisten-Anfragen und Antworten")*
- **Präsens Konjunktiv I Aktiv** ($F_{2, 7223} = 6075$; $p < 0,001^{***}$; mult. $R^2 = 0,6272$; adj. $R^2 = 0,6271$)
 - *Es sei vor allem "Mutti Merkel", die nicht eingestehen wollte, dass sie mit der Nominierung von Christian Wulff zum Staatsoberhaupt ein "unglückliches Händchen" bewiesen habe und sich bis zum Ende nicht von ihm lösen konnte. (Süddeutsche.de, 22.02.2012: "Nach dem gewonnenen Präsidenten-Poker: FDP sehnt Gauck-Effekt herbei")*
- **Futur I Indikativ Passiv mit Modalverb** ($F_{2, 7223} = 2257$; $p < 0,001^{***}$; mult. $R^2 = 0,3846$; adj. $R^2 = 0,3844$)
 - *Der Begriff "politischer Grund" ist ein unbestimmter Rechtsbegriff, dessen Bedeutung im Wege der Auslegung erschlossen werden muss. (Bild.de, Stand: 20.03.2012: "Ehrensold für Christian Wulff: Die Begründung des Bundespräsidialamtes")*
 - *Und daher ist es bedauerlich, wenn ein Präsident erst zum Bedauern getragen werden muss. (Süddeutsche.de, Stand: 16.12.2011: "Bundespräsident in der Kreditaffäre: Wenn Amt und Charakter verschmelzen")*

Die Wulff-Berichterstattung hat sich also bezüglich der Tempusformen mit dem Skandal nicht stark geändert. Über Wulff wird generell stärker in Aktiv- als Passivformen berichtet, abgesehen von Modalverbssätzen im Futur, wo im Wulff-Korpus die Passivformen häufiger sind als im Referenzkorpus. Die Präsens-Aktiv-Formen (und gleichzeitig der verminderte Gebrauch von Passiven) im Wulff-Korpus führen zu einem unmittelbaren, raschen Stil, könnten aber auch mit den vielen Zitaten zusammenhängen, die in direkter Rede berichtet werden. Die Zitate machen sich auch im häufigen Gebrauch von Konjunktiv-I-Formen bemerkbar, die normalerweise in indirekter Rede verwendet werden.

