

# Korpusbasierte Analyse der Semantik von Idiomen\*

Katerina Stathi (Berlin)

---

## Abstract

Idioms are usually defined in semantic terms as non-compositional units whose meaning does not result from the meaning of their parts. This definition implies that idioms have exactly one meaning. But if we examine idioms in context in large electronic corpora, we observe that the meaning of idioms is contextually variable, like that of words. We observe uses that exceed the range of the conventionalized and (lexicographically) codified meaning. A case study (the German idiom *ins Gras beißen*, literally "to bite into the grass", 'to die, to bite the dust') is used as an illustration of the relation between core meaning and contextual variability or meaning extension in idioms and the underlying mechanisms of this process.

---

## 1 Einleitung

Die Korpuslinguistik, die in den 1980er Jahren im englischsprachigen Raum ihren Ursprung fand und sich seitdem rasant entwickelt hat, stellt für die Sprachwissenschaft sowohl eine große Herausforderung als auch eine große Chance dar. Die Korpuslinguistik basiert auf Korpora, d. h. großen Sammlungen von Texten, die als Datenbasis und Evidenz für die Beschreibung und Erklärung sprachlicher Phänomene dienen können.<sup>1</sup> Dadurch zeichnet sich der korpusbasierte Ansatz durch eine Reihe von Merkmalen aus. Korpusbasierte Analysen sind empirisch, d. h. sie beruhen auf Sprachdaten und nicht allein auf der Intuition der Linguisten. Zudem sind die Texte, aus denen Korpora bestehen, Teil natürlicher Sprache und Kommunikation und wurden nicht eigens für linguistische Zwecke kreiert.

Per Definition werden in korpusbasierten Studien Sprachverwendung bzw. Sprache im tatsächlichen Gebrauch untersucht. Die Korpuslinguistik verpflichtet sich somit nicht dem Dualismus zwischen Kompetenz und Performanz, der die Linguistik in den letzten vier

---

\* Diese Studie ist im Rahmen des Projekts "Kollokationen im Wörterbuch" entstanden, das an der Berlin-Brandenburgischen Akademie der Wissenschaften ansässig ist und von der Alexander-von-Humboldt-Stiftung im Rahmen des Wolfgang-Paul-Preises und des Zukunftsinvestitionsprogramms der Bundesregierung finanziert wird.

Manfred Bierwisch, Christiane Fellbaum, Elke Gehweiler, Undine Kramer und Hans-Jürgen Sasse danke ich für Diskussion und Kommentare.

<sup>1</sup> Für einen Überblick über die Entwicklung und die Methoden der Korpuslinguistik cf. u. a. McEnery/Wilson (1996); Biber et al. (1998); Kennedy (1998); Aijmer/Altenberg (1991); Svartvik (1992); Stubbs (1996); Baker et al. (1993).

Jahrzehnten weitgehend dominiert hat. Da der Sprachgebrauch im Vordergrund steht, werden sprachliche Phänomene notwendigerweise als kontextualisierte Phänomene wahrgenommen, d. h. als in ihrem Kontext verankert, wobei dieser sozusagen "mitgeliefert" wird. Wir werden im Folgenden sehen, dass eine Abstrahierung vom Kontext nicht leicht ist. Stubbs (2001: 221) zufolge ist Korpuslinguistik aus diesem Grund inhärent soziolinguistisch, da die Texte, aus denen Korpora bestehen, echte Kommunikationsakte sind, die in einer Diskursgemeinschaft kommunikativen Zwecken dienen oder gedient haben. Eine natürliche Folge davon ist eine inhärente Variation in Korpora.<sup>2</sup>

Korpusbasierte Analysen sind außerdem inhärent textlinguistisch. In Korpora haben wir es nicht mit isolierten Wörtern und Sätzen zu tun, sondern mit zusammenhängenden Sinneinheiten. Für die Erforschung von semantischen und syntaktischen Fragen (z. B. Wortstellung, Informationsstruktur) sind daher die Methoden der Textlinguistik und der Diskursanalyse unverzichtbar.<sup>3</sup> Dabei sind kohärenzstiftende Mechanismen, wie z. B. Elaboration des Themas im Text, von großer Wichtigkeit für die Zuweisung einer Interpretation und der Bestimmung von Bedeutung im Text.

Die Korpuslinguistik ist inhärent quantitativ. Als wesentlicher Grundsatz gilt, dass Frequenz von Bedeutung ist und sich wiederholende Erscheinungen wichtig sind. Das grenzt korpusbasierte Analysen von bisherigen theoretischen Ansätzen ab, insbesondere von generativen. Durch die Wichtigkeit der statistischen Signifikanz wird der Begriff der Norm (cf. Coseriu 1975) zu einem zentralen Begriff: nicht im präskriptiven Sinne, sondern im Sinne von typischen und gewöhnlichen, erwartbaren und vorhersagbaren Erscheinungen. Nach Stubbs (2001: 221) sollte die Untersuchung von "normalen", d. h. usuellen Erscheinungen das erste Ziel der Korpuslinguistik sein. Einmalige und idiosynkratische Erscheinungen sind zu registrieren, doch diese können nur auf dem Hintergrund dessen beurteilt und beschrieben werden, was usuell und erwartbar ist.

Korpusbasierte Analysen gehen über quantitative Aussagen hinaus, da Korpuslinguistik sich nicht in der Aufzählung von Erscheinungen erschöpft, sondern neben quantitativen auch qualitative Methoden einsetzt. Letztere sollen Erklärungen für die durch die quantitative Analyse beobachteten Phänomene liefern (Biber et al. 1998: 9).

Stubbs (2001: 221) zufolge ist die Korpuslinguistik zudem inhärent diachron, denn sie untersucht, was in der Vergangenheit häufig gebraucht worden ist. Ein Korpus wird also als eine Ansammlung von Gebrauchsweisen verstanden, die sich durch die stetige Verwendung über einen längeren Zeitraum als Norm etabliert haben. Der "Jetztzustand" der Sprache wird als das Produkt einer Entwicklung angesehen, die sich in den etablierten Normen, d. h. den

---

<sup>2</sup> Dies hängt selbstverständlich auch von der Zusammensetzung des Korpus ab. Obwohl die meisten Korpora größtenteils Texte der Standardvarietät enthalten, sind doch oft diverse Textsorten vertreten, die durch unterschiedliche Merkmale gekennzeichnet sind. Sogar in Korpora, die hauptsächlich Zeitungstexte enthalten, ist Variation zu beobachten. Damit spiegelt sich in den Korpora die Heterogenität der Sprachgemeinschaft wider.

<sup>3</sup> Da die meisten Korpora hauptsächlich aus geschriebenen Texten erstellt sind, ist ein Nebeneffekt der Korpuslinguistik eine automatische Fokussierung auf geschriebene Sprache. Das heißt, obwohl Korpusstudien auf Sprachgebrauch basieren, haben wir es in den meisten Fällen nicht mit spontan produzierter Sprache zu tun.

statistisch signifikanten Erscheinungen manifestiert. Auch wenn man diese Auffassung nicht teilt, sind die meisten Korpora tatsächlich diachron, da sie Daten aus längeren Zeiträumen beinhalten.

