

Erscheint voraussichtlich 2007 in
den Kongressakten von: Congreso
Internacional de Fraseología y
Paremiología Santiago de
Compostela/Lugo, 19 al 22 de
septiembre de 2006

«Es liegt in der Natur der Sache...» Korpuslinguistische Untersuchungen zu Kollokationen in Argumentationsfiguren

Noah Bubenhofer
Universität Zürich, Schweiz

Abstract

Argumentationsfiguren sind abstrakte Konzepte auf der sprachlichen Tiefenstruktur, die sich aber über Indikatoren auf der Textoberfläche andeuten. Diese Indikatoren sind häufig das, was die Phraseologie unter «Kollokationen» fasst und die Korpuslinguistik als «rekurrente lexikalische Einheiten» beschreibt. Es sind also Wortgruppen, deren Typik sich statistisch berechnen lässt. Der Beitrag zeigt, wie mit einem induktiven («corpus-driven») statt alleinig deduktiven Vorgehen («corpus-based») überraschende Hinweise auf Argumentationsfiguren in Leserbriefen gefunden werden können.

Argumentation schemes are abstract concepts at the linguistic deep structure which, however, are visible at the surface structure of a text via indicators. These indicators often appear in the form of «collocations» (in terms of phraseology) or «recurrent lexical items» (in terms of corpus linguistics). Thus, they are typical clusters of tokens that can be calculated statistically. This paper shows how to find surprising evidence for argumentation schemes in «letters to the editor» by using an inductive («corpus-driven») approach rather than a deductive («corpus-based») approach only.

Key words: Argumentationsfiguren, Korpuslinguistik, corpus-driven, Multi Word Units

1 Einleitung: Oberflächen- und Tiefenstruktur

Die Erforschung von Argumentationsfiguren hat eine lange Tradition, die in der klassischen Rhetorik begründet liegt. Doch auch die Linguistik beschäftigt sich seit Anbeginn mit Argumentation in Texten und später auch in gesprochener Sprache. Grundlegende These dieser Forschung ist zumeist, dass im konkreten Sprechen hervorgebrachte Argumente sich zu abstrakten Typen der Argumentation zusammenfassen lassen (vgl. stellvertretend für die linguistische Sicht: Toulmin 1958, Kopperschmidt 1989, Kienpointner 1992). Diese abstrakten Typen liegen auf der sprachlichen Tiefenstruktur und werden auf der sprachlichen Oberfläche zwar oft ähnlich, aber doch selten gleich repräsentiert.

Die Untersuchung solcher Argumentationsfiguren hat z.B. in der Diskurslinguistik Niederschlag gefunden. Diskurs ist hier im

Foucault'schen Sinn (Foucault 2000) gemeint als ein Geflecht von Texten und Äußerungen, die in einem intertextuellen Bezug stehen. Oft wird dieser intertextuelle Bezug thematisch definiert. So zeigen z.B. Wengeler (2003) oder Spitzmüller (2005) sehr detailliert anhand des Migrations- bzw. Anglizismendiskurses, welche Argumentationsfiguren und Topoi im Diskurs verwendet werden (können).

Unter einer weniger diskurs-, dafür textlinguistischen Perspektive arbeitet Egger (2006) sehr nahe am Text die Argumentationen heraus, die in zwei prototypischen Leitartikeln für und wider den Golfkrieg von 1991 ausgebreitet werden. Er plädiert dafür, Makro- und Mikrostruktur des Textes bei der Analyse zu trennen, und verfolgt letztere besonders genau.

Diesen Ansätzen ist Folgendes gemeinsam: Bei der semantischen Interpretation der Textoberfläche wird diese relativ rasch verlassen, da sofort die Typisierung und Kategorisierung auf der Tiefenstruktur erfolgt. Dieser Schritt ist legitim. Doch bleiben dabei Phänomene auf der Textoberfläche, die für eine argumentationstheoretische Analyse interessant sein könnten, unbeachtet. Dass aber der genaue Blick auf die Textoberfläche lohnt, beweist Egger (2006) letztlich mit seiner textnahen Analyse.

Eines dieser Phänomene auf der Textoberfläche ist der typisierte Sprachgebrauch, der sich in der Rekurrenz von lexikalischen Einheiten (Teubert 2005: 5) zeigt. Die Phraseologie hat dieses Phänomen zu ihrem Untersuchungsgegenstand gemacht und dabei eine differenzierte Systematik entwickelt, mit der sich z.B. nach Burger (1998) Kollokationen, Idiome, Sprichwörter und anderes mehr unterscheiden lassen und die es möglich macht, den semantischen und pragmatischen Wert dieses formelhaften Sprachgebrauchs zu untersuchen.

Der vorliegende Aufsatz möchte zeigen, wie mit dem Wissen der Phraseologie und den Methoden der Korpuslinguistik der Analyse von Argumentationsfiguren gedient werden kann. Damit ist auch ein methodischer Perspektivenwechsel verbunden, der von einer eher deduktiven zu einer induktiven Sicht führt.

2 Text als Daten: Ein korpuslinguistischer Zugang

Zunächst ist die Rekurrenz von lexikalischen Einheiten ein statistisches Phänomen. Die Korpuslinguistik kennt schon seit längerer Zeit Methoden, um anhand der Frequenz von Wörtern und Wortgruppen zu berechnen, welche Einheiten auf der sprachlichen Oberfläche rekurrent sind (Manning/Schütze 2000: 149ff.). Dafür stehen unterschiedliche statistische Maße zur Verfügung. Im einfachsten Verfahren lässt man sich alle im Korpus möglichen Wortgruppen (auch: *Multi Word Units* oder *n-Gramme* genannt) berechnen und nach Frequenz sortieren. In komplexeren Verfahren kann man die so gewonnenen Multi Word Units nach ihrer statistischen Signifikanz gewichten: Je nach Eigenfrequenz der einzelnen Wörter kann die Unwahrscheinlichkeit berechnet werden, mit der genau diese Wortgruppe bei zufälliger Verteilung der Token entstehen würde.

