*Noah Bubenhofer/Joachim Scharloth*

# Kontext korpuslinguistisch

## Die induktive Berechnung von Sprachgebrauchsmustern in großen Textkorpora

Kontexte werden mit Mitteln der sprachlichen Oberfläche konstituiert. Zu diesen Mitteln gehören sprachliche Formeln, die für bestimmte Textsorten oder Textfunktionen typisch sind. Weil diese sprachlichen Formeln ein Phänomen der sprachlichen Oberfläche sind, auf die unmittelbar zugegriffen werden kann, ist eine empirische, korpuslinguistische Analyse dieser Formeln möglich. Mehr: Mit Mitteln der induktiven Statistik lässt sich *berechnen,* welche Elemente der sprachlichen Oberfläche für ganz bestimmte Kontexte typisch sind.

### 1 Sprachgebrauchsmuster als Effekte sozialen Sprachhandelns

Kontext als Bedingung für Sprechen anzunehmen, ist eines der Grundpostulate der sog. „pragmatischen Wende". Allerdings hat sich mit dieser Wende die Blockbildung in der Linguistik - Block 1: die sog. „Systemlinguistik", Block 2: die sog. „Bindestrichlinguistik" - verschärft, und in einem zweiten Schritt wurde die Pragmatik als Kind der Wende im systemlinguistischen Sinn umgedeutet und vereinnahmt: „Austin, Wittgenstein, and Grice were hailed as heroes in the 1970s and their insights were quickly integrated into a systemoriented linguistics looking for universal features of language." (Nerlich 1995: 311) Damit führte die pragmatische Wende zwar zu einer Ausweitung des linguistischen Untersuchungsgegenstandes: Es wurden nun auch Alltagstexte und Gespräche betrachtet. Doch wurde nun parallel zum universalgrammatischen ein universalpragmatisches Erkenntnisinteresse formuliert, das sich auf die Suche nach der Universalität von Sprechakten und deren tiefenstrukturellen Gemeinsamkeiten machte. Feilke (2003: 217) sieht darin aber nur den ersten Teil einer pragmatischen Wende, die sich in den vergangenen 20 Jahren durch eine „zweite pragmatische Wende" unter geänderten Vorzeichen fortsetzte.

Diese zweite Wende setzt ihre Positionen in einigen Bereichen neu. (Feilke 2003: 217 ff.) Besonders relevant für eine neue Sichtweise auf Kontext sind die folgenden beiden Aspekte:

- Formelhaftigkeit der sprachlichen Oberfläche statt sprachliche Universalien der Tiefenstruktur

- Kontextualisierung statt Kontext

Die Fähigkeit, den pragmatischen Wert eines Textes bestimmen zu können, liegt weniger darin, Tiefenstrukturen zu entschlüsseln, sondern über ein Sprach*gebrauchs*wissen zu verfügen. Überspitzt formuliert: Die Art der Formulierung entscheidet darüber, ob es sich bei einem Text um einen Wetterbericht oder eine politische Ansprache handelt.

Vgl. dazu die beiden Gruppen von Formeln:

1. „... Durchzug von ... weitgehend trocken ... vereinzelt ..."[1]
2. „... liebe Mitbürgerinnen und Mitbürger ... unser ist ... aber ... gleichzeitig nimmt ... stetig ab ... jeder ... unabhängig von ... ich will ... unser ... muss ... werden ... mehr Wettbewerb ... solidarisch ... ich bin sicher ..."[2]

Statt den Kontext als Grundlage für das Verstehen zu begreifen, wird Kontext auch als Folge von Text betrachtet: Anscheinend sind bestimmte Elemente in einem Text dafür verantwortlich, dass unterschiedliche LeserInnen und HörerInnen den gleichen Kontext erzeugen. Sprachhandeln wirkt also konstituierend für Entitäten wie Textsorten oder kommunikative Gattungen. Denn wie Feilke (2003: 212) klar zeigt, sind es z.B. bestimmte Formulierungen, wie jene oben im Wetterbericht, die genügen, um LeserInnen oder HörerInnen zu veranlassen, jeweils einen ähnlichen Kontext um die Formeln herum aufzubauen. Damit werden aber solche Formeln zu einem entscheidenden Element, die die Textsorte „Wetterbericht" konstituieren. Notwendig ist demnach eine Art „Sprachgebrauchskompetenz"[3] oder „Sprachgebrauchswissen", das Wissen um typisierten Sprachgebrauch, das den SprecherInnen und HörerInnen erlaubt, „to judge the appropriateness of given utterances in given contexts". (Fillmore 1976: 90)[4,5]

Idiomatische Prägungen sind das Resultat von Konventionalisierungen von Interpretationen, die ausdrucksseitig diese Konventionalisierung widerspiegeln. „So können über die Prägung bestimmter *Ausdrücke* Ressourcen

[1] Vgl. Feilke (2003: 211).

[2] Phrasen aus dem „Video-Podcast der Kanzlerin zur Gesundheitsreform" vom 24. Juni 2006 (Angela Merkel); vgl. http://www.bundeskanzlerin.de/ (26.2.2008).

[3] Mit „Kompetenz" klingt die Chomsky'sche Dichotomie „Kompetenz vs. Performanz" an. Die Sprachgebrauchskompetenz weist mit dieser Kompetenz die Gemeinsamkeit auf, dass sie von den SprecherInnen der Sprache unbewusst angeeignet wurde. Allerdings sind die Sprachgebrauchskompetenzen der Individuen nicht deshalb ähnlich, weil es ein dahinter stehendes „Sprachgebrauchssystem" gibt, sondern weil es sich durch das kooperative Sprachhandeln angeglichen hat. Alternativ zur heiklen Bezeichnung könnte „Sprachgebrauchswissen" verwendet werden.

[4] Deutsch übersetzt als „Welche *sprachliche* Kompetenz ist es [...], die es einem Sprecher ermöglicht, einer gegebenen sprachlichen Äußerung einen *bestimmten* Kontext des Handelns zuzuordnen? " (in Feilke 2000: 78)

[5] Fillmore (1976: 88) beschreibt dies als Zuordnung zu „set of worlds in which the discourse could play a role, together with the set of possible worlds compatible with the message content of the discourse".



des Vorverständigtseins für die Kommunikation geschaffen und gesichert werden." (Feilke 1993: 367) Diese „Typik des Gebrauchs" (Feilke 2003: 209) ermöglicht überhaupt erst pragmatisch brauchbare Sprache.

Die beiden Sammlungen von Formeln aus einem Wetterbericht und einer politischen Ansprache plausibilisieren diese Überlegungen zur Rolle der idiomatischen Prägungen, die, in semasiologischer Perspektive, als Sprachmaterial übereinstimmende Vorstellungen typischen sozialen Handelns ermöglichen. Denn bei den Textbeispielen wurden eben nicht beliebige Textfragmente verwendet, sondern syntagmatische Ausdrücke, die den „konnotativen Mehrwert" (Feilke 2003: 213) bieten, der sich durch die Konventionalisierung ergeben hat. Aus phraseologischer Perspektive bemerkenswert ist die Tatsache, dass es sich dabei nicht ausschließlich um Ausdrücke hoher Idiomatizität handelt, sondern hauptsächlich um Kollokationen oder Routineformeln (Burger 1998: 50-52) mit „gänzlich regulärer Struktur und kompositioneller Bedeutung", die aber „hochgradig konventionell und zeichenhaft" sind. (Feilke 2003: 213) So wird das breite Bedeutungsspektrum von *Durchzug* im ersten Textbeispiel durch die Erweiterung zu *Durchzug von* bereits enorm reduziert, hauptsächlich auf den Kontext Wetter.[6] Allerdings findet Feilke (2003: 214) im DeReKo IDS-Korpus bloß 11 Belege für *Durchzug von,* was zeigt, dass nicht absolute Frequenzen rekurrenten Auftretens eines Ausdrucks dessen Typik bestimmen, sondern zu Textsorten relative Frequenzen. Das ist - um vorzugreifen - eines der Kernprobleme korpuslinguistischer Analysen, das jedoch mit der Hilfe statistischer Methoden gut gelöst werden kann.

Die Kombination mehrerer idomatisch gepräger Ausdrücke verringert die möglichen Kontextualisierungen weiter, wobei unter solchen Ausdrücken nicht zwingend Mehrwortausdrücke zu verstehen sind, wie *vereinzelt* im Textbeispiel oben zeigt. An diesem Wort zeigt sich auch, dass eine bestimmte morphologische Ausprägung (also das Partizip Perfekt von *vereinzeln*) hinsichtlich der Textsorte „Wetterbericht" idiomatisch geprägt ist. Die Form *vereinzeln* wird, was eine Recherche im DeReKo IDS-Korpus zeigt, kaum verwendet.

Die Erkenntnis, dass letztlich Sprachgebrauchsmuster (oder eben: typische Ausdrücke, idiomatische Prägungen) an der Textoberfläche die Kontextualisierung ermöglichen, ist der Schlüssel für einen korpuslinguistischen, induktiven Zugang zu Sprachgebrauchsanalysen. Wenn die Sprachgebrauchsmuster erkannt werden können, sind Rückschlüsse auf ihre typischen Kontextualisierungsleistungen möglich.