Zusammenfassend kann festgestellt werden, dass korpusbasierte Ansätze empirisch und induktiv sind, d. h. von Daten ausgehend zur Theorie gelangen, auf authentischen Daten in natürlicher Kommunikation beruhen und den Gebrauch von Sprache einer sehr großen Menge von Sprechern einer Sprachgemeinschaft widerspiegeln.<sup>4</sup> Die Korpuslinguistik verwirft die Dichotomie zwischen Kompetenz und Performanz und relativiert die Dichotomie zwischen Synchronie und Diachronie.<sup>5</sup>

Obwohl die Korpuslinguistik empirische Untersuchungen in zahlreichen Bereichen der Linguistik ermöglicht und korpuslinguistische Studien auf vielen Gebieten der Linguistik Anwendung finden (Biber et al. 1998: 11), hat die computergestützte Analyse in erster Linie zu einem großen Fortschritt bei der Erforschung des Lexikons und der lexikalischen Semantik beigetragen (cf. dazu Arbeiten zur korpusbasierten Lexikografie und lexikalischen Semantik, z. B. Sinclair (1991) (1996), (1998); Atkins/Zampolli (1994); Hanks (1996); Stubbs (2001); Corréard (2002)).

Die korpuslinguistischen Methoden und die dabei verwendeten computerlinguistischen Techniken haben auch das Interesse an Mehrworteinheiten geweckt und ihre komplexe Untersuchung erheblich erleichtert und stark vorangetrieben, wie z. B. Moon (1998) mit einer korpusbasierten Analyse von Mehrworteinheiten und Idiomen des Englischen belegt.

Auf einen wesentlichen Grundbegriff korpusbasierter Analyse des Lexikons, den Begriff der Norm, habe ich bereits hingewiesen. Die Norm ergibt sich quantitativ durch das Kriterium der Frequenz (statistische Signifikanz), cf. Church et al. (1991); Church et al. (1994); Church/Hanks (1990); Church/Mercer (1994).<sup>6</sup>

Ferner sind *Patterns* relevant, d. h. wiederkehrende Muster in der Kombinatorik eines Wortes auf der syntagmatischen Ebene. Durch computerlinguistische und -gestützte Techniken (z. B. Konkordanzen) haben Korpora Gebrauchsmuster von Wörtern zutage gefördert, die bisher und ohne diese Techniken zwar vermutet, aber nicht verifiziert werden konnten. Patterns spielen zudem bei der Bestimmung von Bedeutung eine große Rolle. Sinclair (1991) war einer der ersten, die behauptet haben, dass jede Bedeutung eines Wortes mit einem Pattern korreliert. Auch Kollokationen konnten in diesem Zusammenhang zu einem wichtigen Untersuchungsgegenstand avancieren.

---

<sup>4</sup> Vor allem bei Korpora, die vorwiegend aus Zeitungstexten bestehen, herrscht keine Einigkeit darüber, wie viele Sprecher produktiv tatsächlich beteiligt sind. Rezeptiv ist die Zahl der Sprecher aber sehr groß.

<sup>5</sup> Dies ist jedoch nicht zwingend die Folge. Die Dichotomie zwischen Synchronie und Diachronie kann vom Interpreten aufgrund der Daten wieder hergestellt werden.

<sup>6</sup> Jedoch ist noch umstritten, wie groß ein Korpus sein muss, um verlässliche Aussagen zu erlauben. Die Frage stellt sich ganz besonders auch bezüglich der korpusbasierten Analyse von Idiomen. Dazu cf. Geyken et al. (2004).

## 2 Korpusbasierte Analyse der Semantik deutscher Idiome

Im Folgenden wird der korpusbasierte Ansatz anhand einer Untersuchung von Idiomen des Deutschen illustriert. Die Analyse beruht auf einem Korpus von ca. 970 Millionen tokens, welches sich über das gesamte 20. Jahrhundert erstreckt. Die Gewichtung – für die Gewinnung größerer Datenmengen für Idiome – liegt auf den 1990er Jahren. Es dominieren Zeitungstexte (cf. Geyken, erscheint).

Die Idiome wurden mit Hilfe von eigens entwickelten Suchanfragen extrahiert. Dieses Verfahren ermöglicht es, einerseits nicht überzugenerieren, andererseits aber auch interessante Daten nicht zu verlieren (cf. Herold, erscheint). Jedes Idiom wurde mit einem Kontext von drei Sätzen vor und drei Sätzen nach dem Satz, der das Idiom enthielt, extrahiert und betrachtet. Wenn für die Interpretation mehr Kontext benötigt wurde, konnte dieser geliefert werden (cf. Neumann/Körner, erscheint). Alle Belege wurden berücksichtigt und manuell annotiert.

Wie schon ausgeführt, ist korpusgestützte Analyse notwendigerweise kontextbasiert und kontextabhängig. Besonders für semantische Fragen bedeutet dies, dass der korpusbasierte Ansatz die Bedeutung von lexikalischen Einheiten zwingend als **Bedeutung im Gebrauch** oder **Bedeutung als Gebrauch** behandelt. Bedeutung ist Bedeutung im Kontext, d. h. in der unmittelbaren, lokalen syntaktischen Umgebung des betreffenden lexikalischen Elementes im Satz, und im Kontext, d. h. im Äußerungskontext. Aus einer solchen Sicht ergeben sich folgende Fragen:

- 1) Was lässt sich über die Bedeutung von Idiomen im Gebrauch beobachten?
- 2) Wie interagiert die Bedeutung des Idioms mit dem Kontext?
- 3) Wie kann die Bedeutung von Idiomen angemessen repräsentiert werden?

Idiome werden in der Regel über ihre Bedeutung definiert, d. h. als Ausdrücke, deren Bedeutung sich nicht regulär aus der Bedeutung ihrer Komponenten ergibt. Regulär bedeutet dabei nach dem Kompositionalitätsprinzip, das in einer Formulierung besagt: "Die Bedeutung eines komplexen Ausdrucks ergibt sich eindeutig aus der lexikalischen Bedeutung seiner Komponenten, aus deren grammatischer Bedeutung und aus seiner syntaktischen Struktur" (cf. Löbner 2002: 20).

Dieses Prinzip, das schon seit den Anfängen der traditionellen Grammatik impliziert wird und in formal-semantischen Ansätzen zum Grundsatz wurde, setzt voraus, dass Grundausrücke (Wörter) mit ihren Bedeutungen zu komplexen Einheiten (zunächst zu Phrasen und dann zu Sätzen) verbunden werden, deren Bedeutung sich aus den Bedeutungen der Grundausrücke aufbaut. Dabei ist die syntaktische Verknüpfung entscheidend, denn *Karl liebt Maria* bedeutet etwas anderes als *Maria liebt Karl*. In diesem Modell verlaufen syntaktische Komposition und semantische Komposition (Interpretation) parallel (Rule-to-Rule-Hypothesis). Schematisch kann dies folgendermaßen dargestellt werden:

<u>Grammatik</u>	<u>Semantik</u>
<b>komplexer Ausdruck</b> (-> Syntax)	<b>kompositionale Bedeutung des komplexen Ausdrucks</b>
<i>Kombination der Ausdrücke</i>	<i>Komposition der Bedeutungen</i>
Grundausdrücke in einer bestimmten Form (-> Morphologie)	kompositionale Bedeutungen der Grundausdrücke in dieser Form
<i>Wahl einer grammatischen Form</i>	<i>Hinzufügen der grammatischen Bedeutung</i>
Grundausdrücke (-> Phonologie)	lexikalische Bedeutung der Grundausdrücke

Abb. 1: Kompositionalitätsprozess (nach Löbner 2002: 19)

Dieses Verfahren ist ein Bottom-up-Prozess, d. h. die Komposition verläuft von den einfachen hin zu den komplexeren Ausdrücken.