Nötige Basis solcher Berechnungen ist jedoch ein genügend großes Korpus an Sprachdaten. Mit der immer einfacheren Verfügbarkeit von elektronischen Sprachdaten (zumindest schriftlicher Sprache) und den für die meisten Zwecke mehr als genügenden Rechen- und Speicherkapazitäten können nun entsprechend große Korpora verwaltet und analysiert werden.

Diese neuen *technischen* Voraussetzungen ermöglichen aber auch neue *methodische* Zugänge. In der Korpuslinguistik werden zwei grundsätzliche Zugänge unterschieden: Untersuchungsfragen können *corpus-based* oder *corpus-driven* angegangen werden. In der (linguistischen) Forschung finden sich viele Beispiele für corpus-based-Vorgehen, bei denen Korpora eine Art Nachschlagewerk für vorher genau definierte Suchanfragen darstellen. So interessiert man sich z.B. für die Verteilung eines Begriffs (in verschiedenen Varianten) in einem diachron ausgerichteten Korpus.

Bei einem corpus-driven-Vorgehen versucht man jedoch gerade ohne im Vorhinein genau definierte Suchanfragen (und damit Kategorien und Klassifikationen) auszukommen:

While corpus linguistics may make use of the categories of traditional linguistics, it does not take them for granted. It is the discourse itself, and not a language-external taxonomy of linguistic entities, which will have to provide the categories and classifications that are needed to answer a given research question. This is the corpus-driven approach.
(Teubert 2005: 4)

Es wäre natürlich naiv zu glauben, bei corpus-driven-Ansätzen käme man gänzlich ohne vorherige Festlegungen aus. Doch der Kontrast zu corpus-based-Untersuchungen ist offensichtlich: In Tabelle 1 sind Beispiele für Suchanfragen aufgeführt.

Tabelle 1: Beispiele für Suchanfragen in corpus-based- bzw. corpus-driven-Verfahren.

	corpus-based	corpus-driven
Suchen	Suche nach Kollokationen zu «Terror» in den Artikeln der Neuen Zürcher Zeitung im Jahr 2005.	Berechnen einer Liste aller statistisch signifikanten Multi Word Units in den Artikeln der Neuen Zürcher Zeitung im Jahr 2005.
Vergleichen	Vergleich der Kollokationen zu «Terror» in den Artikeln der Neuen Zürcher Zeitung im Jahr 2000 verglichen mit dem Jahr 2005.	Vergleich der berechneten Listen aller statistisch signifikanten Multi Word Units in den Artikeln der Neuen Zürcher Zeitung in den Jahren 2000 und 2005.

Wenn man corpus-driven vorgeht, ist das natürlich mit der Hoffnung verbunden, auf interessante Phänomene zu stoßen, an die man zu Beginn der Untersuchung nicht dachte. So führt die Berechnung aller signifikanten Multi Word Units in einem (Teil-)Korpus zu Erkenntnissen zur Art der Formelhaftigkeit der Sprache. Und zwar nicht

nur zur Formelhaftigkeit dort, wo man sie erwartet, sondern auch anderswo.

Die im Folgenden dargestellten Resultate einer Korpusanalyse beruhen auf einer an Stubbs (2003: 230ff.) angelehnten Methode:

1. Bestimmung der Token im Korpus: Ein Token ist durch einen Wortabstand oder ein beliebiges Satzzeichen begrenzt.
2. Alle möglichen Multi Word Units der Länge n im Korpus werden bestimmt. Wahlweise werden auch Lücken zwischen den Token zugelassen, wobei aber die gesamte Multi Word Unit die Satzgrenze nicht überschreitet.
3. Die generierte Liste an Multi Word Units wird alphabetisch sortiert und das Vorkommen der Einheiten gezählt. Nur Multi Word Units mit einer bestimmten Mindestfrequenz (hier: 3) werden berücksichtigt.¹
4. Die fertige Liste wird nach Frequenzen sortiert und in eine Datenbank importiert. Hier können auch die Belege der einzelnen Multi Word Units angezeigt werden.

3 Phraseologismen als Indikatoren für Argumentationsfiguren

3.1 Deduktiv: corpus-based

In argumentativen Texten oder Gesprächen fällt die formelhafte Sprachverwendung oft auf. Intuitiv fallen mir z.B. Phrasen ein wie:

... auf der einen Seite ... auf der anderen Seite ...
... dabei geht völlig vergessen, dass ...
... es war schon immer so, dass ...
... es ist nun einmal so, dass ...
... wenn Ihnen X wirklich wichtig ist, dann ...
usw.

Wir können uns sofort – jeweils ganz unterschiedliche – Kontexte vorstellen, in denen sie verwendet werden könnten. Und diese Kontexte kann man sich so ausmalen, dass man sich damit auch die mit den Phrasen verbundenen Argumentationsfigur vor Augen führen kann. So könnte z.B. die Phrase «... es ist nun einmal so, dass ...» mit dem Zweck verwendet werden, eine strittige Aussage zu rechtfertigen, indem ein Zustand X als Faktum definiert wird, das in den Augen des Sprechers zur Norm werden soll. Eine Gegnerin könnte das als unzulässigen Schluss vom Faktischen zum Normativen (auch *naturalistischen Fehlschluss*) auffassen.

¹ Hier können wahlweise die Frequenzen auch der einzelnen Token und Teilwortgruppen berechnet und statistische Maße wie *Log Likelihood Ratio* oder *Mutual Information* etc. darauf angewandt werden. In unkomplizierter Weise erledigt das Ted Pedersens *Ngram Statistics Package (NSP)*: <http://www.d.umn.edu/~tpederse/nsp.html>.