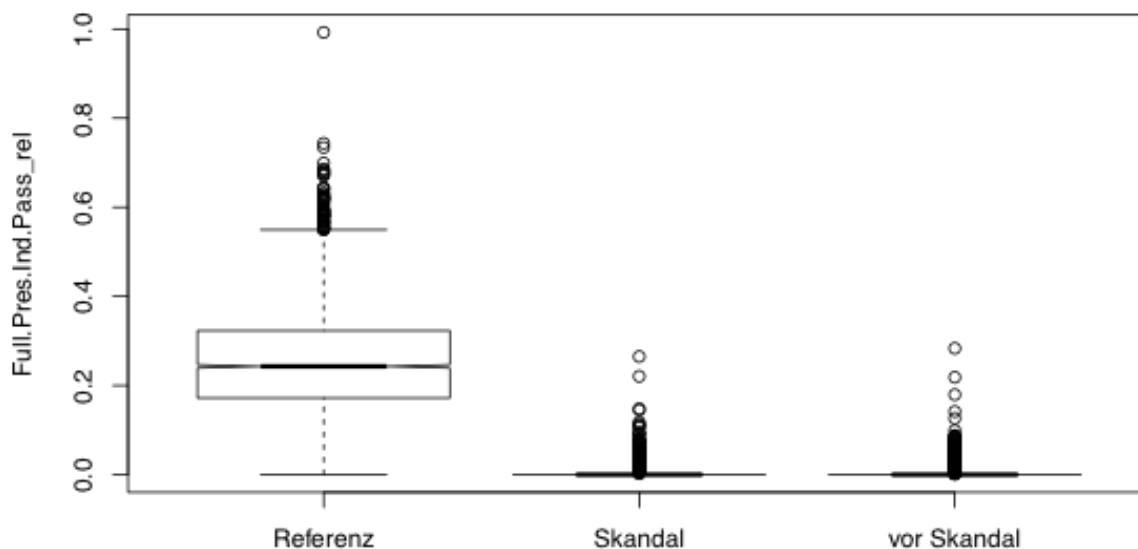


Abbildung 6: Präsens/Indikativ/Passiv-Formen in den drei Korpora

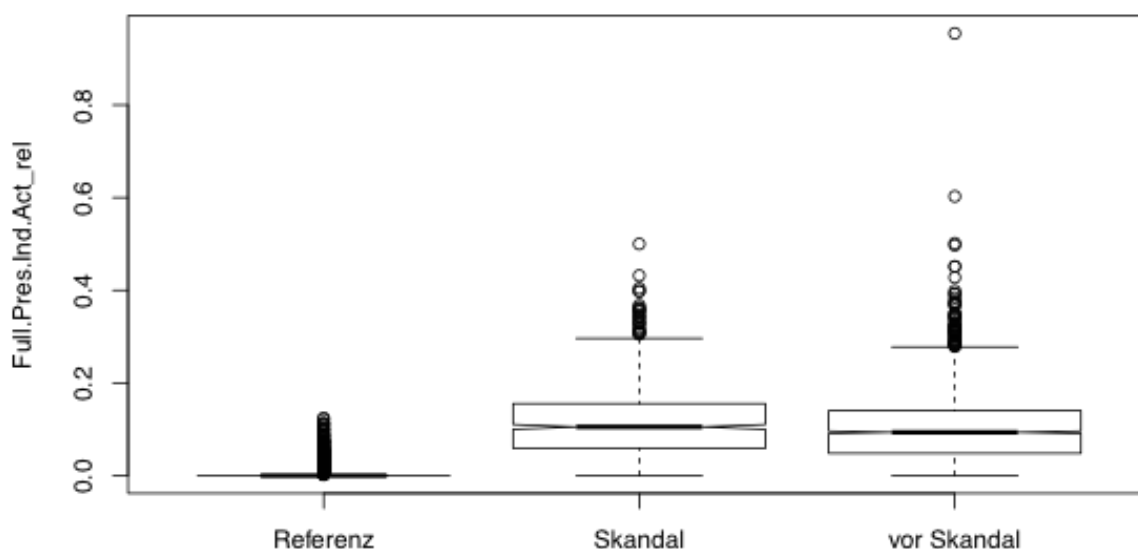


Abbildung 7: Präsens/Indikativ/Aktiv-Formen in den drei Korpora

Die seltenere Verwendung von Futur I Aktiv mit Modalverben im Wulff-Korpus ist schwierig zu plausibilisieren. Diese Form wird im Referenzkorpus häufig in einem kommentierenden, prognostischen Sinn, manchmal auch in direkter Rede verwendet, wie z. B. im folgenden Beleg:

22. *Die Wahrheit ist, niemand **wird** diese Sicherheit **gewährleisten können**.* (Bild.de, Stand: 06.12.2011: "Bombenanschlag in Kabul: Ein Foto zeigt die ganze Afghanistan-Tragödie")
23. *Scheitert ein Folgeabkommen für Kyoto, **wird** Angela Merkel sich in Deutschland massive Kritik **gefallen lassen müssen**.* (Süddeutsche.de, Stand: 17.12.2009: "Kanzlerin in der Krise: Angela Merkel – Chefin verzweifelt gesucht")

Allenfalls ist diese prognostische Komponente im Wulff-Korpus seltener, da Kommentare eher die Vergangenheit und die aktuellen Geschehnisse thematisieren anstelle der weiteren Zukunft. Doch diese Hypothese müsste genauer geprüft werden, ebenso wie die Möglichkeit, dass es andere Korrelationen mit dieser Tempusform gibt, die stärker sind als die Korrelation mit den drei getesteten Korpora.