[6] Neben einer Korpusrecherche, z.B. im DeReKo IDS-Korpus, wird dies auch mittels entsprechender Suchanfragen in einer Suchmaschine wie „Google" deutlich: Die Belegmenge von *Durchzug* wird durch *Durchzug von* um etwa den Faktor 30 verkleinert, wobei die Mehrzahl der Treffer im Kontext von Wetter und Militär *(Durchzug von Militär)* angesiedelt zu sein scheint (Anfrage auf www.google.com vom 6. Juli 2007).

Die Analyse dieser Kontextualisierungsleistungen ist die Grundlage für die Beantwortung weitergehender Untersuchungsfragen. Neben textlinguistischen Fragestellungen zur Typik von Textsorten und Textfunktionen gibt es interessante Anwendungsmöglichkeiten in der Diskurs- und Kulturanalyse, indem typische Sprachgebrauchsmuster nicht nur als Mittel der Kontextualisierung, sondern letztlich der Konstituierung von Diskursen und Kultur angesehen werden. Grundlegende Schritte der Methode liegen darin, die Distribution der Sprachgebrauchsmuster in einem Korpus, deren Veränderungen bezüglich Frequenzen, Form oder bezüglich deren Vorkommen mit anderen Mustern zu beobachten (vgl. dazu ausführlich Bubenhofer 2008, 2009).

Der erste Schritt besteht aber darin, mit korpuslinguistischen Methoden Sprachgebrauchsmuster zu finden. Dabei ist die Erkenntnis wichtig, dass solche Ausdrücke immer nur prägend sind in Bezug auf eine Textfunktion, z.B. eine Textsorte. Dies muss ein statistisches Modell zur Berechnung solcher Ausdrücke berücksichtigen. Ziel der Methode ist es, mit Hilfe der geprägten Ausdrücke den typischen Sprachgebrauch eines Sprachausschnittes zu beschreiben.

## 2 Corpus-driven

„Sind Korpora nur Belegsammlungen oder Zettelkästen in elektronischer Form? Mitnichten! In entsprechender Größe [...] und mit den entsprechenden Analysemethoden eröffnen sie eine eigene Perspektive in der linguistischen Forschung – die korpuslinguistische Perspektive." (Perkuhn/Belica 2006: 2)

Das Zitat verweist auf ein in der Linguistik noch immer häufiges Missverständnis von Korpuslinguistik. Natürlich können Korpora auch als Nachschlagewerke benutzt werden, um zu überprüfen, ob sich darin ein bestimmtes sprachliches Phänomen belegt findet. So kann beispielsweise geprüft werden, ob ein sprachliches Phänomen, von dem vermutet wird, dass es typisch für einen bestimmten Kontext ist, in einem entsprechenden Korpus auch signifikant häufiger vorkommt. Doch gerade mit Hilfe der immer schneller werdenden Computer und der Verfügbarkeit von elektronischen Korpora ist für viele Fragestellungen eine andere Perspektive in der Korpuslinguistik interessanter: die „corpus-driven"-Perspektive:

> 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)

Dieser Zugang zeichnet sich also im Versuch aus, das Korpus als Datenbestand aufzufassen, in dem mit geeigneten Methoden Strukturen sichtbar gemacht werden, die erst im Nachhinein klassifiziert werden.



Damit steht dieser Zugang in Kontrast zur „corpus-based"-Perspektive, bei der mit bestimmten Kategorien und bestehenden Theorien das Korpus analysiert wird und das Interesse dann darin besteht, eine Hypothese zu testen. Letztlich wird corpus-based die Frage verfolgt, ob das gesuchte Phänomen im Korpus auftritt, wenn ja, wo, wie oft und wie:

> [...] the term corpus-based is used to refer to a methodology that avails itself of the corpus mainly to expound, test or exemplify theories and descriptions that were formulated before large corpora became available to inform language study. (Tognini-Bonelli 2001: 65)

Doch Korpora mit ganz bestimmten Theorien als Prämissen zu befragen, birgt die Gefahr, in den Daten nur die Strukturen zu finden, die mit der Theorie kompatibel sind und blind gegenüber Evidenzen zu sein, die quer zu einer Theorie stehen:

> [...] corpus-based linguists adopt a „confident" stand with respect to the relationship between theory and data in that they bring with them models of language and descriptions which they believe to be fundamentally adequate, they perceive and analyse the corpus through these categories and sieve the data accordingly. (Tognini-Bonelli 2001: 66)

So sind im corpus-based-Verfahren Korpora nützlich, um aufgrund der Daten die Hypothesen zu korrigieren und zu quantifizieren. Aber:

> In this case, however, corpus evidence is brought in as an extra bonus rather than as a determining factor with respect to the analysis, which is still carried out according to pre-existing categories; although it is used to refine such categories, it is never really in a position to challenge them as there is no claim made that they arise directly from the data. (Tognini-Bonelli 2001: 66)

So ist es oft nahe liegend, Daten, die einem Modell widersprechen, als durch Sprachvarietät verursachte Ausnahmen zu erklären, die jedoch zu marginal sind, als dass sie das Modell angreifen könnten.[7] Die corpus-driven-Analyse bietet einen anderen Zugang, der die Daten zum Ausgangspunkt der Theoriebildung macht.[8]

---

7 Tognini-Bonelli (2001: 67) spitzt dieses Dilemma zur rhetorischen Frage zu: „A problem then arises, one that has been cropping up throughout the history of linguistics: given that the data is non-negotiable, does the linguist choose to revise the theory and derive it more directly from corpus evidence, or does (s)he opt to insulate the data from the theory?"

8 Das corpus-driven-Paradigma ist nicht neu. Bei Sinclair (1991) bereits angedacht, wird es bei Tognini-Bonelli (2001) mit dem Terminus „corpus-driven Linguistics" (CDL) explizit gemacht. Doch sind es vor allem im deutschsprachigen Raum auch heute erst wenige Linguistinnen und Linguisten, die konsequent so arbeiten. Dazu gehören z.B. Arbeiten von Cyril Belica und Rainer Perkuhn (Belica 1996, Belica 2001-2007; Perkuhn u.a. 2005; Perkuhn und Belica 2006) und Kathrin Steyer (Steyer 2004a, Steyer 2004b; Steyer/Lauer 2007).

Möchte man den typischen Sprachgebrauch in einem Untersuchungskorpus feststellen, ergeben sich aus diesen Überlegungen folgende Konsequenzen:

1. Die corpus-driven Perspektive sollte es ermöglichen, induktiv aus einem Untersuchungskorpus Sprachgebrauchsmuster hervorzubringen, die typisch für dieses Korpus sind.
2. Dabei bleibt bis zu einem gewissen Grad offen, welcher Form die Sprachgebrauchsmuster überhaupt sind. Die Methode ist blind gegenüber semantischen Kategorisierungen und bringt alle Wortkombinationen hervor, die statistisch auffällig sind. Es kann sich dabei um Kollokationen, Phraseologismen, idiomatische Ausdrücke oder komplexe Mehrworteinheiten handeln.

## 3 Typischer Sprachgebrauch

Typischer Sprachgebrauch kann operationalisiert werden als rekurrentes Auftreten von textuellen Einheiten in bestimmten Sprachausschnitten. Solche textuelle Einheiten können Morpheme sein, Lexeme oder aber komplexe Kombinationen von solchen Einheiten. Zudem können sie in Kombinationen mit „Slots" erscheinen, wie z.B. in *in der Nacht auf* [Wochentag]. „Typisch" ist dabei nicht einfach mit „häufig" gleichzusetzen, sondern meint immer „häufig in Bezug auf bestimmte Sprachausschnitte". Ein Sprachgebrauch ist dann typisch, wenn er in einer bestimmten Art von Kontexten, z.B. in Zeitungsartikeln aus dem Auslandsressort, überzufällig häufiger auftritt, als in Referenzkorpora. Wir schlagen vor, dieses Phänomen typischen Sprachgebrauchs „Muster" zu nennen. Von besonderem Interesse ist also das Moment der Gebräuchlichkeit, der Typik (Feilke 1996: 72 ff.) von sprachlichen Mustern. Und diese zeigt sich in der rekurrenten Struktur sprachlicher Einheiten, Einheiten die statistisch auffällig sind.

Es gibt Verfahren, die es auf einfache Weise ermöglichen, rekurrente Wortverbindungen zu berechnen (vgl. z.B. Banerjee/Pedersen 2003; Cheng u.a. 2006 sowie Bubenhofer 2008, 2009). In einem simplen Auszählverfahren werden die Frequenzen aller möglichen Mehrworteinheiten beliebiger Länge, also nicht nur von Wortpaaren, berechnet. Die Mehrworteinheiten können auch Lücken aufweisen und müssen nicht nur aus kontinuierlich aufeinander folgenden Wörtern bestehen. Zusätzlich können statistische Maße angewandt werden, um den Grad der Überzufälligkeit der Wortverbindung auszudrücken. Wichtig ist, dass diese Verfahren keine vorgängige Beschränkung auf ein zu untersuchendes Lemma benötigen, denn es werden grundsätzlich alle möglichen Mehrworteinheiten evaluiert.