Befragen wir ein Korpus: Das von der Suchmaschine Google indizierte Web. Die Suche nach «es ist nun einmal so» führt zu über 33'000 Treffern.² Der Beleg einer der ersten Treffer stammt aus einem Interview und lautet:

Wowereit: Ich halte diese Unterscheidung von Industrie- und Dienstleistungsarbeitsplätzen für unsinnig. Ich kämpfe um jeden Arbeitsplatz, in der Industrie wie in der Dienstleistung und **es ist nun einmal so**, dass weniger Qualifizierte, die früher in der Industrie beschäftigt werden konnten, heute bei Dienstleistern wie Call-Centern oder im Tourismus beschäftigt werden müssen. Was soll also die ganze Diskussion?³

Es ist hier nicht der Ort, eine detaillierte Argumentationsanalyse zu leisten. Doch ist – und das nicht nur bei diesem Beleg – klar, dass die Phrase in einem argumentativen Zusammenhang verwendet wird. Und es handelt sich dabei um einen Schluss vom Faktischen («weniger Qualifizierte müssen heute bei Dienstleistern beschäftigt werden») zum Normativen («eine Diskussion darüber lohnt sich nicht») über die Schlussregel: «Wenn etwas ist, wie es ist, dann muss es auch so sein und man muss es akzeptieren.»

3.2 Induktiv: corpus-driven

Die obige Recherche ist ein Beispiel für einen Zugang, der corpus-based ist. Nun wollen wir den umgekehrten Weg gehen: Basis ist ein Korpus von 1357 Leserbriefen aus dem Schweizer Tages-Anzeiger⁴ (vgl. Tabelle 2).

Tabelle 2: Die Eckdaten des Untersuchungskorpus.

Zeitspanne	Anzahl Artikel	Anzahl Token
1. Januar bis 31. August 2006	1357	318'535

Nach dem in Kapitel 2 beschriebenen Verfahren wurden nun 4-Wortgruppen aus den Daten gewonnen. Tabelle 3 zeigt die ersten 36 Multi Word Units, ohne dass die Daten irgendwie gefiltert worden wären.

Tabelle 3: Die 36 häufigsten Multi Word Units im Leserbrief-Korpus.

Multi Word Unit	absolute Frequenz	Frequenz/Mio. Token
1 Name der Redaktion bekannt	24	75
2 in den letzten Jahren	15	47
3 ist zu hoffen, dass	11	35
4 nicht in der Lage	7	22
5 ein Dorn im Auge	7	22

² Suche in <http://www.google.com> am 26. September 2006.

³ Netzeitung vom 16. September 2006: Linkspartei im Bund ist nicht regierungsfähig (<http://www.netzeitung.de/deutschland/landtagswahlen2006/interviews/440576.html>, gesehen am 26. September 2006). Hervorhebung von mir.

⁴ In der Folge abgekürzt mit TA.

6	an der Zeit, dass	7	22
7	rund um die Uhr	6	19
8	je länger je mehr	6	19
9	ist meiner Meinung nach	6	19
10	Auf der einen Seite	6	19
11	Auf der anderen Seite	6	19
12	Zur Kenntnis nehmen, dass	5	16
13	und schon gar nicht	5	16
14	ist nicht nur ein	5	16
15	in einer Zeit, in	5	16
16	in Tat und Wahrheit	5	16
17	handelt es sich um	5	16
18	ein Ende zu setzen	5	16
19	doch nicht sein, dass	5	16
20	das Vertrauen in die	5	16
21	auf den ersten Blick	5	16
22	auf Teufel komm raus	5	16
23	Wie wäre es wenn	5	16
24	Es ist eine Schande	5	16
25	Krieg im Nahen Osten	5	16
26	zwischen Israel und der	4	13
27	zur Rechenschaft gezogen werden	4	13
28	wie zum Beispiel die	4	13
29	von Kirche und Staat	4	13
30	von Bundesrätin Calmy-Rey	4	13
31	verletzen die religiösen Gefühle	4	13
32	unter die Lupe nehmen	4	13
33	unter den Augen der	4	13
34	stellt sich die Frage	4	13
35	sich nicht wundern wenn	4	13
36	liegt auf der Hand	4	13

Grundsätzlich erlaubt es die geschilderte Methode, Multi Word Units beliebiger Länge zu berechnen. Da es an dieser Stelle nur um eine exemplarische Studie geht, verzichten wir aus Platzgründen auf die Auflistung der Bi- und Trigramme⁵. Es würde sich nämlich zeigen, dass die Listen für unsere Zwecke bei dieser Textsorte wenig hilfreich sind. Ein zu großer Prozentsatz der Daten ist bei Bi- und Trigrammen zu mehrdeutig, als dass sie als Indikatoren für Argumentationsfiguren gelesen werden könnten. Anders bei den 4-Wortgruppen: Ohne komplexe statistische Maße anzuwenden, zeigt Tabelle 3 Multi Word Units, die uns als Leserbrief-typische Formeln erscheinen.

Wählt man jedoch Wortgruppen, die länger sind als vier Wörter, erhält man nicht mehr genügend hohe Frequenzen: Es gibt nur noch wenige Wortgruppen, die mehrmals im Korpus in gleicher Form auftauchen.

Ich kann an dieser Stelle diesen Aspekt nicht vertiefen, doch sei am Rande erwähnt, dass die Antwort auf die Frage, wie lange typischerweise die Wortgruppen sein müssen, um für ein bestimmtes

⁵ Also: Zwei- und Dreiwortgruppen.