Idiome "verletzen" das Kompositionalitätsprinzip, weil sich ihre Gesamtbedeutung nicht aus den Bedeutungen der Wörter ergibt, aus denen sie bestehen. Sie werden demzufolge als nicht-kompositional bezeichnet. Diese Charakterisierung von Idiomen gründet auf der Beziehung zwischen den Idiomkomponenten. Für VP-Idiome, die aus einem Verb und mindestens einer NP (i. d. R. in Objektfunktion) oder PP bestehen, bedeutet dies, dass die VP eine Gesamtbedeutung hat, die sich nicht kompositional aus der Bedeutung des Verbs und seiner Ergänzungen konstruiert. So ist z. B. im Idiom *ins Gras beißen*, das die konventionalisierte Bedeutung 'sterben' hat, allein die Beziehung zwischen seinen Komponenten *beißen* und *ins Gras* für die semantische Nicht-Kompositionalität zuständig. Das Idiom nimmt jedoch als Einheit am Kompositionsprozess des Satzes teil, d. h. die Komposition setzt sich (mindestens) durch das Hinzukommen eines Subjekts fort. Das heißt, Idiome nehmen in zweifacher Weise am Kompositionsprozess teil: erstens, in ihrer internen Struktur als Ausdruck und zweitens, extern in ihrer Beziehung zu anderen Konstituenten des Satzes, wie monolexikalische Verben.<sup>7</sup> Die Kompositionalität von Idiomen ist also eine zweifache; ich werde sie hier als Typ A und B bezeichnen:

<sup>7</sup> In der deutschen Phraseologieforschung ist von interner und externer Valenz bei VP-Idiomen die Rede.

	<i>jmd.</i>	<span style="border: 1px solid black; padding: 2px;"><i>beißt ins Gras</i></span>
	Subjekt-NP	VP-Idiom
Kompositionalität:	Typ B	Typ A

Abb. 2: Typen der Kompositionalität bei VP-Idiomen

Die Definition als nicht-kompositionales Phänomen betrifft die Kompositionalität vom Typ A, als lexikalische Einheit jedoch verbindet sich das Idiom mit anderen Konstituenten (hier mit einer Subjekt-NP) und nimmt an der Kompositionalität vom Typ B teil.<sup>8</sup>

Verfolgt man einen korpusbasierten Ansatz, so ist der eben geschilderte Prozess der Komposition für den Interpreten umgekehrt, man verfährt nun "top-down" statt "bottom-up".<sup>9</sup> Nicht die einfachen Ausdrücke sind primär, sondern der Text und die Interpretation. Die Interpretation erfolgt im Text, und die Bedeutung einfacher und komplexer Ausdrücke ist in einem größeren Zusammenhang verankert. Daher ist Bedeutung in Korpora zwingend kontextualisierte Bedeutung. Was man zunächst bei Idiomen sieht, ist Kompositionalität vom Typ B, d. h. man versteht das Idiom als Einheit in einem Satz und in einem größeren Kontext. Aus diesem Grund werde ich die Untersuchung der Bedeutung von Idiomen aus dieser Perspektive beginnen. Dabei ist mein Ziel, über die Bedeutung von Idiomen im Kontext ihre Eigenschaften als komplexe Ausdrücke zu erhellen, oder anders gesagt: von der Kompositionalität des Typs B zu der des Typs A zu gelangen und eventuell zu erklären.

### 3 Fallstudie

Eines der in der Literatur zu Idiomen häufig angeführten Idiome, *ins Gras beißen*, soll hier zunächst anhand der Korpusdaten illustriert werden. Das *Wörterbuch der deutschen Idiomatik* (Duden Band 11) verzeichnet als Bedeutung des Idioms '[eines gewaltsamen Todes] sterben'. Das *Wörterbuch der Deutschen Gegenwartssprache (WDG)* gibt 'sterben' als einzige Bedeutung an. *Duden Universalwörterbuch* und das *Deutsche Wörterbuch* von Hermann Paul paraphrasieren das Idiom ebenfalls mit 'sterben'.

---

<sup>8</sup> Da Idiome Einheiten des Lexikons sind, stellt sich die Frage, ob es überhaupt sinnvoll ist, sie mit "freien" Konstruktionen zu kontrastieren. Meiner Meinung nach macht dies Sinn, denn obwohl Idiome Wortschatzeinheiten sind, sind sie diskontinuierliche phrasale Einheiten. Gerade Korpusstudien zeigen, dass sie bei ihrer Kontextualisierung im Satz den syntaktischen Regeln des Deutschen folgen, z. B. bezüglich Wortstellung, Modifizierung durch Adjektive etc.

<sup>9</sup> Löbner (2002: 19, Fn. 13) merkt an, dass ein Top-down-Prozess auch in Gang gesetzt wird, wenn man in einem gegebenen Kontext einem unbekanntem Wort begegnet. In solchen Situationen kann man häufig die Bedeutung des Wortes aus der Gesamtbedeutung des Satzes erschließen. Dies ist aber nur dann möglich, wenn sich der übrige Satz kompositional, also bottom-up, interpretieren lässt. Löbner zufolge widerlegen solche Fälle nicht, sondern bestätigen, dass sich die Satzbedeutung bottom-up aus den lexikalischen Bedeutungen ableitet.

Im Korpus wurden 200 Belege gefunden. Davon wurde 147 die Bedeutung 'sterben' zugeordnet (siehe jedoch die Diskussion des Belegs (15) weiter unten). Typische Beispiele (unter Auslassung von Wortspielen und Modifikationen) sind (1)–(4):

- (1) *"Jeder zweite von uns benötigt, bevor er **ins Gras beißt**, mindestens eine Blutkonserve, spenden tut aber nur jeder hundertste."*  
(die tageszeitung 25.05.1991, S. 44)
- (2) *Wenn er ein Soldat wird, muß er **ins Gras beißen**, das ist klar.*  
(Brecht, Bertolt: Mutter Courage und ihre Kinder, o.O. 1939)
- (3) *Bis dahin aber werden wohl noch ein paar Drogensüchtige im Knast wegen Aids **ins Gras beißen** müssen.*  
(die tageszeitung 05.10.1994, S. 17)
- (4) *Für Pelzstolen mußte mehr als ein Fuchselein **ins Gras beißen** [...]*  
(Süddeutsche Zeitung 31.07.1999, S. M/F/B7)

In dieser Bedeutung des Idioms ist das Subjekt belebt, d. h. PERSON (in den meisten Fällen) oder TIER (und in einem Beleg erscheint *Bäume* als Subjekt, also PFLANZE). Das Subjekt referiert in der häufigsten und typischen Bedeutung ('sterben' = 'aufhören zu leben') auf ein BELEBTES Wesen. Der Idiomgebrauch in den übrigen 53 Belegen entspricht nicht dieser Bedeutung. Ich werde diese Belege zunächst unter einem Label 'nicht-sterben' zusammenfassen und eine schrittweise Differenzierung vornehmen:

- (5) *1987/88 **bissen** bei den Saarländern Uwe Klimaschefski und Gert Schwickert **ins Gras** – Homburg stieg ab.*  
(Süddeutsche Zeitung 03.04.1996, S. 24)
- (6) *'Einer wird **ins Gras beißen**', sagt Bender, 'ich werde es nicht sein.*  
(Süddeutsche Zeitung 14.02.1997, S. 44)
- (7) *Der Wimbledonssieger von 1991 kann frühestens im Viertelfinale auf einen gesetzten Spieler treffen, nachdem der Schwede Jonas Björkman und French Open-Sieger Gustavo Kuerten bereits **ins Gras bissen**.*  
(Frankfurter Rundschau 25.06.1997, S. 17)

Aufgrund unseres Weltwissens ist uns klar, dass die Personen in (7) zum Zeitpunkt der Äußerung noch am Leben sind. Dass wir es hier nicht mit der Bedeutung 'sterben' zu tun haben, wird auch durch den Hauptsatz nahe gelegt, der den Äußerungskontext festlegt: Es handelt sich um ein Tennisturnier, bei dem die genannten Spieler ausgeschieden sind, d. h. das Turnier ist für sie beendet. (6) kann ohne Kontext – so wie es hier angeführt wird – 'sterben' bedeuten. Aber das ist nicht der Fall. Der Satz ist in einem fußballerischen Kontext geäußert worden, wobei in einer Mannschaft zwei Spieler (Bender und Pele) um eine Position kämpfen. Bender sagt, dass einer von ihnen seine Position verlieren wird. (5) bezieht sich auf

eine Fußballmannschaft (*Homburg*). Der Text zählt alle Trainer auf, die von dieser Mannschaft entlassen wurden. Diese haben also ihren Job verloren.

In der Einleitung wurde der Begriff des Pattern eingeführt. Durch die Korpusanalyse lassen sich zunächst zwei Pattern erkennen, die quantitativ überwiegen: Das erste wurde in (1)–(3) gezeigt und es bezieht sich auf die Bedeutung 'sterben' ('aufhören zu leben'). Es wird in (i) dargestellt:

- (i) PERSON<sub>subj</sub> + IDIOM-VP 'sterben'

Wenn man nun die restlichen 53 Belege der 'nicht-sterben'-Gruppe betrachtet, so lassen sich auch dort Pattern feststellen. Darin dominiert hinsichtlich der Frequenz das im folgenden Beleg illustrierte Pattern:

- (8) *Hunderte von Kleinunternehmen in dieser Branche **beißen** jährlich **ins Gras**.*  
(Süddeutsche Zeitung 06.04.1994, S. 28)

Dieser Gebrauch findet sich in 16 Belegen.<sup>10</sup> In diesem Fall ist die Subjekt-NP ein UNTERNEHMEN (z. B. *Kleinunternehmen, Gesellschaften, Mittelstandsbetriebe* usw.). Die Bedeutung hier ist 'schließen', 'zugrunde gehen', 'pleite gehen'. Das Pattern ist:

- (ii) UNTERNEHMEN<sub>SUBJ</sub> + IDIOM-VP (*Kontext: Wirtschaft*)  
'schließen', 'zugrunde gehen', 'pleite gehen'

Ein weiteres Pattern konnte durch 15 Belege im Korpus nachgewiesen werden, z. B. wie in folgendem:

- (9) (*Von den Zweiten der Gruppen 1, 2 und 4 muß der Punktschlechteste **ins Gras beißen**.*)  
(die tageszeitung 28.04.1989, S. 10)

In den Belegen dieser Gruppe ist die Bedeutung variabel, 'ausscheiden', 'verlieren' oder 'absteigen'. Alle Fälle können auch allgemeiner paraphrasiert werden als 'das Nachsehen haben', 'leer ausgehen', 'scheitern'. Das Pattern kann folgendermaßen angesetzt werden:

- (iii) MANNSCHAFT<sub>SUBJ</sub> + IDIOM-VP (*Kontext: Sport*)  
'ausscheiden', 'verlieren', 'absteigen'

---

<sup>10</sup> Der Übersichtlichkeit halber werden zu den Belegen auch diejenigen gezählt, in denen das Subjekt des Satzes ein Pronomen oder ein anderer Ausdruck ist, dessen Antezedens aber ein Nomen ist, das auf ein Unternehmen referiert. 3 Belege finden sich in Titeln.

Bisher konnten – wie gezeigt – die beiden Bedeutungen 'sterben' und 'nicht-sterben' aufgrund der mit ihnen assoziierten Pattern auseinandergelassen werden. Die Daten gestalten sich jedoch etwas komplizierter, denn es verbleiben weitere 15 Belege der 'nicht-sterben'-Gruppe, in denen das Subjekt auf eine Person oder eine Gruppe von Personen referiert.<sup>11</sup> Wir haben zwar Pattern (i) PERSON<sub>SUBJ</sub> + IDIOM-VP, das für 'sterben' steht, aber man versteht, dass im Text nicht die Bedeutung 'sterben' vorliegt. Beispiele dafür waren (5)–(7). Es gibt zwei Belege mit der Bedeutung 'gefeuert werden', 'den Job verlieren', z. B. (5). Drei weitere Belege erscheinen mit der Bedeutung 'seine Position in der Mannschaft verlieren', z. B. (6). Fünf Belege, von denen zwei identisch sind, illustrieren einen Gebrauch wie in (7), nämlich 'verlieren', 'in einem Turnier ausscheiden', 'scheitern'. Ein weiterer Beleg zeigt den Gebrauch im Sinne von 'den Job verlieren', diesmal bei einem Fernsehsender. Auf eine Gruppe von Personen referiert das Subjekt im folgenden Beleg:

- (10) *Bei den Behörden ist man nicht bereit, mehr zu zahlen, aber ebenso beharren auch die Lehrer auf ihrer Position, [...] "Ins Gras beißen", so Hobunn, müssen derweil zu seinem Bedauern die TeilnehmerInnen der Aufbaukurse. Nachdem sie jetzt schon drei Semester lang geackert und insgesamt 660 Mark investiert haben, können sie ihr Lernziel, das bundesweit anerkannte VHS-Zertifikat, nun nicht mehr erreichen, wodurch sie sich in ihrem beruflichen Werdegang blockiert sehen.*

(die tageszeitung 08.09.1989, S. 22)

In (10) ist gemeint, dass die Teilnehmer der Kurse Opfer von Sparmaßnahmen sind. Meine Unterstreichungen im Beleg verdeutlichen, wie diese Bedeutung im Textverlauf weiter elaboriert und expliziert wird.

Es erscheinen zwei weitere Belege mit der Bedeutung 'scheitern' außerhalb von Sport-Kontexten. Die Subjekt-NP ist in beiden Fällen eine Person:

- (11) *Steinbrück mein Name [...] akademisch ins Gras gebissen, abberufen worden, [...]*  
(Burger 1982)

- (12) *Als zuvor [...] Camillo Castiglioni, mit allen seinen Unternehmungen ins Gras beißen mußte, [...]*  
(Kölnische Zeitung (1. Morgenblatt) 01.01.1926, S. 1)

Diese Belege instantiieren Pattern (i), PERSON<sub>SUBJ</sub> + IDIOM-VP, aber von 'sterben' ist hier nicht die Rede. Um diese Interpretation zu vermeiden, ist im Kontext eine zusätzliche Markierung notwendig. In (11) modifiziert das Adverb *akademisch* die VP und spezifiziert die Domäne, auf die der Sachverhalt des Verbs zutrifft. Im Kontext wird dies weiter elaboriert (*abberufen worden*). In (12) verhindert die PP *mit allen seinen Unternehmungen*, dass es zur Interpretation 'sterben' kommt. Eine etwas andere Situation liegt im folgenden Beleg vor:

<sup>11</sup> Unternehmen und Mannschaften sind zwar auch Gruppen von Personen, sie werden aber in Pattern (ii) und (iii) nicht als ein Kollektiv von Personen behandelt, sondern in ihrer institutionellen Funktion als Einrichtungen.