Der Vorteil dieser Methode zeigt sich darin, dass sie weitgehend dem corpus-driven-Paradigma entspricht: Es ist der Versuch, Strukturen in einem Korpus sichtbar zu machen, ohne die Suche schon zu Beginn mit der Definition bestimmter Suchbegriffe einzuschränken. Mit diesem Versuch ist



die Hoffnung verbunden, auf Strukturen zu stoßen, an die man zu Untersuchungsbeginn gar nicht gedacht hatte.[9]

Der Nachteil kann in technischen Beschränkungen liegen: Der Preis für flexible Definitionen von Kollokationen (Diskontinuität, freie Reihenfolge, Berücksichtigung des gesamten Korpus) ist hoch, was die nötige Rechenleistung betrifft. Und die entstehenden Datenmengen sind sehr umfangreich, so dass für die weitere Verarbeitung Verfahren der Filterung oder Fokussierung nötig sind.

## 4 Typikprofile erstellen und analysieren

Die besondere Herausforderung bei einer corpus-driven-Analyse liegt darin, potenziell interessante von restlichen Daten zu scheiden. Denn mit den Berechnungen entstehen umfangreiche Daten. Der Datenumfang hängt von den Parametern ab, die für die Berechnung der Mehrworteinheiten gewählt wurden. Es würde der Philosophie des corpus-driven-Paradigmas jedoch widersprechen, wenn bereits dort besonders rigide Parameter gewählt würden. Deshalb muss der Schritt des Filterns der Daten möglichst spät erfolgen.

Um die sprachlichen Muster zu finden, die für ein Korpus typisch sind, werden die Listen von Mehrworteinheiten, die separat für das Untersuchungs- und das Referenzkorpus berechnet worden sind, verglichen. Dazu können statistische Methoden verwendet werden. Mit einem Signifikanztest wird geprüft, welche Mehrworteinheiten typisch für das Untersuchungskorpus sind. Dazu dient z.B. der Log-Likelihood-Test ($G_2$), der in einer Reihe von korpuslinguistischen Arbeiten für Korpusvergleiche verwendet wird (vgl. für Vergleiche mit anderen Verfahren und einer ausführlichen Diskussion Bubenhofer 2009).

Durch die Kontrastierung der Listen von Mehrworteinheiten für Untersuchungs- und Referenzkorpus – oder für verschiedene Teilkorpora generell – werden aus diesen Listen die für die jeweiligen Korpora typischen Mehrworteinheiten extrahiert. Diese Zusammenstellungen von für ein Teilkorpus typischen Mehrworteinheiten (im Kontrast zu einem oder mehreren anderen Teilkorpora) nennen wir „Typikprofil". Natürlich können weitere Kontrastierungen mit Typikprofilen anderer Korpora folgen. So kann es interessant sein, das Typikprofil von Texten eines Korpus A mit denen eines Korpus B zu vergleichen. Oder es kann für einen beliebigen Text berechnet

[9] Diese Hoffnung teilen z.B. auch Cheng u.a. (2006: 415), die mit der Software „Concgram" die flexibelste Form der Wortgruppenberechnung anbieten: „The fully automated capability of the search engine, i.e. the absence of any form of prior intervention by the user, makes it a truly ‚corpus-driven' methodology [...], and so further increases the likelihood that the concgram searches will enable the researcher to discover not only a more extensive description of patterns of collocation and their meanings, but also, and more importantly, new patterns of language use."

werden, wie stark er mit einem bestimmten Typikprofil übereinstimmt: Je mehr Mehrworteinheiten eines Typikprofils in einem Text vorkommen, desto eher entspricht der Text dem Typikprofil.

Je nach Korpusgröße kann das Typikprofil relativ umfangreiche Datenmengen enthalten. Das Untersuchungsinteresse entscheidet nun darüber, ob und wie diese Datenmenge eingeschränkt werden kann. So kann der Fokus auf spezifischen lexikalischen oder syntaktischen Mehrworteinheiten liegen oder die Mehrworteinheiten werden nach Kriterien der Textfunktion oder Pragmatik gruppiert und weiter analysiert.

Diese Analyseschritte markieren auch den Moment, bei dem das corpus-driven-Verfahren in eine corpus-based-Perspektive umschlägt. Wie in klassischen korpuslinguistischen Arbeiten stellen die berechneten Wortgruppen die Basis für weiterführende Korpusanalysen dar. Es werden die Verwendungsweisen der Wortgruppen im Untersuchungs- und Referenzkorpus untersucht oder Varianten geprüft.

Die Selektion und Klassifikation von Mehrworteinheiten für die weitere Analyse richtet sich nach unterschiedlichen Kriterien:

*Selektionskriterium 1: Statistische Typik*
1. Mehrworteinheit ist typisch für bestimmte Teilkorpora (z.B. das Untersuchungskorpus)
2. Mehrworteinheit ist über alle Teilkorpora ähnlich verteilt

*Selektionskriterium 2: Syntaktische Festigkeit*
1. Tendenziell syntaktisch fest
2. Tendenziell syntaktisch variabel

*Typologisierendes Kriterium 1: Kontinuität der Wortfolge*
1. Tendenziell diskontinuierliche Mehrworteinheiten
2. Tendenziell kontinuierliche Mehrworteinheiten

*Typologisierendes Kriterium 2: Phraseologischer Typus*
1. Referentielle Phraseologismen
2. Strukturelle Phraseologismen
3. Kommunikative Phraseologismen
4. Sonderformen

Das primäre Selektionskriterium ist die statistische Typik der Mehrworteinheit für ein bestimmtes Teilkorpus. Einerseits können Einheiten in den Blick genommen werden, die besonders typisch für eine Textsorte, einen Zeitabschnitt etc. sind, oder aber solche, die es gerade nicht sind. Ist man an grundsätzlichen sprachlichen Mustern interessiert, die unabhängig von Zeit und Textsorte vorkommen, würde man die Mehrworteinheiten auswählen, die über mehrere Korpora – Untersuchungs- und Referenzkorpora – ähnlich verteilt sind.

Als sekundäres Selektionskriterium gilt die syntaktische Festigkeit: Die Mehrworteinheiten im Typikprofil werden zunächst danach untersucht, ob



sie syntaktisch eher fest oder eher variabel sind. Eine Mehrworteinheit ist dann syntaktisch fest, wenn der größte Teil der Belege für die Mehrworteinheit demselben syntaktischen Muster folgt. Syntaktisch feste Mehrworteinheiten sind für die weiteren Analysen interessanter als variable, da sie auf Verwendungskontexte zurückgehen, die zumindest syntaktisch, aber wahrscheinlich auch semantisch, ähnlich sind.

Die Mehrworteinheiten können nach weiteren Kriterien klassifiziert werden. Nützlich ist beispielsweise eine Klassifikation, wie sie Burger (1998: 35 ff.) für Phraseologismen vorschlägt.

## 5 Anwendungsbeispiel

Im Folgenden soll in aller Kürze eine exemplarische Analyse von Sprachgebrauchsmustern zu *Kampf gegen …* präsentiert werden, die besonders unter dem Aspekt der unterschiedlichen Kontextualisierungsleistungen interessant sind. Die Analyse ist Teil einer umfangreicheren Arbeit (Bubenhofer 2009), bei der mit einem 27,9 Mio. Wörter umfassenden Zeitungskorpus gearbeitet wurde. Die folgende Analyse zeigt dabei auch Zusammenhänge zwischen Kontextualisierungsleistungen und Diskursen auf. Bestimmte Sprachgebrauchsmuster erzeugen Kontextualisierungen im Rahmen von Diskursen und konstituieren diese damit mit.[10]

### 5.1 Korpus

Das Untersuchungskorpus enthält Zeitungsartikel der Neuen Zürcher Zeitung (NZZ). Als Grundgesamtheit sind alle redaktionellen Artikel aus dem Zeitraum von 1995 bis einschließlich 2005 definiert. Daraus wurde eine Zufallsstichprobe von 44.843 Artikeln gezogen, die insgesamt einen Umfang von 27,9 Mio. Wörtern umfasst. Diese Stichprobe deckt damit knapp 6,5 % der 693.213 Artikel der Grundgesamtheit ab. In Tabelle 1 sind die detaillierten Angaben pro Jahr aufgeführt.

| Jahr | Grundges. | Stichprobe | | | Zeitspanne |
|---|---|---|---|---|---|
| | Artikel | Artikel | Wörter | % der GG | |
| 1995 | 56514 | 3666 | 2458651 | 6,49 | 3. Jan. bis 30. Dez. 1995 |
| 1996 | 57708 | 3711 | 2463270 | 6,43 | 3. Jan. bis 31. Dez. 1996 |
| 1997 | 57141 | 3827 | 2557410 | 6,7 | 3. Jan. bis 31. Dez. 1997 |
| 1998 | 54543 | 3616 | 2557811 | 6,63 | 3. Jan. bis 31. Dez. 1998 |