Erkenntnisinteresse brauchbar zu sein, (auch) von der Textsorte abhängt. Da die Frequenzen der unterschiedlich langen Wortgruppen den Grad der Formelhaftigkeit der Texte ausdrücken, wird klar, dass Formelhaftigkeit textsortenspezifisch ist. Und mit der Auswertung der Frequenzen der unterschiedlich langen Wortgruppen haben wir ein statistisches Mittel in der Hand, um Formelhaftigkeit auszudrücken.

3.3 Von der Oberfläche in die Tiefe – und zurück: Interpretation

Doch stimmt die These, dass sich hinter vielen formelhaften Wortgruppen auf der Textoberfläche Argumentationsfiguren verbergen? Dass es also möglich sein müsste, formelhafte Wortgruppen als Indikatoren für Argumentationsfiguren zu lesen? Nachdem in einem ersten Schritt mit einem corpus-driven-Verfahren die Multi Word Units gewonnen wurden, folgt nun der zweite Schritt, bei dem wieder eine corpus-based-Perspektive eingenommen werden muss. Denn es muss nun geprüft werden, wie die gewonnenen Multi Word Units kontextualisiert sind und ob sich dahinter tatsächlich abstraktere Argumentationsfiguren verbergen. Wieder exemplarisch betrachten wir uns einige Einheiten detaillierter.⁶

3.3.1 ein Dorn im Auge

Dabei handelt es sich um einen metaphorischen Phraseologismus, der als Teil einer Argumentationsfigur kontextualisiert werden könnte. Allerdings sind mindestens zwei leicht unterschiedliche Argumentationsfiguren denkbar, was sich auch in den prototypischen Belegen zeigt:

Hickhack um Beschwerde, TA vom 31. 1.
Bautätigkeit wird nicht verhindert.
Seit langem sind den rechtsgerichteten Kreisen das Verbandsbeschwerderecht sowie die Lex Koller (Limitierung von Wohneigentum durch Ausländer) **ein Dorn im Auge**. Man scheut sich dabei nicht, mit Unwahrheiten die breite Masse zu manipulieren. Nur gerade ein Prozent aller Verwaltungsgerichtsbeschwerden stammt von Umweltschutzorganisationen. Die Behauptung, die Verbände würden die Bautätigkeit behindern, ist damit klar widerlegt. (1. 3. 2006)⁷

Die Aussage kann so generalisiert werden: A (die rechtsgerichteten Kreise) stört sich an Gegenstand x (Verbandsbeschwerderecht und Lex Koller). Für den Verfasser/die Verfasserin B ist x jedoch positiv konnotiert.

Doch es lässt sich eine weitere Nuance der Argumentation vermuten: Wenn B von A behauptet, x sei ihm «ein Dorn im Auge», wird A auch eine unlautere Argumentation unterstellt («fehlende Wahrhaftigkeit», Eggler 2006: 45f.), die eine pauschale Abneigung

⁶ Ich danke Marcel Eggler für seine argumentationstheoretische Hilfe und kritischen Anregungen besonders in diesem Kapitel.

⁷ Für alle Belege gilt: Hervorhebungen durch Fettschrift von mir. Für alle Datumsangaben gilt: Erschienen zum genannten Datum im Tages-Anzeiger, Ressort Leserforum.

gegenüber x verschleiert. Im Beleg oben könnte die Argumentation von B so ausformuliert werden: «Die rechtsgerichteten Kreise wollen (wahrscheinlich aus neoliberalen Gründen) die hinderlichen Gesetze abschaffen. Sie machen das mit einer Argumentation, die vordergründig plausibel ist und nur Details dieser Gesetze betreffen. Bei genauerer Betrachtung sind diese Argumente aber unlauter; eine pauschale Opposition aus ideologischen Gründen.»

In fünf von sechs Belegen des Leserbrief-Korpus wird vom *Gegner* behauptet, ihm sei etwas ein Dorn im Auge. In einem Fall jedoch ist das anders:

Grundsätzlich ist mir unsere (nach wie vor immer noch zu liberale, weil allzu oft missbrauchte) Asylpolitik **ein Dorn im Auge**. Der Dok-Film [ein Dokumentarfilm, NB] aber ging mir tief unter die Haut und gab mir sehr zu denken. (4. 3. 2006)

Ungewöhnlich ist hier, dass Gegenstand x für den Verfasser/die Verfasserin B selbst als störend empfunden wird; damit ist x nicht mehr positiv konnotiert und A und B fallen zusammen.

Die beiden Analysen zeigen, dass «ein Dorn im Auge» in argumentativer Funktion eine weitreichendere Bedeutung hat, als es z.B. der Duden-Band der Redewendungen (Duden 2002: «Dorn») beschreibt. Dort wird die Bedeutung dieser Formel mit «jmdm. ein Ärgernis, unerträglich sein» angegeben. Es ist wahrscheinlich, dass die oben skizzierte Kontextualisierung als komplexe Argumentationsfigur typisch für Leserbriefe ist und in anderen Textsorten nicht in dieser Weise verwendet wird.

3.3.2 je länger, je mehr ...

Als Mutter und Hundefan sehe und lese ich tagtäglich besorgt, dass der Mensch **je länger, je mehr** jeglichen Bezug zur Realität verliert. (4. 2. 2006)

Die Religion kann sie [die spirituellen und ethischen Werte, NB] scheinbar **je länger, je mehr** nicht mehr vermitteln. (17. 2. 2006)

Mit dieser Multi Word Unit wird in den Belegen eine negative Entwicklung behauptet. Die Semantik des Ausdrucks würde es zweifellos zulassen, ihn auch im Zusammenhang mit positiven Entwicklungen zu verwenden. Doch lässt sich vermuten, dass bei Leserbriefen typischerweise genau das nicht passiert und die Wortgruppe «je länger, je mehr ...» in den meisten Fällen als Indikator für eine Argumentationsfigur gelesen werden kann, die behauptet, etwas sei eine negative Entwicklung («Laudatio temporis acti»).