- (13) *"Mit Geld werden wir in Schach gehalten, denn unser gesamtes Leben müssen wir versuchen, so viel davon zu erwerben wie möglich, da wir ansonsten **ins Gras beißen**. "So schrieb Johnny in derselben Hausarbeit. Er und seine Freunde aus der 10. Klasse [...] stünden vor der Wahl, sich für einen Beruf zu entscheiden oder aufs Abitur hinzuarbeiten. "Wir fragen uns bloß noch, wie kriegen wir das hin, nicht so sehr auf die Schnauze zu fallen, arbeitslos zu werden."*

(Frankfurter Rundschau 03.02.2000, S. 30)

In (13) ist die Bedeutung des Idioms nicht eindeutig. Nach dem Kontext könnte es 'scheitern' bedeuten (vgl. *arbeitslos zu werden*), aber es könnte auch 'zugrunde gehen' oder – was weniger plausibel ist – 'sterben' bedeuten. Hier liegt ein typischer Fall von Vagheit vor, d. h. verschiedene Bedeutungen überlappen sich, sie sind gleichzeitig präsent. Es ist eines der wesentlichen Erkenntnisse von korpusbasierter lexikalischer Analyse, dass sich bei polysemen Lexemen die verschiedenen Bedeutungen im Gebrauch überschneiden können. Das gilt ebenso für Idiome.

Bisher wurden auf der Grundlage von Korpusdaten drei wesentliche Pattern des Idioms identifiziert. Pattern (i) überwiegt quantitativ und ist mit der Bedeutung 'sterben' assoziiert. Pattern (ii) und (iii) sind insgesamt weniger häufig, aber signifikant.<sup>12</sup> Sie können jeweils mit der Bedeutung 'verlieren' bzw. 'zugrunde gehen' verbunden werden. Die Bedeutung 'sterben' ist von allen Bedeutungen die am wenigsten durch zusätzliche syntaktische oder lexikalische Mittel markierte, sie ist als Kernbedeutung des Idioms konventionalisiert und erscheint sozusagen als Default-Bedeutung im Fall einer Subjekt-NP mit dem Merkmal BELEBT.<sup>13</sup> Obwohl die Subjekt-NP in den meisten Fällen das Merkmal PERSON hat, ist die Bedeutung 'sterben' nur im Fall einer Subjekt-NP mit dem Merkmal TIER oder PFLANZE eindeutig.

Die Bedeutungen 'verlieren', 'zugrunde gehen', etc. werden dadurch unterschieden, dass sie typischerweise mit anderen Subjekten auftreten, nämlich MANNSCHAFT und UNTERNEHMEN und evtl. weiteren Spezifizierungen durch den Kontext. Die beiden bisher ermittelten Bedeutungen ('sterben' und 'nicht-sterben') sind also auf diese Weise distinkt. Interessant ist die Vermischung von Pattern der einen Bedeutung mit Pattern der zweiten Bedeutung. Dies wurde für PERSON<sub>SUBJ</sub> illustriert. Im Sport-Kontext erfolgt die Interpretation 'ein Spiel (bei Einzelsport)/seine Position (im Team) verlieren', 'ausscheiden' oder aber 'den Job verlieren'. Diese Belege können mit Rückgriff auf das stark vertretene Pattern MANNSCHAFT<sub>SUBJ</sub> erklärt werden. Außerhalb des Sport-Kontextes findet sich mit PERSON<sub>SUBJ</sub> die Bedeutung 'scheitern'. Die Belege mit diesem Pattern sind niedrig frequent. Als Ursache dafür nehme ich Folgendes an: Da PERSON<sub>SUBJ</sub> stark an die Bedeutung 'sterben' gekoppelt ist, wird eine Verbindung mit einer anderen Bedeutung eher vermieden. Wenn sie dennoch erfolgt, dann ist eine zusätzliche Markierung im Kontext wie in (11) und (12) notwendig, oder es erfolgt eine Umschreibung

<sup>12</sup> Das Erscheinen von rund 30 Belegen mit diesen beiden Pattern, die in den Kontexten Sport und Wirtschaft erscheinen, ist auf die Beschaffenheit des Korpus zurückzuführen, in dem Zeitungstexte überwiegen. Dies ändert jedoch nichts an der Tatsache, dass das Idiom außer seiner kodifizierten Bedeutung neue Bedeutungen erwerben kann. Es geht uns hier darum, diese zu erklären.

<sup>13</sup> Sie ist auch die spontane Interpretation von einem Satz ohne Kontext; vgl. auch (6), das aus seinem Kontext entfernt wurde.

durch synonyme Ausdrücke und Elaboration im weiteren Kontext wie in (10). Fehlt eine solche zusätzliche Markierung, was selten ist, so bleibt die Bedeutung vage wie in (13).

Die bisher illustrierten Verwendungsweisen verdeutlichen, dass jede Bedeutung typischerweise mit einem Pattern assoziiert ist. Wir haben eine Kernbedeutung ermittelt, 'sterben', und eine weitere Gruppe von Bedeutungen, die wir – zugegebenermaßen etwas unelegant – als 'nicht-sterben' bezeichnet haben. Innerhalb dieser Gruppe haben wir eine Reihe von weiteren Bedeutungen ermitteln können. Darin überwiegen Bedeutungen, die aus den beiden dominanten Pattern (ii) und (iii) resultieren. Aus den Beispielen geht hervor, dass sich die Pattern der zweiten Bedeutung ('nicht-sterben') von den Pattern der ersten Bedeutung ('sterben') "fernhalten". Sobald sich ein Pattern Merkmale eines anderen Pattern zu Eigen macht, muss der Unterschied durch zusätzliche Markierung verdeutlicht werden, in den meisten Fällen sogar durch multiple Markierung in der unmittelbaren syntaktischen Umgebung und im weiteren Kontext. Dadurch entsteht Redundanz, die aber vermutlich notwendig ist, wenn nicht-typische oder marginale Verwendungsweisen einer lexikalischen Einheit vorliegen und Vagheit ausgeschlossen werden soll.

Die Bedeutung 'sterben' unterscheidet sich vom Cluster der Bedeutungen 'nicht-sterben' noch durch eine weitere Eigenschaft: Nur Letztere werden in den Daten durch Adjektive modifiziert. Die Adjektive modifizieren syntaktisch das Nomen *Gras*, aber semantisch die gesamte VP. Ihre Funktion ist die Spezifizierung des Äußerungskontextes. Vgl.:

- (14) *Immerhin überlebte Tony die Salve, während Oskar, wie wir wissen, **ins politische Gras biß**.*  
(Frankfurter Rundschau 12.05.1999, S. 22)

*Ins politische Gras beißen* meint 'politisch ins Gras beißen' (d. h. 'politisch scheitern'). Dadurch wird, wie schon in (11) und (12), die Bedeutung 'sterben' ausgeschlossen. Dies ist eine weitere Strategie, um die beiden Bedeutungen voneinander zu absentieren und die Bedeutung 'sterben' zu blockieren, die in (14) durch die Verwendung des Pattern (i) PERSON<sub>SUBJ</sub> entstünde.