[10] Vgl. für eine ausführliche Diskussion zum Zusammenhang von Diskursen und Sprachgebrauchsmustern Bubenhofer (2009).

| | | | | | |
|---|---|---|---|---|---|
| 1999 | 66788 | 4248 | 2631274 | 6,36 | 4. Jan. bis 31. Dez. 1999 |
| 2000 | 69170 | 4496 | 2740517 | 6,5 | 1. Jan. bis 30. Dez. 2000 |
| 2001 | 68511 | 4373 | 2621799 | 6,38 | 3. Jan. bis 31. Dez. 2001 |
| 2002 | 68363 | 4385 | 2582855 | 6,41 | 3. Jan. bis 31. Dez. 2002 |
| 2003 | 65063 | 4235 | 2520628 | 6,51 | 3. Jan. bis 31. Dez. 2003 |
| 2004 | 64157 | 4155 | 2405419 | 6,48 | 3. Jan. bis 31. Dez. 2004 |
| 2005 | 65255 | 4131 | 2406747 | 6,33 | 3. Jan. bis 31. Dez. 2005 |
| **Total** | **693213** | **44843** | **27946381** | **6,47** | **1995 bis 2005** |

*Tabelle 1*
*Datengrundlage der Untersuchung: Artikel der Neuen Zürcher Zeitung zwischen 1995 und 2005 (Zufallsstichprobe, 6 Ausgaben/Woche)*

Um in den Korpusdaten die frequenten und/oder statistisch signifikanten Mehrworteinheiten zu berechnen, wurde die Software „Ngram Statistics Package" („NSP") verwendet (Banerjee/Pedersen 2003). Mit NSP wurden für verschiedene Teilkorpora die frequenten Mehrworteinheiten berechnet. Die Teilkorpora orientieren sich entlang der Kategorien „Zeit" und „Ressort". Für die vorliegende Untersuchung ist nur das Teilkorpus der Auslandsartikel relevant, das in verschiedene Zeitabschnitte aufgeteilt wurde, um Mehrworteinheiten berechnen zu können, die für bestimmte Epochen zwischen 1995 und 2005 typisch sind.

Für die Berechnung der Mehrworteinheiten müssen eine Reihe von Parametern gesetzt werden wie Länge der Mehrworteinheit oder maximale Spannweite zwischen erstem und letztem Wort der Mehrworteinheit. Als Standardverfahren haben sich folgende Parameter als gewinnbringend für die vorliegenden Analysezwecke erwiesen:

**Länge:** 3 Wörter
**Spannweite:** 10 Wörter (max. Distanz zwischen erstem und letztem Wort)
**Basis:** Wortformen (die Mehrworteinheiten *die Kämpfe gegen* und *der Kampf gegen* werden als zwei verschiedene Mehrworteinheiten gezählt)
**Mindestfrequenz:** 6 pro Mio. Wörter

Im Anschluss wurden die Listen der für ein Teilkorpus berechneten Mehrworteinheiten durch Kontrastverfahren untereinander verglichen, um die für die Teilkorpora typischen Mehrworteinheiten zu finden. Dafür diente ein eigenes Programm, das jeweils zwei Listen von Mehrworteinheiten wahlweise mittels der Statistiken „Log-Likelihood Koeffizient", „Chi-Quadrat-Test" oder „Mann-Whitney-Rank-Test" vergleicht.[11] Das Programm ermöglicht aber nicht nur den Vergleich zweier Listen miteinander, sondern den Vergleich einer mit mehreren Listen. So können die für ein Teilkorpus im Vergleich zu allen anderen Teilkorpora typischen Mehrworteinheiten eruiert werden.



## 5.2 Resultate

| | $G^2$ | MWE 1995–1997 | | $G^2$ | MWE 2003–2005 |
|---|---|---|---|---|---|
| 1 | 47,15 | der bosnischen Serben | 36 | -72,89 | im Irak die |
| 2 | 27,90 | die bosnischen Serben | 37 | -59,77 | und im Irak |
| 3 | 39,52 | und der Opposition | 38 | -40,82 | gegen den Irak |
| 4 | 35,56 | Aung San Suu | 39 | -27,7 | amerikanischen im Irak |
| 5 | 30,3 | San Suu Kyi | 40 | -52,48 | der den USA |
| 6 | 32,93 | der der Menschenrechte | 41 | -34,99 | der USA im |
| 7 | 31,71 | der in Bosnien | 42 | -48,11 | Abzug aus dem |
| 8 | 30,3 | der ersten Runde | 43 | -43,74 | aus dem Irak |
| 9 | 28,98 | in der Serben | 44 | -41,82 | gegen den Terrorismus |
| 10 | 27,66 | vor den Wahlen | 45 | -24,49 | Kampf gegen den |
| 11 | 27,66 | die immer wieder | 46 | -21,87 | Kampf gegen Terrorismus |
| 12 | 26,34 | die in gegen | 47 | -21,87 | Krieg gegen den |
| 13 | 26,34 | Der Präsident der | 48 | -37,91 | den neunziger Jahren |
| 14 | 26,34 | in der Slowakei | 49 | -35,28 | hiess es in |
| 15 | 26,34 | am Abend der | 50 | -33,53 | aus dem Gazastreifen |
| 16 | 25,36 | der Nato und | 51 | -30,62 | der in Kosovo |
| 17 | 25,03 | am in Paris | 52 | -29,16 | Präsident Bush in |
| 18 | 23,71 | die absolute Mehrheit | 53 | -26,24 | den besetzten Gebieten |
| 19 | 23,71 | russische Präsident Jelzin | 54 | -24,78 | auf in Bagdad |

[11] Vgl. für die Details dieser statistischen Verfahren Bubenhofer (2009).

| | | | | | |
|---|---|---|---|---|---|
| 20 | 22,39 | der in Sarajewo | 55 | -24,78 | die Kontrolle über |
| 21 | 22,39 | die Koalition der | 56 | -24,78 | im Weissen Haus |
| 22 | 22,39 | der frühere der | 57 | -24,78 | der Internationalen Atomenergieagentur |
| 23 | 22,39 | sich die Frage | 58 | -23,33 | der Ministerpräsident Sharon |
| 24 | 21,07 | Beginn der in | 59 | -21,87 | Usama bin Ladin |
| 25 | 21,07 | in den Reihen | 60 | -20,3 | der Suche nach |
| 26 | 19,76 | die Fortsetzung der | 61 | -18,95 | die in Darfur |
| 27 | 17,12 | Kim Dae Jung | 62 | -18,95 | Debatte über die |
| 28 | 15,81 | Kim Young Sam | 63 | -17,49 | in den Tod |
| 29 | 17,12 | die Wege geleitet | 64 | -16,48 | eine Lösung des |
| 30 | 15,81 | gegen die Korruption | 65 | -16,04 | die Opfer der |
| 31 | 15,81 | der in Grosny | 66 | -16,04 | Saddam Hussein der |
| 32 | 15,81 | in der DDR | 67 | -16,04 | Zahl der auf |
| 33 | 15,81 | Bundeskanzler Kohl in | 68 | -16,04 | der gemässigten Islamisten |
| 34 | 15,81 | Präsident Clinton und | 69 | -15,3 | hat den letzten |
| 35 | 15,65 | die Tatsache dass | | | |

*Tabelle 2*
*Typische Mehrworteinheiten im Ressort „Ausland" in den Zeitperioden 1995–1997 (links) und 2003–2005 (rechts)*[12]

Tabelle 2 zeigt eine Auswahl an Mehrworteinheiten, die für die Periode 1995–1997 bzw. 2003–2005 jeweils signifikant sind. Der corpus-driven-Zugang bringt einerseits Mehrworteinheiten hervor, bei denen sofort einleuchtet, dass sie typisch für die ältere oder neuere Zeitperiode sind. Dazu gehören Ausdrücke, die die Jugoslawienkriege widerspiegeln oder die damaligen politischen Akteure nennen. Oder aber die Mehrworteinheiten, die auf den Terrorismus nach dem 11. September 2001 und den Irakkrieg verweisen.

Daneben gibt es Ausdrücke, deren Typik nicht unmittelbar einleuchten mag: *die Fortsetzung der, die Tatsache dass, die Kontrolle über, Zahl der auf* etc.

[12] Es handelt sich dabei um eine manuelle Auswahl für die weitere Analyse. Die vollständige Version ist unter http://www.bubenhofer.com/korpusanalyse/ zu finden.



Corpus-based-Analysen müssten aufzeigen, warum diese Ausdrücke für die eine oder andere Periode typisch sind.