Diese Vermutung deckt sich mit einer Korpusrecherche in Leserbriefen der Neuen Zürcher Zeitung: In der Zeit vom 3. Januar 1993 bis zum 8. Oktober 2006 finden sich 20 Belege für «je länger, je mehr ...», die eine Entwicklung negativ einschätzen, 7 Belege mit unklarer/neutraler, 2 mit klar positiver Einschätzung. Die Zahlen zeigen

aber auch, dass es sich um eine sehr niedrigfrequente Wortgruppe handelt.

3.3.3 Auf der einen Seite/Auf der anderen Seite

Auf der einen Seite die Überheblichkeit derjenigen, die sich im Recht wähnen, in einem richtigen Amoklauf ein ganzes Land in Schutt und Asche zu legen, **auf der anderen Seite** die verzweifelte Gegenwehr derjenigen, die wissen, dass für ihre berechtigten Anliegen keine gerechte Lösung in Sicht ist. (9. 8. 2006)

Diese Formel wird in allen Belegen nur paarig verwendet. Sie folgt dem Muster: Es gibt x , es gibt (aber auch) y . Dabei scheint y von der Sprecherin B in der Mehrzahl der Belege als von A unterschätzt, vernachlässigt oder benachteiligt eingeschätzt zu werden.

3.3.4 und schon gar nicht

Weil sich heute eben gar niemand um die Leute ab 50 **und schon gar nicht** mehr ab 55 kümmert! (5. 1. 2006)

Einmal mehr wird das Stereotyp bemüht, wonach in der Schweiz die Medikamente zu teuer sind. Ob die präsentierten Daten aktuell und somit relevant sind, scheint weder Santésuisse noch den Tages-Anzeiger zu kümmern. Wer die beliebte Botschaft zu teuer verkündet, wird nicht hinterfragt **und schon gar nicht** kritisiert. Ich wünschte mir kritischere Journalisten und ausgewogenere Artikel [...]. (19. 6. 2006)

Sachverhalt x ist nicht der Fall. Sachverhalt y ist noch viel weniger der Fall («Argumentum a fortiori»). Evtl. mit dem Zusatz: Es wird aber oft behauptet, dass x und y der Fall seien.

3.3.5 in Tat und Wahrheit

Die momentanen Lohnerhöhungen sind **in Tat und Wahrheit** dazu da, Missstände zu beheben, die zu Lohnklagen hätten führen können. In Wirklichkeit sind die Einstiegs-, aber auch die Endlöhne jetzt tiefer als früher. (25. 1. 2006)

Das Unternehmen ist ein Staatsbetrieb wie im ehemaligen Osten. Der [Betriebschef, NB] soll seine Zügel straff in den Händen halten. **In Tat und Wahrheit** hat der Betriebschef aber seinen Job nicht gemacht! (11. 3. 2006)

Diese Multi Word Unit wird in Argumentationsfiguren verwendet, bei denen ein nach Auffassung des Sprechers (evtl. absichtlich) vernachlässigtes Faktum betont oder dem Gegner ein unehrliches Handeln unterstellt wird: Es scheint x , eigentlich ist aber y .

Im Kontext der untersuchten Leserbriefe wird «in Tat und Wahrheit» also nicht bloß mit der Bedeutung «in Wirklichkeit» verwendet, wie dies der Duden beschreibt (Duden 2002: «Tab»). Darüber hinaus handelt es sich um einen Sprechakt der Unterstellung unlauteren

Handelns, also der Unterstellung fehlender Wahrhaftigkeit (Eggler 2006: 45f).

3.3.6 auf Teufel komm raus

Denn der Konzern muss wachsen ..., schreibt der Tagi. Was soll die Schnapsidee vom ewigen Wachstum **auf Teufel komm raus**? Auch ein kleines Unternehmen kann sehr gut existieren, wenn es ein gutes Produkt, eine gute Leistung zu günstigem Preis bei kulantem Verhalten zu bieten hat. (26. 1. 2006)

Dieser idiomatische Phraseologismus wird im Sinne verwendet: x ist starrsinniges Verfolgen einer Idee. Die Durchsetzung dieser Idee lohnt sich aber nicht oder ist gar unangebracht. (In der Schweiz nennt man das «Zwängerei».) Auch hier unterscheidet sich die Verwendung in Leserbriefen wieder von der neutraleren Semantik, wie sie der Duden beschreibt: «mit allen Kräften, rückhaltlos», als «letztes, verzweifelteres Mittel» (Duden 2002, «Teufel»). In den untersuchten Leserbriefen wird immer der *andere* der «Zwängerei» bezichtigt und damit dessen Handeln als *falsch* eingeschätzt. Eine Aussage, in der ein Verfasser/eine Verfasserin von *sich selber* sagt, er/sie handle «auf Teufel komm raus» – und handle damit *richtig* –, findet sich nicht in den Belegen.

3.3.7 wie wäre es, wenn

Natürlich gibt es Schmarotzer. Aber **wie wäre es, wenn** sich Parlament und Bundesrat einmal um die wirklich grossen Schmarotzer kümmern würden, den Tausenden, die seit Jahrzehnten in Saus und Braus leben, hoch angesehen sind und dem Staat, wenn überhaupt, nicht viel mehr als die Personalsteuer entrichten? (24. 3. 2006)

Wie wäre es, wenn die internationale Gemeinschaft und die Uno eine internationale Streitmacht an den Grenzen Israels aufstellen würden, um die Palästinenser und die Libanesen vor der extremistisch zionistischen IDF [Israeli Defence Force, NB] zu schützen? (10. 8. 2006)

Hinter der Verwendung dieser Multi Word Unit verbirgt sich folgende Argumentationsfigur: A macht x . Es wäre besser/angebracht/konsequent, auch oder stattdessen y zu tun. Der Sprecher B schätzt es aber als hypothetisch ein, dass A das machen wird. Auch hier: Die Semantik dieses Ausdrucks ist eigentlich freier (z.B.: «Wie wäre es, wenn wir heute Abend ins Kino gingen?»). In Leserbriefen scheint damit aber eine Argumentation verbunden zu sein, die auf eine detailliertere Semantik des Ausdrucks schließen lässt.