3 Fazit

Die Berechnung von typischen komplexen n-Grammen machte auf Formulierungsmuster aufmerksam, die für die Wulff-Affäre in den Zeitungen Bild und Süddeutsche typisch waren. Die Analysen zu den Wortarten- und Tempusverteilungen offenbarten Tempusformen, die in diesen Zeitungstexten signifikant häufiger vorkommen, hingegen keine belastbaren Resultate zum Wortformengebrauch. Aus den Ergebnissen lassen sich Schlüsse für die typischen linguistischen Merkmale von Presstexten zu Skandalen ziehen:

1. Floskelhafte Sprache: Im Skandalkorpus gibt es häufiger Sprachgebrauchsmuster mit niedriger Produktivität, die also mit immer ähnlichen Floskeln realisiert werden. Das hängt damit zusammen, dass bestimmte Aussagen (z. B. des Skandalisierten) immer wieder zitiert werden. Die Vermutung liegt nahe, dass diese Aussagen selber eine Skandalisierungsfunktion einnehmen können.
2. Sprachgebrauchsmuster mit Skandalisierungsfunktion: Dazu gehört beispielsweise das Muster "Ein NN_{Personenbezeichnung}, der", das im Fall des Wulff-Skandals häufig verwendet wurde und zeigen soll, dass der Skandalisierte seine Aufgabe als Bundespräsident nicht mehr erfüllen kann.
3. Hinweise auf kommentierende Strukturen: Einige der typischen n-Gramme sind Teil von kommentierenden und bewertenden Sprechakten.
4. Hinweise auf einen unmittelbaren, mit vielen Zitaten und direkter Rede durchsetzten Schreibstil im gesamten Wulff-Korpus: Dies lässt die Verteilung der Tempusformen vermuten, wo im Wulff-Korpus signifikant häufiger Aktiv- statt Passivformen vorkommen.
5. Verhalten der Zeitungen: Die beiden sehr unterschiedlichen Zeitungen Bild und Süddeutsche berichten normalerweise unterschiedlich ausführlich über politische Themen. Während des Skandals jedoch engagierten sich beide ähnlich stark und setzten zu den gleichen Zeitpunkten besonders starke Akzente.

Die vorliegende Studie wollte Anregungen dazu geben, welche korpuslinguistischen Methoden zu den ansonsten gängigen wie Schlagwortanalyse, Kollokationsanalysen, (halb)manuelle Kategorisierung von Belegen etc. zusätzlich fruchtbar gemacht werden können, um der linguistischen Besonderheiten der Skandalisierung habhaft zu werden. Interessant wäre aber im Anschluss die Frage, inwiefern sich die identifizierten Mittel generalisieren lassen und auch für andere Skandale gelten. Dafür müsste man das Korpus erweitern und die Berichterstattung zu weiteren Skandalen hinzufügen.

Diese Generalisierung könnte ein grundsätzliches Problem der präsentierten Studie lösen: Gemäß den Kriterien der linguistischen Skandalforschung kann man zwar davon ausgehen, dass es sich bei der Wulff-Affäre um einen Skandal handelt. Trotzdem wird damit eine starke Prämisse für die weitere korpuslinguistische Analyse gesetzt, die nicht zwingend ist, vor

allem was die Grenzen des Skandals betrifft: Wann genau beginnt der Skandal? Ist er mit dem Rücktritt des Präsidenten zu Ende? Vielleicht zeigen sich Tendenzen der sprachlichen Skandalisierung auch in Bereichen, die nicht als Skandal definiert werden würden. Um dies herauszufinden, ist es jedoch unumgänglich, anhand eines eng definierten Trainingskorpus, von dem angenommen wird, dass es einen prototypischen Skandal repräsentiert, die spezifischen linguistischen Merkmale festzustellen. Dann können weitere Skandale hinzugezogen und die Liste der Merkmale kann ergänzt und korrigiert werden, um – und dies wäre die eigentlich interessante Aufgabe – mit maschinellen Methoden Skandalisierungstendenzen zu erkennen, bevor der Skandal als Skandal wahrgenommen wird.