Es lohnt sich aber auch ein genauerer Blick auf vermeintlich sofort einleuchtende Mehrworteinheiten. So stellt sich beispielsweise die Frage, ob der Ausdruck *die bosnischen Serben* zu einem allgemeineren Muster „Ethnienbezeichnungen" abstrahiert werden kann. Denn alternativ zur Nennung von Ethnien (*Bosnier, Serben, Jugoslawen, bosnische Serben* etc.), könnten auch Länder genannt werden: *Bosnien, Serbien, Jugoslawien* etc. Dass Ethnien statt Staatsnamen verwendet werden, ist bei einem zerfallenden Staatengebilde nicht verwunderlich. Detailliertere Untersuchungen erlauben aber die Hypothese, dass Ethnienbezeichnungen ganz allgemein in kriegerischen Kontexten wichtig sind – und deren Ausbleiben ein Indikator für nicht-kriegerische Kontexte ist (vgl. dazu ausführlich Bubenhofer 2009).

Ein genauerer Blick lohnt sich aber auch auf die Mehrworteinheiten *Kampf gegen Terrorismus, im Kampf gegen* etc., die motivieren, die Kontextualisierungsprofile dieser Ausdrücke genauer zu analysieren. Es scheint zwar plausibel, dass *Kampf gegen Terror**[13] ein Ausdruck ist, der typisch für die Zeit nach dem 11. September 2001 ist, wie Tabelle 3 zeigt. Doch allgemeiner ist zu fragen nach dem Gebrauch des Musters *Kampf gegen X* und den Füllungen des „Slots" X.

| Jahr | Kampf(s/es) gegen ... Terror* | | Kampf(s/es) gegen | | Total |
|---|---|---|---|---|---|
| | # | % | # | % | # |
| 1995 | 2 | 0,05 | 62 | 1,69 | 3666 |
| 1996 | 3 | 0,08 | 58 | 1,56 | 3711 |
| 1997 | 1 | 0,03 | 55 | 1,44 | 3827 |
| 1998 | 3 | 0,08 | 58 | 1,60 | 3616 |
| 1999 | 5 | 0,12 | 44 | 1,04 | 4248 |
| 2000 | 2 | 0,04 | 61 | 1,36 | 4496 |
| 2001 | 19 | 0,43 | 82 | 1,88 | 4373 |
| 2002 | 22 | 0,50 | 67 | 1,53 | 4385 |
| 2003 | 16 | 0,38 | 70 | 1,65 | 4235 |
| 2004 | 17 | 0,41 | 71 | 1,71 | 4155 |

[13] Der Stern (*) steht für beliebig viele zusätzliche Zeichen (außer Leerzeichen). Hier und im Folgenden wird damit angezeigt, dass in der Suche auch andere Endungen des Wortes gefunden würden. *Terror** findet *Terror, Terrorismus, Terroristen* etc.

| | | | | | |
|---|---|---|---|---|---|
| 2005 | 18 | 0,44 | 65 | 1,57 | 4131 |
| **Total** | **108** | | **693** | | **44843** |
| | $X^2 = 64{,}222$ | | $X^2 = 13{,}403$ | | |
| | $p < 0{,}001$ (signifikant) | | $p > 0{,}10$ (nicht signifikant) | | |
| | $V = 0{,}038$ | | $V = 0{,}017$ | | |
| | $\varnothing = 9{,}8 \pm 8$ | | $\varnothing = 63 \pm 9{,}4$ | | |

*Tabelle 3*
*Verteilungen und Statistiken (Einheit: Artikel) für die Mehrworteinheiten KAMPF(S/ES) GEGEN ... TERROR* bzw. KAMPF(S/ES) GEGEN ... im NZZ-Korpus (alle Ressorts; für die Statistiken gilt: df = 10)*

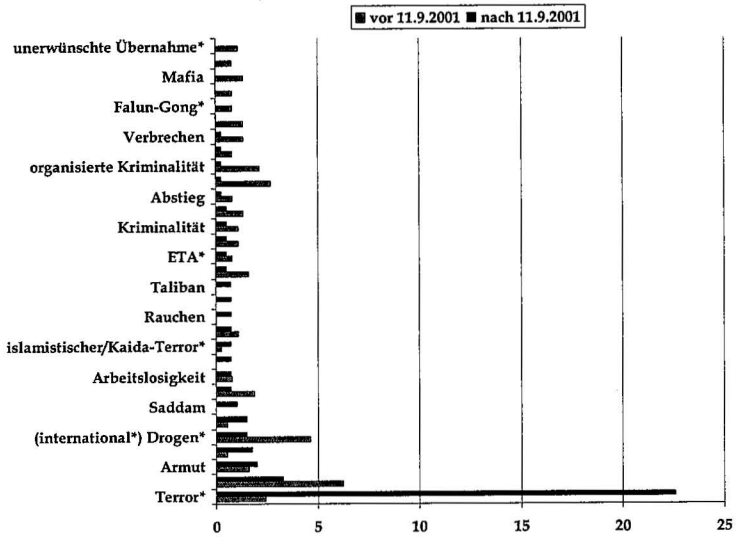


*Abbildung 1*
*Die Füllungen von X im Muster KAMPF GEGEN X [Artikel] mit Mindestfrequenz 3 für die Zeit vor und nach dem 11. September 2001*

Tabelle 3 zeigt die Frequenzen von *Kampf gegen* im NZZ-Korpus. Die Verteilung ist nicht signifikant, die Verwendung also über die Jahre ungefähr ähnlich. Abbildung 1 zeigt die Füllungen von X im Muster *Kampf gegen X* für die Perioden vor und nach dem 11. September 2001.

Für die weitere Analyse interessiert im Besonderen, ob der Ausdruck je nach Zeit oder Thema unterschiedliche Kontextualisierungsleistungen erbringt. Im Beispiel von *Kampf gegen X* wäre also zu fragen:

1. Gibt es zu bestimmten Zeiten oder in bestimmten Kontexten typische Füllungen des Slots X im Muster *Kampf gegen X?*
2. Welche alternativen Ausdrucksmöglichkeiten zum Muster *Kampf gegen X* existieren? Und wann werden sie verwendet?



Die notwendigen Analysen, um die erste Fragestellung zu verfolgen, sind einsichtig: In erster Linie geht es darum, die Füllung von X im Muster *Kampf gegen X* in Abhängigkeit von Zeit, Thema, Textsorte etc. zu untersuchen. Die Analysen um die zweite Fragestellung zu verfolgen, können in folgende Richtung gehen:

Als engere alternative Ausdrucksmöglichkeiten zu *Kampf gegen X* liegen folgende Varianten auf der Hand: *Kampf mit X* und *Kampf dem/der* [Dativobjekt].[14] Es ist zu fragen, ob der propositionale Gehalt der drei Varianten gleich ist. Die möglichen syntaktischen Einbettungen der Ausdrücke zeigen, dass die Ausdrucksvarianten erst als ausgebaute Satzteile syntaktisch austauschbar werden, da sie teilweise syntaktische Restriktionen aufweisen:

1. *Kampf gegen den Terrorismus*
- Jemand fordert den Kampf gegen den Terrorismus.
- Jemand befindet sich im Kampf gegen den Terrorismus.

2. *Kampf mit dem Terrorismus*
- Jemand fordert den Kampf mit dem Terrorismus.
- Jemand befindet sich im Kampf mit dem Terrorismus.

3. *Kampf dem Terrorismus*
- Jemand fordert: Kampf dem Terrorismus!
- Heute gilt: Kampf dem Terrorismus!

Die (erfundenen) Beispielsätze lassen vermuten, dass die Ausdrucksvarianten nicht alle dieselbe Konnotation aufweisen müssen und dass vor allem der Ausdruck *Kampf dem Terrorismus* eine ganz bestimmte pragmatische Funktion aufweist, nämlich als Parole wirkt.

Das NZZ-Korpus enthält keine Belege für die Varianten 2 und 3, und nur wenige Belege für die von Varianten 2 und 3 abstrahierten Muster *Kampf mit X* und *Kampf dem/der* [Dativobjekt]. Die Füllungen für X in den beiden Mustern sind in Tabelle 4 aufgeführt. Dabei wird sofort klar, dass sich in den untersuchten Daten deutliche Unterschiede im Sprachgebrauch zeigen, wobei es aber schwierig ist, die Unterschiede zu benennen. Vielleicht lässt es sich so fassen: *Kampf dem/der* [Dativobjekt] wird überhaupt nicht, und *Kampf mit X* [Artikel] nur vereinzelt in kriegerischen Kontexten verwendet. Etwas anders zeigt sich das Bild bei *Kampf gegen X* (vgl. Abbildung 1). Bei diesem Muster fallen die hochfrequenten Kollokatoren wie *Terror*, Saddam, islamistischer/Kaida-Terror*, Taliban, ETA* etc. auf, die in kriegerischen Kontexten zu finden sind.