3.3.8 Name der Redaktion bekannt

Diese Wortgruppe ist zunächst einmal spezifisch für die Textsorte der Leserbriefe. Sie erlaubt es Verfasserinnen und Verfassern gegenüber der Öffentlichkeit anonym zu bleiben, wobei die Redaktion trotzdem Authentizität behauptet.

Die Belege im Korpus lassen folgende Gründe für den Wunsch nach Anonymität vermuten:

- Verbotene oder heikle Nennung von Interna⁸
- Vertreten einer Minderheitsmeinung oder einer politisch unkorrekten Meinung⁹
- Anklage¹⁰
- Sich wehren gegen Anschuldigungen¹¹
- Explizites Äußern von Scham¹²

Die Verfasserinnen und Verfasser der Leserbriefe scheinen bei einer Namensnennung Konsequenzen zu fürchten, weil sie z.B. glauben, eine politisch unkorrekte Meinung zu vertreten, sich schämen oder jemanden angreifen. Allerdings lassen sich aus diesem Modus des Leserbriefschreibens auch weitergehende Facetten der Argumentation ableiten: Die Tatsache, dass eine Kritik anonym vorgebracht wird, kann als weiteres Argument dafür gelesen werden, wie berechtigt die Kritik ist. «Wenn es einem Assistenzarzt», könnte geschlossen werden, «nicht einmal möglich ist, seine Kritik offen gegenüber seinen Vorgesetzten zu formulieren, müssen die Zustände tatsächlich besonders schlimm sein!».

Oder aber, wenn eine Minderheitsmeinung oder eine politisch unkorrekte Meinung vertreten wird, kann der Verfasser/die Verfasserin versuchen, für sich Sympathien zu gewinnen und gleichzeitig den Gegner in Misskredit zu bringen: «Ich vertrete eine Meinung, die man ja nicht offen sagen darf. Das beweist, wie ungerecht die herrschende Schicht ist. Und das wiederum zeigt, dass ihre Meinung falsch sein muss!».

Der folgende Leserbrief mag das illustrieren:

Wie lange brauchen unser Staat und unsere Gesellschaft, um zu begreifen, dass es sich beim Kampf gegen die «bösen Drogen(konsumenten)» um ein aussichtsloses Unterfangen handelt? Der Wille, sich zu berauschen, ist im Gehirn des Menschen verankert und mit keiner Form von Repression zu ändern. Nur die legale Droge Alkohol darf für gute Stimmung an Partys sorgen, denn damit verdienen die Wirtschaft und unser Staat viel Geld. E. Sch. **(Name der Redaktion bekannt)** (10. 12. 2006)

⁸ Typische Beispiele: Arbeit der Assistenzärzte (11. 8. 2006); fragwürdige Behandlung auf Sozial- (8. 7. 2006) oder Arbeitsamt (20. 6. 2006); fragwürdige Zustände im Altersheim etc.

⁹ Typische Beispiele: Israel verteidigt sich zu recht gegen Hizbollah-Angriffe (18. 7. 2006, 4. 7. 2006); Serbin möchte ihr Volk nicht nur als Täter, sondern auch als Opfer verstanden wissen (17. 3. 2006) etc.

¹⁰ Typische Beispiele: Unterstellung schlechter Geschäftsführung (5. 5. 2006); Kritik hoher Gewinne eines Pharmaunternehmens (6. 2. 2006) etc.

¹¹ Typisches Beispiel: «Längst nicht alle [Heimkinder] sind kriminell und süchtig» (5. 5. 2006).

¹² Typisches Beispiel: Scham, schon lange arbeitslos zu sein (6. 1. 2006).

Der Verfasser/die Verfasserin behauptet, Drogenkonsum liege in der Natur des Menschen und könne deshalb nicht verhindert werden. Trotzdem werde das jedoch versucht, außer bei Alkohol. Und dies – die zweite Behauptung – liege an den Interessen von Wirtschaft und Staat. Der Satz «Name der Redaktion bekannt» suggeriert, eine solche Meinung könne man nur anonym äußern, es sei also die «nicht gern gehörte Wahrheit», die von der «Wirtschaft» und dem «Staat» vertuscht wird. Gleichzeitig wird aber der «Wirtschaft» und dem «Staat» fehlende Wahrhaftigkeit unterstellt (Eggler 2006: 45f.) und damit präventiv jegliche mögliche Gegenargumentation demontiert. Überspitzt: «Dass die Wahrheit [es geht bei der Zulassung von Alkohol nur um den Profit] anonym geäußert werden muss, zeigt: Der Gegner arbeitet mit unfairen Mitteln. Man riskiert Repressalien, wenn man die Wahrheit sagt.»

Man könnte also dem anonymen Verfasser/der anonymen Verfasserin eines Leserbriefs unterstellen, dass er/sie die Bitte um Anonymität, die sich letztlich durch das redaktionell gesetzte «Name der Redaktion bekannt» am Ende des Textes manifestiert, als rhetorisches Mittel verwendet. Dieses rhetorische Mittel beeinflusst die Argumentationskette im Text: Der Satz «Name der Redaktion bekannt» wird eine Aussage, die eine ganze argumentative Sequenz (im Leserbrief) untermauert.