Zu Beginn der Fallstudie wurde darauf hingewiesen, dass 147 von 200 Belegen die Bedeutung 'sterben' zugewiesen werden konnte. Darunter befinden sich drei Belege, für die 'sterben' nicht die angemessene Bedeutungsangabe ist, sondern für die sehr allgemein 'aufhören zu existieren' paraphrasiert werden muss. Die Subjekt-NP ist in einem Fall *der Staat*, im anderen eine *Kolumne*. Entsprechend ist auch der nächste Beleg zu interpretieren, in dem die Subjekt-NP UNBELEBT ist:

- (15) *Die Kritische Theorie **hat ins Gras gebissen**, weil Habermas sich nicht an die Theorie [...] gehalten hat, welche er doch selbst entworfen hat.*

(DIE ZEIT 07.10.1999, S. 62)

Aus welchem Grund werden diese Belege der 'sterben'-Bedeutung zugerechnet und nicht der zweiten Gruppe? Wir haben es hier immerhin mit einer Subjekt-NP zu tun, die etwas UNBELEBTES bezeichnet. Ist das nicht ein Fall, der mit Pattern (ii) in Verbindung gebracht

werden kann oder muss? Die Antwort auf diese Frage ist nicht einfach und verdeutlicht zudem eine wichtige Eigenschaft der internen Struktur von Lexemen. *Staat* in Subjektposition kann mit UNTERNEHMEN in Pattern (ii) korreliert werden. Dieses könnte um das Merkmal INSTITUTION erweitert werden. Es handelt sich in beiden Fällen um eine metonymische Extension des Konzepts PERSON. Gleichzeitig ist der Staat etwas Abstrakteres als die Subjekte, die sich in den Belegen von Pattern (ii) finden. Auf diese Weise wird die interne Struktur des Clusters deutlich: Einige Verwendungsweisen gehören enger zusammen und sind durch ein Merkmal *x* mit der einen, mit einem Merkmal *y* mit anderer Verwendungsweise verbunden. So entsteht ein enges Geflecht von Beziehungen innerhalb eines Lexems. Auch die Belege mit *Theorie* und *Kolumne* in Subjektposition wurden auf diese Weise interpretiert. Meine anfängliche Annahme, dass diese Beispiele zur Bedeutung 'sterben' zu rechnen sind, ist zu revidieren: Obwohl Bedeutungen aufgrund einer Reihe von interagierenden Faktoren (semantische Klassen der Argumente, unmittelbare syntaktische Umgebung, weiterer Kontext etc.) signifikant häufig erkannt und voneinander getrennt werden können, gibt es immer auch marginale Verwendungsweisen, die "zwischen" den etablierten Bedeutungen liegen und darauf hindeuten, dass die Grenzen zwischen den Bedeutungen fließend sind. Das wird durch (15) verdeutlicht, ebenso wie durch (13) illustriert wurde, dass Bedeutungen sich überschneiden können. Die Zuordnung zur einen oder anderen Bedeutung – sofern sie zwingend ist, z. B. in der Lexikografie – ist nur durch Festlegung zu entscheiden.

Die genannten Bedeutungen, die mithilfe der Korpusrecherche ermittelt wurden, können als Netzwerk (Abb. 3) dargestellt werden. Darin sind nur die bisher attestierten Bedeutungen enthalten. Potenzielle Bedeutungserweiterungen sind auf dieser Grundlage erklärbar.

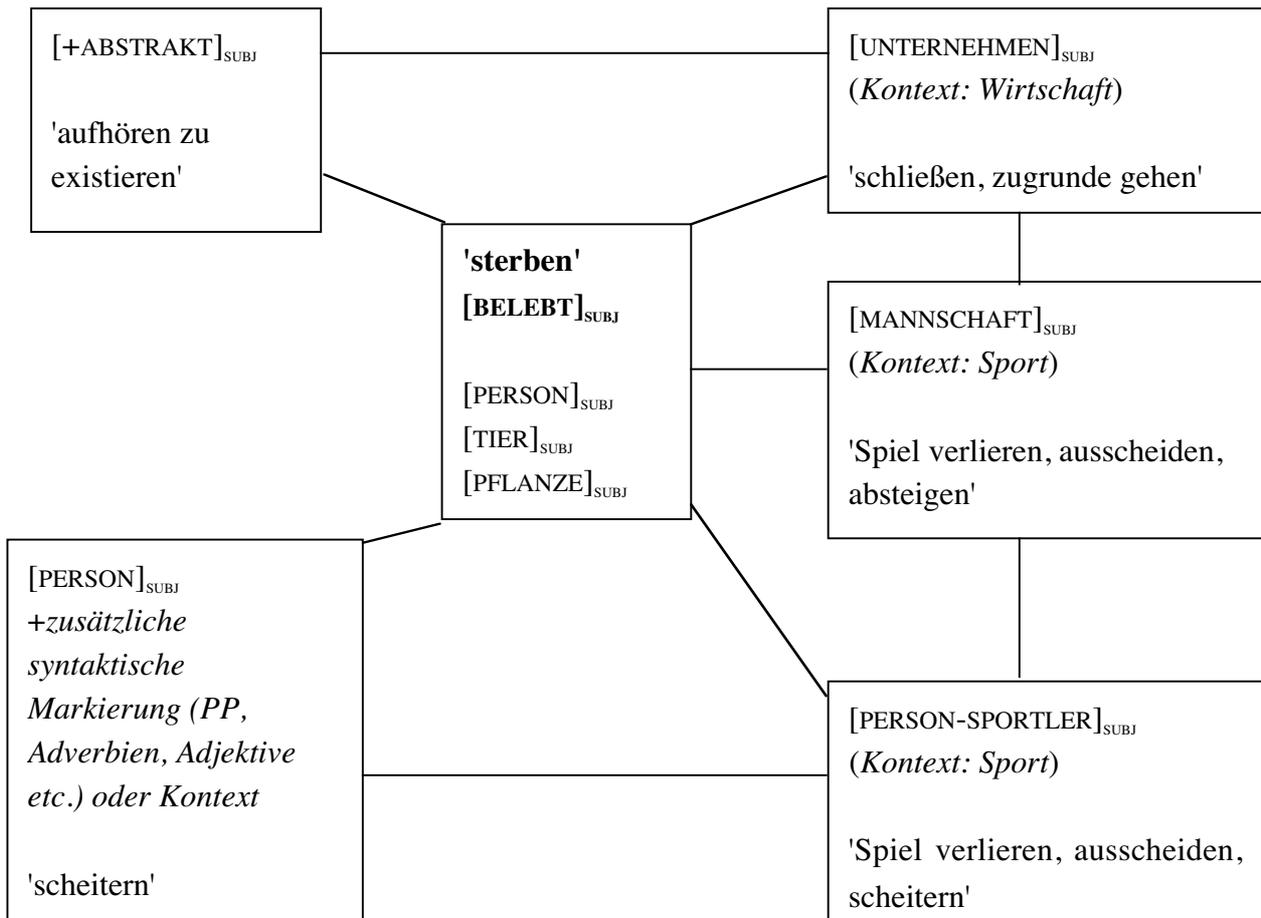


Abb. 3. Bedeutungsbeziehungen des Idioms *ins Gras beißen*.