[14] Weitere Ausdrucksvarianten könnten mit dem Lexem *Krieg* gebildet werden. Der Einfachheit halber soll davon abgesehen werden.

| *Kampf mit X* [Artikel][15] | *Kampf dem/der* [Dativobjekt] |
|---|---|
| allen Mitteln gegen Israel; Aussicht auf Erfolg; baskischen Separatisten und Terroristen; Bayern München; Behörden; Berg; Depression; Drachen; drohenden Überschuldung; eigenen Schwächen; Elementen; erbarmungslosen Feind; Gegnern; gleich langen Spiessen; gleicher Waffe; Grammatik; Great-West Lifeco; harten Bandagen; HIV; ihnen; ihren Texten; Immunsystem; Klavierschülergeschlecht; Konkurrenz; Leuten; Mächten einer niederdrückenden Ideologie; nassen Element; Neat-Gegnern; Schönheit; sehr jungen Gegnern; sich selber; sich selbst; Trieb; ungewissem Ausgang; vorpreschenden Spunden; Wettbewerbskommission; Wind, Wasser und Wellen; x Toren; Zeit | „Bussen- und Gebührenterror"; Dengue-Fieber; Diskriminierung von HIV-Positiven; Feinstaubbelastung; Guineawurm; Korruption; Kriminalität; Lobbying; Minenplage; Neoliberalismus; Schuldenmisswirtschaft; Schwefel; Spam; Stau; Straflosigkeit |

*Tabelle 4*
*Die Füllungen von X in den Mustern KAMPF MIT X [Artikel] und KAMPF DEM/DER [Dativobjekt] im NZZ-Korpus*

Dieser Befund steht aufgrund der Datenlage auf schwachen Füßen. Es gibt im NZZ-Korpus zu wenig Belege für die beiden Muster *Kampf mit X* und *Kampf dem/der* [Dativobjekt]. Es bietet sich deshalb an, ein weiteres Korpus, beispielsweise das DeReKo IDS-Korpus des Instituts für Deutsche Sprache, als Referenzkorpus heranzuziehen. Grundlage für die Recherche waren alle im IDS-Korpus öffentlich verfügbaren Tageszeitungen aus Deutschland, Österreich und der Schweiz im Zeitraum von 1991 bis 2006.[16] Die Tabellen 5 und 6 zeigen die häufigsten Füllungen für X in den drei Mustern.[17]

| *Kampf(s/es) gegen X* | # | % | *Kampf(s/es) mit X* | # | % |
|---|---|---|---|---|---|
| Abstieg | 453 | 4,53 | [Sportresultat X:Y] | 344 | 11,21 |
| Terror(ismus) | 329 | 3,29 | Behörde(n) | 32 | 1,04 |

[15] Bis auf *Kampf mit dem Berg* (vier Vorkommen) gibt es für jede Mehrworteinheit jeweils nur einen Beleg im Korpus.

[16] Das so zusammengestellte Korpus enthält damit 4.129.847 Texte mit insgesamt 960.395.973 Wörtern. Folgende Zeitungen sind in unterschiedlichen Zeiträumen vertreten: Berliner Morgenpost, Die Presse, Frankfurter Rundschau, Hamburger Morgenpost, Kleine Zeitung, Mannheimer Morgen, Neue Kronen-Zeitung, Oberösterreichische Nachrichten, Salzburger Nachrichten, St. Galler Tagblatt, Tiroler Tageszeitung, Vorarlberger Nachrichten, Zürcher Tagesanzeiger.

[17] Für das Muster *Kampf dem/der* [Dativobjekt] mussten im IDS-Korpus Belege mit Dativobjekten mit femininem Genus *(Kampf der Arbeitslosigkeit)* ignoriert werden, da das Korpus nicht syntaktisch annotiert ist und deshalb Dativobjekte nicht von Genitivobjekten *(der Kampf der Regierung)* unterschieden werden konnten.



| | | | | | |
|---|---|---|---|---|---|
| organisiert* | 225 | 2,25 | sich selbst/sich selber | 31 | 1,01 |
| Doping | 183 | 1,83 | Drachen | 26 | 0,85 |
| Korruption | 161 | 1,61 | Natur | 21 | 0,68 |
| Armut | 129 | 1,29 | Bürokratie(n) | 20 | 0,65 |
| Drogen | 119 | 1,19 | Tücke(n) des/der | 17 | 0,55 |
| Geldwäsche/ Geldwäscherei | 105 | 1,05 | allen Mitteln | 16 | 0,52 |
| Kriminalität | 96 | 0,96 | Elementen | 16 | 0,52 |
| Uhr | 94 | 0,94 | Konkurrenten/ Konkurrenz | 16 | 0,52 |
| Mafia | 90 | 0,90 | Waffe(n) | 16 | 0,52 |
| AIDS | 79 | 0,79 | | | |
| illegale(n) | 77 | 0,77 | | | |
| Windmühle(n) | 77 | 0,77 | | | |
| internationale(n) | 76 | 0,76 | | | |
| Krebs | 62 | 0,62 | | | |
| seine(n) | 61 | 0,61 | | | |
| Rassismus | 58 | 0,58 | | | |
| Israel | 51 | 0,51 | | | |
| Verbrechen | 51 | 0,51 | | | |
| **Total** | **10000** | **100,00** | **Total** | **3070** | **100,00** |

*Tabelle 5*
*Die Füllungen von X in den Mustern KAMPF GEGEN X und KAMPF MIT X im IDS-Korpus, deren Frequenzen mindestens 0,5 % aller Kollokatoren ausmachen*

| *Kampf dem X* | # | % | (Fortsetzung ...) | # | % |
|---|---|---|---|---|---|
| Krebs | 33 | 5,45 | Faschismus | 4 | 0,66 |
| Stau | 32 | 5,28 | Hochwasser | 4 | 0,66 |
| Hunger | 21 | 3,47 | Müll | 4 | 0,66 |
| Dickdarmkrebs | 13 | 2,15 | Schimmel* | 4 | 0,66 |
| Brustkrebs | 10 | 1,65 | Tod | 4 | 0,66 |
| Elektrosmog | 9 | 1,49 | Übergewicht | 4 | 0,66 |
| Fahrraddiebstahl | 9 | 1,49 | Atomtod | 3 | 0,50 |
| Herztod | 8 | 1,32 | blauen Dunst | 3 | 0,50 |

| | | | | | |
|---|---|---|---|---|---|
| Schlaganfall | 8 | 1,32 | Chaos | 3 | 0,50 |
| Terror* | 8 | 1,32 | Darmkrebs | 3 | 0,50 |
| Feuerbrand | 7 | 1,16 | Feinstaub | 3 | 0,50 |
| Rassismus | 7 | 1,16 | Grünen Star | 3 | 0,50 |
| Sterilen und Leblosen | 7 | 1,16 | Hanf | 3 | 0,50 |
| Alkohol | 6 | 0,99 | Kommunismus | 3 | 0,50 |
| Kinderkrebs | 6 | 0,99 | Mißbrauch | 3 | 0,50 |
| Pusch | 6 | 0,99 | Orgasmus | 3 | 0,50 |
| Doping | 5 | 0,83 | Proporz | 3 | 0,50 |
| Kindsmissbrauch | 5 | 0,83 | Schilderwald | 3 | 0,50 |
| Verbrechen | 5 | 0,83 | Sozialmissbrauch | 3 | 0,50 |
| Winterspeck | 5 | 0,83 | Verkehrschaos | 3 | 0,50 |
| Drogen* | 4 | 0,66 | | | |
| **Total** | | | | **606** | **100,00** |

*Tabelle 6*
*Die Füllungen von X im Muster KAMPF DEM X im IDS-Korpus, deren Frequenzen mindestens 0,5 % aller Kollokatoren ausmachen*

Mit Hilfe dieser Analysen lässt sich der oben formulierte Befund präzisieren:

1. *Kampf gegen X:* Mit den dominanten Füllungen verweist der propositionale Gehalt des Ausdrucks auf eine internationale, politische Dimension, die wohl als „gesellschaftlich bedeutend" umrissen würde. Ausnahme davon: Der *Kampf gegen den Abstieg,* der im sportlichen Kontext verwendet wird.
2. *Kampf mit X:* Mit den dominanten Füllungen beschreiben die Ausdrücke eher Anekdotisches *(Kampf mit den Behörden,* der *Bürokratie,* den *Tücken* [der Technik]), oder referieren auf die eigene Persönlichkeit *(Kampf mit sich selbst).* Die Verwendung ist aber sehr heterogen. An der Spitze stehen auch hier Verwendungen im Zusammenhang mit Sportresultaten: *Den Kampf mit 2 : 0 (gewonnen).*
3. *Kampf dem X:* Auffallend sind hier die dominanten Verwendungen mit dem Lexem *Krebs-* und Bezeichnungen für andere Krankheiten. Mehrheitlich bewegen sich die Ausdrücke eher auf der individuellen und durchaus tragischen Ebene, daneben werden sie aber auch in Kontexten der Innenpolitik oder von Problemen im Bereich „Naturgewalten" und „Verkehr" genannt.

Neben diesen Präferenzen für Füllungen von X gibt es auch Belege, die zeigen, dass die Muster teilweise austauschbar sind. So finden sich sowohl



Belege für *Kampf gegen das Verbrechen* als auch *Kampf dem Verbrechen* oder *Kampf gegen Rassismus* und *Kampf dem Rassismus* und andere mehr.