Ausgehend von einem Argumentationsschema nach Toulmin (1958) kann man eine solche Argumentation über eine Argumentation als *A-Argument* bezeichnen (vgl. Eggler 2006 bzw. Grewendorf 1975). Das A-Argument («Eine Argumentation ist ganz besonders plausibel, da sie anonym geäußert werden muss.») wirkt somit mindestens als verstärkendes «Argumentum a fortiori» («Erst-recht-Argument», Eggler 2006: 60f.), maximal aber ergänzt es die offensichtliche Argumentation um versteckte Facetten. Die Abbildung 1 zeigt eine generalisierte Form dieser Argumentationsfigur in Toulmin'scher Darstellung.

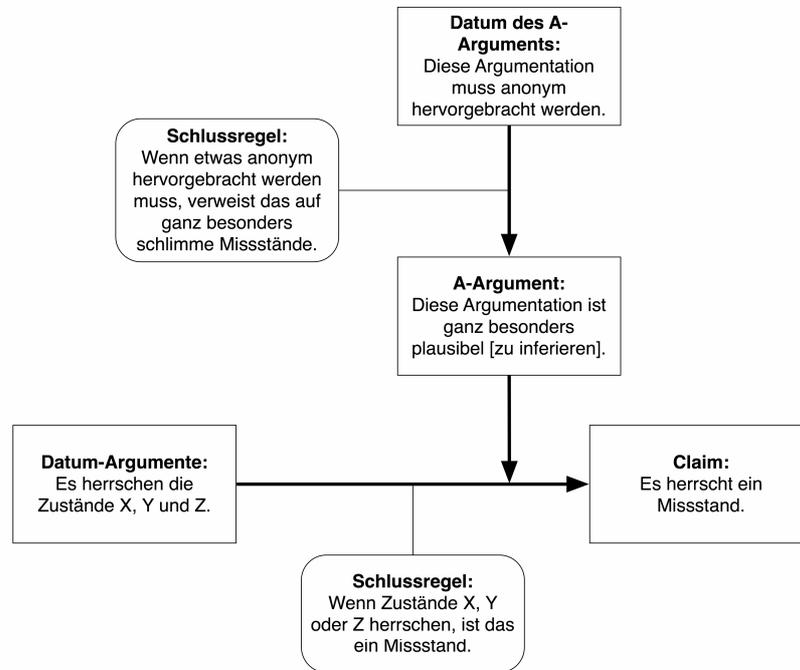
Es muss betont werden, dass es nur *eine* mögliche Interpretation ist, «Name der Redaktion bekannt» als A-Argument zu lesen. Und was den Verfasser/die Verfasserin *tatsächlich* dazu bewog, sich anonym zu geben, können wir nicht wissen.

4 Fazit und Ausblick

Bei der Analyse der Multi Word Units in Leserbriefen hat sich gezeigt, dass viele davon Teil relativ komplexer Argumentationsfiguren sind. Es ist also möglich, mit einem statistischen Verfahren, das «semantisch blind» ist, mit hoher Trefferquote Multi Word Units zu ermitteln, die Indikatoren für teilweise komplexe Argumentationsfiguren sind.

Für eine genauere Analyse könnte nun die Kategorisierung weitergetrieben werden in Richtung noch formalisierterer und abstrakterer Argumentationsfiguren, wie wir das ansatzweise bei «Name der Redaktion bekannt» gezeigt haben. Weiter müssten diese Argumentationsfiguren mit bestehenden Sammlungen rhetorischer Stilmitteln abgeglichen werden. Dieses Abgleichen der induktiv

Abbildung 1: Die erweiterte Argumentationsfigur eines anonym publizierten Leserbriefs: Aus «Name der Redaktion bekannt» kann die Leserin ein A-Argument inferieren, das die grundsätzliche Argumentation noch verstärkt.



gewonnenen Kategorien mit deduktiv gesetzten ist aus zweierlei Gründen wichtig:

1. Die statistische Analyse des Korpus bezüglich Multi Word Units ist nur *eine* mögliche Methode, um Indikatoren für Argumentationsfiguren zu gewinnen. Es fallen dabei zwangsweise viele weitere Indikatoren durch die Maschen.
2. Es gibt wahrscheinlich Argumentationsfiguren, die auf der Textoberfläche überhaupt schwer fassbar sind und zu denen keine eigentlichen (und relativ eindeutigen) Indikatoren gefunden werden können.

Auf der anderen Seite führt die corpus-driven-Methode, die induktiv vorgeht, zu Kategorisierungen, die für ein bestimmtes Korpus vielleicht überraschend sind. Überraschend bezüglich ihrer Existenz überhaupt, aber auch hinsichtlich der Frequenz: Die Statistik gewichtet anders, als der deduktiv denkende Mensch es tut.

Von den oben detailliert präsentierten Multi Word Units werden nur wenige als Einheiten in Wörterbüchern präsentiert. So finden sich im Duden (2002) nur die Phraseologismen «ein Dorn im Auge», «in Tat und Wahrheit» und «auf Teufel komm raus». Ihre argumentative – und damit pragmatische – Funktion in Leserbriefen verschafft ihnen zudem eine weit komplexere Bedeutung, als sie der Duden beschreibt. Darüber

hinaus gewinnen wir durch das induktive Verfahren Multi Word Units wie «Name der Redaktion bekannt» oder «wie wäre es, wenn», die nicht unter die Duden-Definition von Redewendung fallen.