Literaturverzeichnis

- Adler, Joseph (2010): *R in a Nutshell*. Sebastopol: O'Reilly.
- Beckmann, Susanne (2006): "Der Skandal – ein komplexes Handlungsspiel im Bereich öffentlicher Moralisierungskommunikation". In: Girnth, Heiko/Spieß Constanze (eds.) (2006): *Strategien politischer Kommunikation. Pragmatische Analysen*. Berlin, Erich Schmidt: 61–78.
- Bergmann, Jörg/Luckmann, Thomas (1999): *Kommunikative Konstruktion von Moral. Band 1: Struktur und Dynamik der Formen moralischer Kommunikation*. Wiesbaden: VS Verlag.
- Berlin-Brandenburgische Akademie der Wissenschaften (2011): *DWDS*. Berlin: www.dwds.de, Stand: 29. Juni 2012.
- Bubenhofer, Noah (2009): *Sprachgebrauchsmuster. Korpuslinguistik als Methode der Diskurs- und Kulturanalyse*. Berlin/New York: de Gruyter. (= *Sprache und Wissen* 4).
- Bubenhofer, Noah/Scharloth, Joachim (2013): "Korpuslinguistische Diskursanalyse: Der Nutzen empirisch-quantitativer Verfahren". In: Warnke, Ingo/Meinhof, Ulrike/Reisigl, Martin (eds.) (2013): *Diskurslinguistik im Spannungsfeld von Deskription und Kritik*. Berlin, Akademie-Verlag: 147–168.
- Bulkow, Kristin/Petersen, Christer (2011): *Skandale: Strukturen und Strategien öffentlicher Aufmerksamkeitserzeugung*. Wiesbaden: VS Verlag.
- Burkhardt, Armin/Pape, Kornelia (eds.) (2003): *Politik, Sprache und Glaubwürdigkeit: Linguistik des politischen Skandals*. Göttingen: Vandenhoeck & Ruprecht.
- Culpeper, Jonathan (2009): "Keyness: Words, parts-of-speech and semantic categories in the character-talk of Shakespeare's *Romeo and Juliet*". *International Journal of Corpus Linguistics* 14/1: 29–59.
- Eggler, Marcel (2006): *Argumentationsanalyse textlinguistisch. Argumentative Figuren für und wider den Golfkrieg von 1991*. Tübingen: Niemeyer. (= *Reihe Germanistische Linguistik* 268).
- Evert, Stefan/The OCWB Development Team (2010): *The IMS Open Corpus Workbench (CWB) CQP Query Language Tutorial*. cwb.sourceforge.net/documentation.php, Stand: 29. Juni 2012.
- Gries, Stefan Thomas (2008): *Statistik für Sprachwissenschaftler*. Göttingen: Vandenhoeck & Ruprecht. (= *Studienbücher zur Linguistik* 13).
- Hein, Katrin/Bubenhofer, Noah (im Druck): "Korpuslinguistik konstruktionsgrammatisch. Diskursspezifische n-Gramme zwischen statistischer Signifikanz und semantisch-pragmatischem Mehrwert". In: Lasch, Alexander/Ziem, Alexander (eds.) (im Druck):