Interessant im Zusammenhang mit *Terror* ist besonders die seltene Verwendung von *Kampf dem Terror** im Vergleich zu anderen Füllungen des Ausdrucks *Kampf dem X.* Gibt es, neben den jeweiligen syntaktischen Restriktionen, einen semantischen oder pragmatisch-funktionalen Unterschied zwischen *Kampf dem Terror** und *Kampf gegen den Terror**? Fordert der politische Terror-Diskurs eine bestimmte Sprachgebrauchsvariante?

Alle acht Belege für *Kampf dem Terror** im IDS-Korpus sind Verwendungen in Titeln oder als Schlagzeilen von Zeitungsartikeln und haben Parolen-Charakter. Bei dreien können direkte oder indirekte Zitate dahinter vermutet werden:

(1) Vranitzky, Schüssel: **Kampf dem Terror,** ob von links oder rechts! *(Neue Kronen-Zeitung vom 27. April 1995, 2)*

(2) US-Gegenoffensive mit Atomwaffen? **„Kampf dem Terrorismus** mit allen verfügbaren Mitteln" *(Salzburger Nachrichten vom 24. August 1998, Ressort: „Weltpolitik")*

(3) **Kampf dem Terrorismus.** Der neue FBI-Chef Louis Freeh bezeichnete es als eine seiner wichtigsten Aufgaben, scharf gegen den Terrorismus vorzugehen. *(Die Presse vom 4. September 1993, Ressort „In Kürze")*

Bei übersetzten Zitaten können sprachliche Feinheiten, wie die Unterscheidung von *Kampf gegen* und *Kampf dem,* verloren gehen. Doch aus diskursanalytischer Sicht ist es trotzdem interessant zu sehen, welche Sprachgebrauchsvariante auch in der Übersetzung für einen bestimmten Zweck verwendet wird. Aufgrund der gemachten Beobachtungen könnte z.B. die These aufgestellt werden, dass mit *Kampf dem Terror** eine Sprachgebrauchsvariante verwendet wird, die politische Rhetorik, und vielleicht sogar Distanzierung des Verfassers/der Verfasserin gegenüber dieser Rhetorik, signalisieren soll. Weitere Evidenz für diese These ergibt eine unsystematische Suche in deutschsprachigen Tageszeitungen, wie folgende Belege zeigen:

Die konservative Zeitung fordert mehr **Kampf dem Terrorismus:**

(4) Deutschland wird in den kommenden Jahren weiter massiv in den Ausbau der Sicherheitsarchitektur investieren müssen. Es wäre gut, wenn die staatstragenden Parteien, zu denen man neben Union und SPD auch Grüne und FDP zählen möchte, sich darüber gemeinsam Gedanken machten. *(SonntagsZeitung vom 9. September 2007, 23)*

(5) Geiger ist sich sicher, dass für den Finanzplatz Schweiz die Bedeutung des Bankgeheimnisses steigen wird. Denn mit der Begründung **„Kampf dem Terrorismus"** nähmen die Zugriffsbegehrlichkeiten auf die Privatsphäre in vielen Ländern zu. *(Süddeutsche Zeitung vom 29. November 2006, V2/2)*

(6) Ein weiteres Beispiel für Abzocke in großem Stil: die Gebührenerhöhung für den neuen Reisepass um 127 Prozent, natürlich unter dem Deckmantel **„Kampf dem Terrorismus".** *(Stuttgarter Zeitung vom 31.10.2005, Ressort „Leserbriefe", 40)*

(7) Das Schlagwort **„Kampf dem Terrorismus"** muss herhalten, die Folgen dieses Kriegs um Einfluss und Öl zu rechtfertigen. Sind nicht auch die Deutschen mitschuldig, wenn die amerikanischen Stützpunkte im Land für Militärtransporte etc. benützt werden? *(Stuttgarter Zeitung vom 16. November 2004, Ressort „Leserbriefe", 8)*

Die Liste ließe sich fortsetzen. Es zeigt sich, dass die Verwendung von *Kampf dem Terror** hier (z.B. in Leserbriefen) in argumentativer Funktion verwendet wird, um diesen „Kampf" zu kritisieren.

Die bisherigen Analysen führen zu folgenden Hypothesen:

1. Die Muster *Kampf gegen X, Kampf mit X* und *Kampf dem X* können grundsätzlich denselben propositionalen Gehalt aufweisen. Trotzdem finden sich klare Präferenzen, in welchen Kontexten (mit welchen X) die Muster verwendet werden. Diese Präferenzen verändern sich diachron und unterscheiden sich synchron (z.B. bezüglich Textsorte, kommunikativer Funktion etc.). In welcher Form und syntaktischen Umgebung sie erscheinen, ist abhängig von Diskursen.
2. Für *Kampf gegen (den) Terror** und *Kampf dem Terror** gelten in den untersuchten Daten relativ eindeutige Präferenzen für bestimmte kommunikative Zwecke.
3. In den untersuchten Daten wird der Ausdruck *Kampf gegen (den) Terror* weit häufiger verwendet, als die ebenfalls möglichen Alternativen *Kampf dem Terror* oder *Kampf mit dem Terror*. Die Präferenz für eine Variante ist diskursiv begründet. Die vermehrte Verwendung von *Kampf dem Terror* würde z.B. auf einen veränderten Status von *Terror* schließen lassen, einen ähnlichen Status, wie er heute z.B. auch für *Krebs* oder *Stau (Kampf dem Krebs/Stau)* gilt. Der propositionale Gehalt *gegen den Terror kämpfen* bliebe zwar bestehen, die Konnotation wäre jedoch von der heutigen verschieden.

Die Analyse der Kontextualisierungsprofile der drei Muster *Kampf gegen X, Kampf mit X* und *Kampf dem X* fördert also nicht nur thematisch, sondern auch pragmatisch definierte Kontexte hervor. So erscheint die Parole *Kampf dem Terror* oft in Leserbriefen und weist dort eine argumentative Funktion auf. Diese typischen pragmatischen Funktionen gelten aber oft, das haben die Analysen ebenfalls gezeigt, nur für bestimmte Textsorten und können nicht unbedingt generalisiert werden. Der Schlüssel, um die diskursiven Formierungen freizulegen, liegt in der Suche nach Veränderungen in den Kontextualisierungsprofilen.

## 6 Fazit

Es scheint plausibel, Sprachgebrauchsmuster als eigentliche Keimzellen von Kontextualisierung zu verstehen. Sprachgebrauchsmuster sind Kombinationen von Elementen der sprachlichen Oberfläche, die in genau diesen



Kombinationen typisch für einen bestimmten Sprachausschnitt sind. In der vorliegenden Untersuchung wurden nur Wortformen als Elemente von Sprachgebrauchsmustern beachtet. Es wäre nahe liegend, auch Wortarten, Lemmata und weitere syntaktische oder semantische Elemente in die Berechnung von Sprachgebrauchsmustern zu integrieren.[18]

Wenn Sprachgebrauchsmuster als Phänomen an der Textoberfläche verstanden werden, dann sind sie damit empirisch fassbar. Damit wird es möglich, mit statistischen Methoden zu berechnen, welche Sprachgebrauchsmuster in bestimmten Korpora im Vergleich zu anderen Korpora typisch sind. Solche quantitativen Methoden leisten weit mehr, als pure Auftretensfrequenzen von Entitäten an der sprachlichen Oberfläche zu berechnen. Statistische Tests erlauben es, Abweichungen vom Sprachgebrauch eines Referenzkorpus anzuzeigen, egal ob es sich dabei um niedrig- oder hochfrequente Phänomene handelt. Um aber corpus-driven auf den typischen Sprachgebrauch eines Untersuchungskorpus zu stoßen, sind nicht in erster Linie raffinierte statistische Methoden notwendig, sondern geschickte und systematische Vergleiche mit möglichst umfassenden Referenzkorpora. Je umfangreicher die Datengrundlage, desto genauer können die typischen Sprachgebrauchsmuster beschrieben werden.

## Literatur

BANERJEE, Satanjeev/PEDERSEN, Ted (2003): „The Design, Implementation, and Use of the Ngram Statistics Package". In: Alexander Gelbukh (Hrsg.): *Computational Linguistics and Intelligent Text Processing. Proceedings of the Fourth Annual Conference on Intelligent Text Processing and Computational Linguistics.* Berlin: Springer, 370-381 (= Lecture Notes in Computer Science 2588).

BELICA, Cyril (1996): „Analysis of Temporal Changes in Corpora". In: *International Journal of Corpus Linguistics,* Volume 1, Nr. 1, 61–73.

BELICA, Cyril (2001-2007): „Kookkurrenzdatenbank CCDB. Eine korpuslinguistische Denk- und Experimentierplattform für die Erforschung und theoretische Begründung von systemisch-strukturellen Eigenschaften von Kohäsionsrelationen zwischen den Konstituenten des Sprachgebrauchs". In: *http://corpora.ids-mannheim.de/ccdb/.*

BUBENHOFER, Noah (2008): „Diskurse berechnen? Wege zu einer korpuslinguistischen Diskursanalyse". In: Jürgen Spitzmüller/Ingo H. Warnke (Hrsg.): *Methoden der Diskurslinguistik. Sprachwissenschaftliche Zugänge zur transtextuellen Ebene.* Berlin: de Gruyter, 407-434 (= Linguistik – Impulse & Tendenzen 31).