Tabelle 4: Heuristiken für einen korpuslinguistischen Zugang

	corpus-driven	corpus-based
<i>Multi Word Units</i>	berechnen	Verteilung prüfen
<i>Ähnlichkeit und Differenz</i>	berechnen (Basis: Multi Word Units)	setzen und charakterisieren
<i>Konzepte (Argumentationsfiguren, Topoi etc.)</i>	ableiten (Basis: Multi Word Units)	Verteilung prüfen

Besonders zielführend ist deshalb ein Vorgehen, das im Wechsel zwischen corpus-driven- und corpus-based-Perspektiven liegt. Tabelle 4 stellt die möglichen Schritte dar: Berechnete Multi Word Units auf der Basis eines bestimmten Korpus bilden den Beginn der Recherche. Im Detail können ihre Verteilung und ihre unterschiedlichen Kontexte analysiert werden. Sie bilden aber wiederum die Basis, um z.B. Vergleiche mit anderen Korpora oder einem Teilkorpus zu erstellen: Welche Multi Word Units in Leserbriefen kommen auch in Zeitungsartikeln des Ressorts Ausland vor? Welche nicht?

Die Multi Word Units – ggf. im Kontrast zu einem anderen Korpus gefiltert – bilden ebenfalls die Basis für die Ableitung von Konzepten beliebiger Art, z.B. Argumentationsfiguren oder Topoi. Diese Kategorisierung erfolgt wiederum mit Rückgriff auf Korpusrecherchen.

Die kleine korpuslinguistische Untersuchung an Leserbriefen soll aber auch demonstrieren, dass diese Methode ebenso für ganz andere Erkenntnisinteressen eingesetzt werden kann. So können z.B. textlinguistische Fragestellungen zu Spezifika von Textsorten verfolgt werden, oder die Analysen können die Basis für diskurslinguistische Fragestellungen sein (Bubenhofer [in Vorbereitung]). Hier spielen auch die Argumentationsfiguren wieder eine interessante Rolle: Mit welchen Inhalten werden sie wann und von wem verknüpft? Welches sind die diskursiven «Evergreens», die uns immer wieder in alten Formeln frische Inhalte verkaufen?

Um diese Fragestellungen empirisch angehen zu können, wären Operationalisierungen komplexer Konzepte wie *Argumentationsfigur*, *Topos*, *Einstellung*, *Mentalität* etc. nötig. Das ist schwierig und gefährlich: Denn oft findet man in den Daten das, wonach man suchte, und sieht somit zu leicht vorher festgelegte Kategorien bestätigt.

Statistische Verfahren ermöglichen ein Arbeiten, das corpus-driven ist. Mit vergleichsweise einfach operationalisierbaren phraseologischen Kategorien wie *Kollokation* oder *typisiertem Sprachgebrauch* leiten wir aus den Daten induktiv komplexe Kategorien ab. Wir finden damit das Abstrakte, schwer Fassbare hinter dem Text dort, wo es letztlich sein muss: Auf der Textoberfläche.

5. Literaturverzeichnis

- BUBENHOFER, NOAH (in Vorbereitung): «Diskurse berechnen? Wege zu einer computerunterstützten Diskursanalyse.» In: *Methoden der Diskurslinguistik nach Foucault*.
- BURGER, HARALD (1998): *Phraseologie. Eine Einführung am Beispiel des Deutschen*. Schmidt (Grundlagen der Germanistik; 36).
- DUDEN (2002): *Redewendungen. Wörterbuch der deutschen Idiomatik*. Mannheim, Leipzig, Wien, Zürich: Dudenverlag, 2., neu bearbeitete und aktualisierte Aufl.
- EGGLER, MARCEL (2006): *Argumentationsanalyse textlinguistisch. Argumentative Figuren für und wider den Golfkrieg von 1991*. Tübingen: Niemeyer (Reihe Germanistische Linguistik; 268).
- FOUCAULT, MICHEL (2000): *Die Ordnung des Diskurses*. Fischer TB Verlag.
- GREWENDORF, GÜNTHER (1975): *Argumentation und Interpretation. Wissenschaftstheoretische Untersuchungen am Beispiel germanistischer Lyrikinterpretationen*. Kronenberg (Taunus): Scriptor (Wissenschaftstheorie und Grundlagenforschung; 2).
- KIENPOINTNER, MANFRED (1992): *Alltagslogik. Struktur und Funktion von Argumentationsmustern*. Stuttgart-Bad Cannstatt: Frommann-Holzboog (Problemata; 126).
- KOPPERSCHMIDT, JOSEF (1989): *Methodik der Argumentationsanalyse*. Stuttgart-Bad Cannstatt: Frommann-Holzboog (Problemata; 119).
- MANNING, CHRISTOPHER D./SCHÜTZE, HINRICH (2000): *Foundations of Statistical Natural Language Processing*. Cambridge, Massachusetts: The MIT Press, fifth printing 2002.
- SPITZMÜLLER, JÜRGEN (2005): *Metasprachdiskurse. Einstellungen zu Anglizismen und ihre wissenschaftliche Rezeption*. Berlin: De Gruyter (Linguistik – Impulse & Tendenzen; 11).
- STUBBS, MICHAEL (2003): «Two quantitative methods of studying phraseology in English.» In: *International Journal of Corpus Linguistics* 7: 2, S. 215–244.
- TEUBERT, WOLFGANG (2005): «My version of corpus linguistics.» In: *International Journal of Corpus Linguistics* 10: 1, S. 1–13.
- TOULMIN, STEPHEN (1958): *The Uses of Argument*. Cambridge University Press.
- WENGELER, MARTIN (2003): *Topos und Diskurs: Begründung einer argumentationsanalytischen Methode und ihre Anwendung auf den Migrationsdiskurs (1960 – 1985)*. Tübingen: Niemeyer (Reihe Germanistische Linguistik; 244).

Kontakt

Noah Bubenhofer, noah.bubenhofer@access.unizh.ch
www.bubenhofer.com
 Deutsches Seminar, Schönberggasse 9, CH-8001 Zürich