Die Verbindung der Kernbedeutung 'sterben', 'aufhören zu leben' zu 'aufhören zu existieren' ist einfach. Das gilt auch für die Bedeutung 'zugrunde gehen' für UNTERNEHMEN – diese schließen entweder oder sind finanziell ruiniert – was einer Zerstörung gleichkommt und eine Art zu "sterben", eben die typische Art für UNTERNEHMEN ist. Die Verbindung zu 'sterben' ist deutlich erkennbar. Die Bedeutung 'aus einem Turnier ausscheiden' ist bereits 1928 belegt, und zwar zunächst nur für den Mannschaftssport. Auch das kann mit 'sterben' verbunden werden, da das Ausscheiden aus einem Turnier 'nicht mehr präsent zu sein' heißt. Diese Verwendung wird dann offenbar auf Einzelsportler ausgeweitet. Ähnlich geschieht es mit der Bedeutung 'absteigen': Eine Mannschaft, die absteigt, ist nicht mehr in der ursprünglichen Liga präsent. Um auszuschneiden oder abzusteigen muss man mehrmals verlieren – die Bedeutung des Idioms erfährt hier eine Spezialisierung (Einschränkung) der Bedeutung auf den kausalen Hintergrund des 'Sterbens'. Aus häufigen Niederlagen entsteht vermutlich eine Verbindung zum Konzept des 'Scheiterns', das sich ebenfalls auf andere Kontexte ausweitet. Schließlich ist die Bedeutung 'den Job oder die Position verlieren' dadurch zu erklären, dass die betreffende Person nicht mehr gebraucht wird, unwichtig, nicht existent ist und gleichzeitig in ihrer Aufgabe gescheitert ist. Hier kann man sogar davon ausgehen, dass beide semantische Merkmale zu dieser Bedeutungsextension führen.

#### 4 Zusammenfassung

Ich habe alle Belege des Idioms *ins Gras beißen* in einem 970 Millionen tokens umfassenden Korpus untersucht. Dabei konnte zunächst ein wesentliches Pattern ermittelt werden, das mit der Bedeutung 'sterben' verbunden ist. Außerdem zeigten sich zwei weitere quantitativ relativ stark vertretene Pattern, die ganz klar nicht mit physischem Tod verbunden sind ('nicht-sterben'), sondern als 'verlieren', 'zugrunde gehen' usw. paraphrasiert wurden. Die beiden Pattern unterscheiden sich in der semantischen Klasse ihrer Subjekte (BELEBT vs. MANNSCHAFT/UNTERNEHMEN als Organisationen bzw. Einrichtungen). Die zweite Gruppe ist zusätzlich durch den Kontext als einer bestimmten Domäne zugehörig markiert (Sport, Wirtschaft), während die erste Bedeutung häufig unmarkiert ist.<sup>14</sup> Die Bedeutung 'sterben' tritt drei Mal so häufig auf wie die Bedeutung 'nicht-sterben'. Sie repräsentiert die konventionalisierte, kodifizierte und ältere Bedeutung des Idioms. Die Bedeutung 'nicht-sterben' datiert im Korpus aus dem Jahre 1926 und ist in Zeitungstexten der letzten zwei Jahrzehnte vertreten. Sie kann als ein Cluster von mehreren Bedeutungen angesehen werden, die erstens durch die Kernbedeutung und zweitens durch den Kontext determiniert werden.

Deutlich wurde, dass bei einer korpusbasierten Analyse die Bedeutung eines Idioms im Kontext variabel sein kann und nicht auf die konventionalisierte und kodifizierte Bedeutung beschränkt sein muss. Was bedeutet diese Erkenntnis nun für die Kompositionalität vom Typ B? Daraus folgt, dass das Idiom als lexikalische Einheit an der Komposition der Satzbedeutung teilnimmt. Der Beitrag, den es dabei leistet, kann nicht unabhängig von seiner Umgebung betrachtet werden. Dieser Sachverhalt ist sehr komplex und seine Erklärung ist eine große Herausforderung für die Semantik. Wir sind noch weit davon entfernt, diesen Prozess erklären zu können, aber Korpora können dazu beitragen, ihn für eine große Anzahl von lexikalischen Einheiten zu beschreiben und auf dieser Grundlage nach systematischen Prozessen zu fragen. Durch Korpora wird der Kompositionsprozess von der Perspektive des komplexen Ausdrucks hin zu den Grundausdrücken, von der Textbedeutung zur lexikalischen Bedeutung betrachtet. Dabei zeigt sich, dass sich lexikalische Einheiten semantisch durch Flexibilität auszeichnen, die ihre Einbettung in neue Kontexte erleichtert oder sogar ermöglicht.<sup>15</sup>

Wie lässt sich auf dem Hintergrund dieser Ergebnisse die Bedeutung von Idiomen angemessen beschreiben? Die lexikalische Bedeutung von Idiomen weist ebenso wie die von "Einzellexemen" eine interne Struktur auf, die ähnlich wie eine Kategorie im Sinne der Prototypentheorie aufgefasst werden kann. Es kann eine Kernbedeutung ermittelt werden, die sich im Zentrum befindet. Diese ist die häufigste, statistisch signifikante Bedeutung und in diesem Beispiel auch die historisch ältere. Die anderen Bedeutungen ergeben sich aus der Kernbedeutung durch Profilierung oder Unterdrückung semantischer Merkmale. Diese Bedeutungen teilen mehr oder weniger Merkmale untereinander und können als Cluster von

---

<sup>14</sup> Auch für diese Bedeutung finden sich im Korpus wiederkehrende Kontexte: in literarischen Texten wird das Idiom häufig verwendet, wenn von *Krieg* die Rede ist.

<sup>15</sup> Hanks (persönliche Mitteilung) spricht vom Bedeutungspotenzial ("meaning potential") lexikalischer Ausdrücke.

Bedeutungen, wie Familienähnlichkeiten, angesehen werden (cf. Wittgenstein 1953; Lakoff 1987). Dies wurde für das Idiom *jmd. beißt ins Gras* in Abb. 3 illustriert und ausführlich erläutert. Die Verbindung dieser Bedeutungen zur Kompositionalität vom Typ A ist nicht mehr transparent.<sup>16</sup> Die Bedeutungsextensionen basieren in diesem Fall nicht auf dem zugrunde liegenden Bild, sondern auf der lexikalischen Bedeutung des Idioms.<sup>17</sup>

Aus diesen Ausführungen ist hoffentlich deutlich geworden, dass eine korpusbasierte Herangehensweise eine neue Sichtweise auf Idiome ermöglicht. Insbesondere bezüglich der Semantik wird durch Korpusdaten die Behandlung von Idiomen als Lexikoneinheiten ermöglicht. Statt den Blick auf die Komposition der Bedeutung innerhalb des Idioms zu richten, habe ich mich auf die Verwendung der Idiome als Lexeme im engen syntaktischen und im weiteren textuellen Kontext konzentriert und gezeigt, dass Idiome als Lexeme eine komplexe Bedeutungsstruktur und eine Bedeutungsvielfalt aufweisen, die auf der konventionalisierten Bedeutung basiert, diese jedoch überschreitet.