- Konstruktionsgrammatik IV: Konstruktionen als soziale Konventionen und kognitive Routinen*. Tübingen: Stauffenburg.
- Hitzler, Ronald (1989): "Skandal ist Ansichtssache. Zur Inszenierungslogik ritueller Spektakel in der Politik". In: Ebbighausen, Rolf/Neckel, Sighard (eds.) (1989): *Anatomie des politischen Skandals*. Frankfurt am Main, Suhrkamp: 334–354.
- Holly, Werner (2003): "Die Ordnung des Skandals. Zur diskursanalytischen Beschreibung eines "Frame" am Beispiel der "CDU-Spendenaffäre"". In: Burkhardt, Armin/Pape, Kornelia (eds.) (2003): *Politik, Sprache und Glaubwürdigkeit: Linguistik des politischen Skandals*. Göttingen, Vandenhoeck & Ruprecht: 47–68.
- Institut für Deutsche Sprache (o. J.): *Das Deutsche Referenzkorpus DeReKo*. www.ids-mannheim.de/kl/projekte/korpora/, Stand: 29. Juni 2012.
- Liedtke, Frank (2003): "Entschuldigung – ein sprachliches Ritual für Skandalisierte". In: Burkhardt, Armin/Pape, Kornelia (eds.) (2003): *Politik, Sprache und Glaubwürdigkeit: Linguistik des politischen Skandals*. Göttingen, Vandenhoeck & Ruprecht: 69–86.
- Manning, Christopher D./Schütze, Hinrich (2002): *Foundations of Statistical Natural Language Processing*. Cambridge (MA): The MIT Press.
- Neckel, Sighard (1989): "Das Stellhölzchen der Macht. Zur Soziologie des politischen Skandals". In: Ebbighausen, Rolf/Neckel, Sighard (eds.) (1989): *Anatomie des politischen Skandals*. Frankfurt am Main, Suhrkamp: 55–80.
- Scharloth, Joachim/Bubenhof, Noah (2011): "Datengeleitete Korpuspragmatik: Korpusvergleich als Methode der Stilanalyse". In: Felder, Ekkehard/Müller, Marcus/Vogel, Friedemann (eds.) (2011): *Korpuspragmatik. Thematische Korpora als Basis diskurslinguistischer Analysen von Texten und Gesprächen*. Berlin/New York, de Gruyter: 195–230.
- Schiller, Anne/Teufel, Simone/Thielen, Christine (1995): *Guidelines für das Tagging deutscher Textcorpora mit STTS*. Stuttgart: Universität Stuttgart.
- Schmid, Helmut (1994): *Probabilistic Part-of-Speech Tagging Using Decision Trees*. Stuttgart: Universität Stuttgart.
- Schmid, Helmut (1995): *Improvements in Part-of-Speech Tagging with an Application to German*. Stuttgart: Universität Stuttgart.
- Schmid, Helmut/Laws, Florian (2008): *Estimation of Conditional Probabilities With Decision Trees and an Application to Fine-Grained POS Tagging*. Stuttgart: Universität Stuttgart. www.ims.uni-stuttgart.de/projekte/gramotron/PAPERS/COLING08/Schmid-Laws.pdf, Stand: 29. Juni 2012
- Schneider, Jens (2012): "Eklat bei Journalisten-Ehrung: SZ-Redakteure lehnen Henri-Nannen-Preis ab". *Süddeutsche Zeitung* 15. Mai 2012. www.sueddeutsche.de/medien/eklat-bei-journalisten-ehrung-sz-redakteure-lehnen-henri-nannen-preis-ab-1.1355532, Stand: 29. Juni 2012.
- Stefanowitsch, Anatol/Gries, Stefan Thomas (2003): "Collostructions: Investigating the Interaction of Words and Constructions". *International Journal of Corpus Linguistics* 8/2: 209–243.
- Toulmin, Stephen (1958): *The Uses of Argument*. Cambridge: Cambridge University Press.
- Wall, Larry/Christiansen, Tom/Orwant, Jon (2000): *Programming Perl: There's more than one way to do it*. Sebastopol: O'Reilly.

Noah Bubenhofer: Skandalisierung korpuslinguistisch. Ein empirisch-linguistischer Blick auf 55
die Berichterstattung zur "Wulff-Affäre"

Wikimedia Foundation (2012): *Lemma "Wulff-Affäre"*. de.wikipedia.org/wiki/Wulff-Aff%C3%A4re, Stand: 29. Juni 2012.