[18] Vgl. dazu das Konzept der „Collostructions" von Stefanowitsch und Gries (2003). Gegenwärtig wird zudem im Forschungsprojekt „semtracks, Tracking Meaning on the Surface: A Data-Driven Approach to Semantic Imprints of Texts" die Methode einer umfassenden Clusteranalyse entwickelt, die eine Vielzahl von Phänomenen der Textoberfläche miteinbezieht. Vgl. dazu http://www.semtracks.com sowie Bubenhofer/Klimke/Scharloth (2008) als Beispiel für eine linguistische Analyse des US-Wahlkampfes Obama/McCain 2008 bzw. Bubenhofer/Klimke/Scharloth (2009) als Analysen zur deutschen Bundestagswahl 2009.

BUBENHOFER, Noah (2009): *Sprachgebrauchsmuster. Korpuslinguistik als Methode der Diskurs- und Kulturanalyse.* Berlin: de Gruyter (= Sprache und Wissen 4).

BUBENHOFER, Noah/KLIMKE, Martin/SCHARLOTH, Joachim (2008): „The Word War: ‚Yes, He Did' - How Obama Won the (Rhetorical) Battle for the White House". In: *http://www.isn.ethz.ch/isn/Current-Affairs/Special-Reports/The-Word-War-Yes-He-Did/Analysis.*

BUBENHOFER, Noah/KLIMKE, Martin/SCHARLOTH, Joachim (2009): „semtracks political tracker - Bundestagswahl '09. Eine Semantische Matrixanalyse". In: *http://semtracks.com/politicaltracker/bundestagswahl/.*

BURGER, Harald (1998): *Phraseologie. Eine Einführung am Beispiel des Deutschen.* Berlin: Erich Schmidt (= Grundlagen der Germanistik 36).

CHENG, Winnie/GREAVES, Chris/WARREN, Martin (2006): „From N-Gram to Skipgram to Concgram". In: *International Journal of Corpus Linguistics,* Volume 11, Nr. 4, 411–433.

DEREKO - Das Deutsche Referenzkorpus. In: *http://www.ids-mannheim.de/kl/projekte/dereko/.*

FEILKE, Helmuth (1993): „Sprachlicher Common sense und Kommunikation. Über den ‚gesunden Menschenverstand', die Prägung der Kompetenz und die idiomatische Ordnung des Verstehens". In: *Der Deutschunterricht,* Nr. 6/93: Gefrorene Sprache, 6–21.

FEILKE, Helmuth (1996): *Sprache als soziale Gestalt. Ausdruck, Prägung und die Ordnung der sprachlichen Typik.* Frankfurt a. M.: Suhrkamp.

FEILKE, Helmuth (2000): „Die pragmatische Wende in der Textlinguistik". In: Klaus Brinker/Gerd Antos/Wolfgang Heinemann/Sven F. Sager (Hrsg.): *Text- und Gesprächslinguistik. Ein internationales Handbuch zeitgenössischer Forschung/Linguistics of Text and Conversation. An International Handbook of Contemporary Research.* 1. Halbband. Berlin: de Gruyter, 64–82 (= HSK 16.1).

FEILKE, Helmuth (2003): „Textroutine, Textsemantik und sprachliches Wissen". In: Angelika Linke/Hanspeter Ortner/Paul R. Portmann (Hrsg.): *Sprache und mehr. Ansichten einer Linguistik der sprachlichen Praxis.* Tübingen: Niemeyer, 209–230 (= RGL 245).

FILLMORE, Charles (1976): „Pragmatics and the Description of Discourse". In: Siegfried J. Schmidt (Hrsg.): *Pragmatik/Pragmatics 2. Zur Grundlegung einer expliziten Pragmatik.* München: Fink, 83–104 (= Kritische Information 25).

NERLICH, Brigitte (1995): „The 1930s - at the Birth of a Pragmatic Conception of Language". In: *Historiographia Linguistica,* Volume 22, Nr. 3, 311–334.

PERKUHN, Rainer/BELICA, Cyril (2006): „Korpuslinguistik - das unbekannte Wesen. Oder Mythen über Korpora und Korpuslinguistik". In: *Sprachreport,* Nr. 1/2006, 2–8.

PERKUHN, Rainer/BELICA, Cyril/WADI, Doris al/LAUER, Meike/STEYER, Kathrin/WEISS, Christian (2005): „Korpustechnologie am Institut für Deutsche Sprache". In: Johannes Schwitalla/Werner Wegstein (Hrsg.): *Korpuslinguistik deutsch: synchron – diachron – kontrastiv. Würzburger Kolloquium 2003.* Tübingen: Niemeyer, 57–70.

SINCLAIR, John (1991): *Corpus, Concordance, Collocation.* Oxford: Oxford University Press.

STEFANOWITSCH, Anatol/GRIES, Stefan T. (2003): „Collostructions: Investigating the Interaction of Words and Constructions". In: *International Journal of Corpus Linguistics,* Volume 8, Nr. 2, 209–243.



STEYER, Kathrin (2004a): „Kookkurrenz. Korpusmethodik, linguistisches Modell, lexikografische Perspektiven". In: Kathrin Steyer (Hrsg.): *Wortverbindungen – mehr oder weniger fest.* Berlin: de Gruyter, 87–116 (= Jahrbuch des Instituts für Deutsche Sprache 2003).

STEYER, Kathrin, Hrsg. (2004b): *Wortverbindungen – mehr oder weniger fest.* Berlin: de Gruyter (= Jahrbuch des Instituts für Deutsche Sprache 2003).

STEYER, Kathrin/LAUER, Meike (2007): „Corpus-Driven: Linguistische Interpretation von Kookkurrenzbeziehungen". In: Ludwig M. Eichinger/Heidrun Kämper (Hrsg.): *Sprach-Perspektiven. Germanistische Linguistik und das Institut für Deutsche Sprache.* Tübingen: Narr, 493–509 (= Studien zur deutschen Sprache 40).

TEUBERT, Wolfgang (2005): „My Version of Corpus Linguistics". In: *International Journal of Corpus Linguistics,* Volume 10, Nr. 1, 1–13.

TOGNINI-BONELLI, Elena (2001): *Corpus Linguistics at Work.* Amsterdam: Benjamins (= Studies in Corpus Linguistics 6).

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.d-nb.de abrufbar.

© 2010 · Narr Francke Attempto Verlag GmbH + Co. KG
Dischingerweg 5 · D-72070 Tübingen

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Verlages unzulässig und strafbar. Das gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.
Gedruckt auf säurefreiem und alterungsbeständigem Werkdruckpapier.

Internet: http://www.narr.de
E-Mail: info@narr.de

Umschlagbild: Paul Klee, Labiler Wegweiser, 1937
Druck und Bindung: Ilmprint, Langewiesen
Printed in Germany

ISSN 1860-7373
ISBN 978-3-8233-6490-0

# Inhaltsverzeichnis

## I Theoretische Aspekte zur Interdependenz von Kontext - Text - Kontext

*Paul R. Portmann-Tselikas/Georg Weidacher*
Nicht nur zur Begrifflichkeit.
Kontexte, Kommunikation und Kompetenzen ..... 9

*Alexander Ziem*
Welche Rolle spielt der Kontext beim Sprachverstehen?
Zum Stand der psycholinguistischen und kognitionswissenschaftlichen Forschung ..... 59

*Noah Bubenhofer/Joachim Scharloth*
Kontext korpuslinguistisch.
Die induktive Berechnung von Sprachgebrauchsmustern in großen Textkorpora ..... 85

*Mechthild Habermann*
Historische Redeweisen über Texte.
Zur hermeneutischen Macht fester Kontexte ..... 109

## II Kontexte und literale Praktiken

*Angelika Linke*
Textsorten als Elemente kultureller Praktiken.
Zur Funktion und zur Geschichte des Poesiealbumeintrags als Kernelement einer kulturellen Praktik ..... 127

*Helmuth Feilke*
Kontexte und Kompetenzen - am Beispiel schriftlichen Argumentierens ..... 147

*Torsten Steinhoff*
Kontexttransposition.
Studentisches Schreiben zwischen Journalismus und Wissenschaft ..... 167

Europäische Studien zur Textlinguistik

herausgegeben von
Kirsten Adamzik (Genf)
Martine Dalmas (Paris)
Jan Engberg (Aarhus)
Wolf-Dieter Krause (Potsdam)
Arne Ziegler (Graz)

Band 8

Peter Klotz / Paul R. Portmann-Tselikas
Georg Weidacher (Hrsg.)

# Kontexte und Texte

## Soziokulturelle Konstellationen literalen Handelns

narr
VERLAG