### Literaturangaben

- Aijmer, Karin/Altenberg, Bengt (eds.) (1991): *English Corpus Linguistics*. London.
- Atkins, B. T. Sue/Zampolli, Antonio (eds.) (1994): *Computational Approaches to the Lexicon*. Oxford.
- Baker, Mona/Francis, Gill/Tognini-Bonelli, Elena (eds.) (1993): *Text and Technology. In Honour of John Sinclair*. Amsterdam.
- Biber, Douglas/Conrad, Susan/Reppen, Randi (1998): *Corpus Linguistics*. Cambridge.
- Burger, Hermann (1982): *Die künstliche Mutter*. Frankfurt a.M.
- Church, Kenneth et al. (1991): "Using statistics in lexical analysis." In: Zernik, Uri (ed.): *Lexical Acquisition*. Englewood Cliffs, NJ, Erlbaum: 115–64.
- Church, Kenneth et al. (1994): "Lexical substitutability". In: Atkins, B. T. Sue/Zampolli, Antonio (eds.): *Computational Approaches to the Lexicon*. Oxford: 153–77.
- Church, Kenneth/Hanks, Patrick (1990): "Word association norms, mutual information and lexicography". *Computational Linguistics* 16/1: 22–9.
- Church, Kenneth/Mercer, Robert (1994): "Introducing to the special issue on computational linguistics using large Corpora". In: Armstrong, Susan (ed.): *Using Large Corpora*. Cambridge, MA: 1–24.

---

<sup>16</sup> Das Idiom hat laut Duden 11 (S. 294) seinen Ursprung in der Tatsache, "dass Verwundete im Todeskampf in das Gras oder ins Erdreich beißen, um sich die Schmerzen zu verbeißen. Diese Vorstellung findet sich schon im Altertum (vgl. Ilias 2, 418 und Aeneis 11, 118)."

<sup>17</sup> Anders jedoch bei Idiomen, bei denen die metaphorische Basis transparent ist, wie z. B. bei *sich zwischen alle Stühle setzen*, wo Bedeutungsextensionen auch aus dem zugrunde liegenden Bild hervorgehen können. Bei *ins Gras beißen* ist eine einmalige Verwendung mit der Bedeutung 'sich ärgern' festzustellen, die durch einen synonymen Ausdruck im Kontext hinreichend spezifiziert wird: *Wäre er im Strafraum liegengeblieben, hätte er ins Gras gebissen oder sich sonstwie über den nicht gepfiffenen Elfmeter echauffiert*. (Frankfurter Rundschau 06.07.1998, S. 30). Diese Bedeutungserweiterung ist nur aus dem Bild zu interpretieren: die beschriebene Geste impliziert Ärger.

- Corréard, Marie-Hélène (ed.) (2002): *Lexicography and Natural Language Processing. A Festschrift in Honour of B. T. S. Atkins*. Stuttgart.
- Coseriu, Eugenio (1975): *Sprachtheorie und allgemeine Sprachwissenschaft*. München.
- Cruse, D. Alan (1986): *Lexical Semantics*. Cambridge.
- Duden 11. *Redewendungen. Wörterbuch der deutschen Idiomatik* (<sup>2</sup>2002). Mannheim.
- Duden. *Deutsches Universalwörterbuch* (<sup>4</sup>2001). Mannheim.
- Fellbaum, Christiane (ed.) (erscheint): *Idioms and Collocations*. Birmingham.
- Francis, Gill (1994): "A corpus-driven approach to grammar: principles, methods and examples". In: Baker, Mona/Francis, Gill/Tognini-Bonelli, Elena (eds.) (1993): *Text and Technology. In Honour of John Sinclair*. Amsterdam: 137–56.
- Geyken, Alexander (erscheint): "The DWDS Corpus". In: Fellbaum, Christiane (ed.): *Idioms and Collocations*. Birmingham.
- Geyken, Alexander et al. (2004): "What is the optimal corpus size for the study of idioms?" *Vortrag auf der DGfS-Jahrestagung. Mainz (D) 25.–27.2. 2004*.
- Halliday, Michael A. K. (1992): "Language as system and language as instance: the corpus as a theoretical construct". In: Svartvik, Jan (ed.) *Directions in Corpus Linguistics*. Berlin: 61–77.
- Halliday, Michael A. K./Hasan, Ruqaiya (1976): *Cohesion in English*. London.
- Hanks, Patrick (1987): "Definitions and explanations". In: Sinclair, John (ed.): *Looking Up*. London: 116–36.
- Hanks, Patrick (1996): "Contextual dependency and lexical sets". *International Journal of Corpus Linguistics* 1/1: 75–88.
- Hanks, Patrick (2004): *The Syntagmatics of Metaphor and Idiom*. Unveröffentlichtes Manuskript.
- Herold, Axel (erscheint): "Corpus queries". In: Fellbaum, Christiane (ed.): *Idioms and Collocations*. Birmingham.
- Kennedy, Graeme (1998): *An Introduction to Corpus Linguistics*. London/New York.
- Klappenbach, Ruth/Steinitz, Wolfgang (eds.) (1964-1977): *Wörterbuch der deutschen Gegenwartssprache. Berlin*.: <http://www.dwds.de/woerterbuch>
- Lakoff, George (1987): *Women, Fire, and Dangerous Things. What the Categories Reveal about the Mind*. Chicago.
- Lakoff, George/Johnson, Mark (1980): *Metaphors We Live By*. Chicago.
- Löbner, Sebastian (2002): *Semantik: Eine Einführung*. Berlin.
- Lyons, John (1995): *Linguistic Semantics. An Introduction*. Cambridge.
- McEnery, Tony/Wilson, Andrew (1996): *Corpus Linguistics*. Edinburgh.
- Moon, Rosamund (1998): *Fixed Expressions and Idioms in English. A Corpus-Based Approach*. Oxford.
- Neumann, Gerald/Körner, Fabian (erscheint): "The corpus of examples". In: Fellbaum, Christiane (ed.): *Idioms and Collocations*. Birmingham.
- Nunberg, Geoffrey/Sag, Ivan/Wasow, Thomas (1994): "Idioms". *Language* 71/3: 491–538.
- Paul, Hermann (<sup>10</sup>2002). *Deutsches Wörterbuch*. Überarbeitet und erweitert von Helmut Henne, Heidrun Kämper und Georg Objartel. Tübingen.
- Saussure, Ferdinand de (1916): *Cours de Linguistique Générale*. Paris.

- Sinclair, John (1991): *Corpus, Concordance, Collocation*. Oxford.
- Sinclair, John (1996): "The search for units of meaning". *Textus* 9: 75–106.
- Sinclair, John (1998): "The lexical item". In: Weigand, Edda (ed.): *Contrastive Lexical Semantics*. Amsterdam: 1–24.
- Sinclair, John et al. (1998): "Language independent statistical software for corpus exploration". *Computers and the Humanities* 31/3: 229–55.
- Stubbs, Michael (1996): *Text and Corpus Analysis*. Oxford.
- Stubbs, Michael (2001): *Words and Phrases. Corpus Studies of Lexical Semantics*. Oxford.
- Wittgenstein, Ludwig (1953): *Philosophische Untersuchungen*. Frankfurt a.M.