

PHRASEOLOGIE IM ZEITALTER DER INFORMATIK UND DES INTERNETS

Dirk Hofmann, Universidad de Valladolid (Spanien)

- 1 Die Erstellung von Phraseologiedatenbanken
- 2 Das SZ-Archiv
- 3 Eurodicautom
- 4 Textkorpora des Instituts für deutsche Sprache
- 5 Ausblick
- Bibliographie

In diesem Artikel möchte ich zeigen, inwieweit die Informatik und das Internet der Phraseologie als Hilfsmittel dienen können. Es werden einige Datenbanken sowie deren Nutzen für die Phraseologie vorgestellt.

1 Die Erstellung von Phraseologiedatenbanken

Der Einsatz des Computers in der Linguistik und auch im Fremdsprachenunterricht ist nichts Neues. Besonders im Bereich der Lexikographie und der Übersetzung hat der Computer schon vor langer Zeit eine wichtige Rolle übernommen¹. Auf dem Gebiet der Übersetzung setzen sich die maschinelle und computergestützte Übersetzung immer mehr durch und auch in der Phraseologie böte sich der Computereinsatz, wie die Beispiele im Laufe dieses Kapitels zeigen, geradezu an.

Man könnte eine Phraseologiedatenbank erstellen, die sich aus Einträgen mit verschiedenartigen Informationen zu bestimmten Phraseologismen zusammensetzt. Des Weiteren müssten sich die Informationen der Datenbank leicht erweitern und modifizieren lassen. Datenbanken verfügen außerdem über ein größeres Volumen als traditionelle Wörterbücher. Es können somit mehr Varianten und zahlreichere Beispiele aufgenommen werden.

Erste Versuche möchte ich mit Hilfe des *SZ-Archivs*, eine Datenbank der *Süddeutschen Zeitung*, des *Eurodicautom* und der Textkorpora des *IDS* veranschaulichen. Ziel kann aber nur sein, mit Hilfe dieser Informationen, eine eigene Phraseologiedatenbank zu erstellen. Die Datenbank müsste aus mehreren Masken bestehen, die u.a. Angaben über Herkunft, Gebrauch, Stilmerkmale und Übersetzungsmöglichkeiten in zahlreiche europäische Sprachen enthielten. Außerdem müsste sie ständig aktualisiert werden und nur authentische Beispiele aus verschiedenen Textsorten enthalten, die sowohl den Bereich der schriftlichen als auch der mündlichen Sprache abdecken. Solch eine Datenbank wäre mit einem enormen Arbeitsaufwand verbunden und wäre nur innerhalb eines mehrjährigen Forschungsprojekts zu realisieren. Diesen enormen Arbeitsaufwand belegt auch das folgende Zitat von Dobrovol'skij (1989: 527):

Die primäre Aufgabe der computergestützten Phraseographie besteht u. E. nicht in der Übertragung der bereits kodifizierten Informationen in maschinenlesbare Formen, sondern in der Schaffung prinzipiell neuer Arten von Nachschlagewerken, die ohne Computereinsatz überhaupt nicht denkbar sind.

Dobrovol'skij hat bereits eine Idiomdatei für die deutsche Sprache erstellt, mit der er eine möglichst vollständige Informationsquelle über die Idiome zur Verfügung stellen

¹ Eine gute Übersicht über verschiedene computergestützte Untersuchungen im Bereich der Linguistik bietet Batori et al. (1995).

möchte. Der Nachteil der Datenbank ist die geringe Anzahl (1000) der gespeicherten Idiome. In der Maske erscheinen verschiedene Deskriptoren, die dem Benutzer u.a. stilistische und emotive Angaben der einzelnen Redewendungen vermitteln. Das Eingabefeld 'Belege' dient dazu Belegmaterial zu speichern.

Zu erwähnen ist auch das Lexikondatenbanksystem *Phraseo-Lex*², das von Martina Keil entwickelt wurde. *Phraseo-Lex* richtet sich an Phraseologieforscher, denen es als Hilfsmittel zur Erfassung und Erforschung verbaler Phraseologismen dienen soll und an Lexikologen, denen es als phraseographische Informationsquelle zur Seite stehen soll. Das Datenbanksystem ist ein Hilfsmittel für die Erfassung verbaler Phraseologismen. Phraseologismen werden vor ihrer Aufnahme in die Datenbank computergestützt analysiert. Laut Barbara Wotjak (1992: 88) ist *Phraseo-Lex* als ein einheitlicher Filter zu verstehen, den eine jede Einheit zu durchlaufen hat. Die Informationen, die der Benutzer über den jeweiligen Phraseologismus erhält sind sehr ausführlich. Der Phraseologismus wird zuerst mittels einer Paraphrase erläutert, dann folgen Angaben zur semantischen Struktur, die u.a. auch Hinweise zur Modifikation und zur Synonymie und Antonymie erhalten. Hilfreich sind auch die diatopischen, diaphasischen und diastratischen Angaben.

Nun möchte ich an dieser Stelle auf die von mir durchgeführten Untersuchungen mit den bereits erwähnten Datenbanken eingehen.

2 Das SZ-Archiv

Zum Suchen von Phraseologismen eignet sich bedingt die Suchmaschine der Datenbank der *Süddeutschen Zeitung* „SZ-Archiv“ (www.sz-archiv.de). Hier kann man alle Texte der *Süddeutschen Zeitung* der letzten 30 Tage kostenlos einsehen. Recherchen mit älteren Texten sind jedoch kostenpflichtig. Die Suchfunktion ist nicht speziell für Phraseologismen konzipiert, wodurch sich die Suche als mühsam und zeitaufwendig herausstellt. Phraseologismen sind mehrgliedrige Einheiten, die in ihrer Mehrzahl nur als Ganzes einen Sinn ergeben. Die Suchmaschine ist jedoch nicht in der Lage, sie als Ganzes zu erkennen, sondern gibt alle Texte an, in denen ein Element der Wendung vorkommt. Modifikationen von Phraseologismen können von dieser Datenbank, aber häufig auch von speziell für Phraseologismen konzipierte Datenbanken, nicht lokalisiert werden, da Modifikationen auf Grund ihrer vielfältigen Assoziationsmöglichkeiten eine enorme Komplexität aufweisen. Die Suchmaschine des *SZ-Archivs* zeigt daher oft 50 Treffer an, aber nur bei einem oder zwei handelt es sich um Phraseologismen. Die Aufgabe eines Informatikers wäre nun, ein Programm so zu konzipieren, dass die Suchmaschine imstande ist, feste Wortverbindungen von freien zu unterscheiden; fast unmöglich wird wohl auch weiterhin das Erkennen von modifizierten Phraseologismen sein.

Trotzdem gelingt es durchaus zu befriedigenden Ergebnissen zu kommen. Der Fremdsprachenlerner hat die Möglichkeit, sich ein Archiv mit authentischen und vor allem aktuellen Beispielen anzulegen. Ein weiterer Vorteil ist, dass die Beispiele in einem Kontext und nicht isoliert präsentiert werden, wodurch dem Fremdsprachenlerner mehrere Beispiele in verschiedenen Kontexten vermittelt werden. Hierdurch ist er im Nachhinein in der Lage, die Phraseologismen in ähnlichen Kontexten korrekt anzuwenden, d.h. das *SZ-Archiv* ist besonders bei der Kontextfrage behilflich, ohne jedoch konkrete stilistische Angaben zu machen.

Natürlich bringt das *SZ-Archiv* auch einige Nachteile mit sich, was teilweise damit zusammenhängt, dass das Archiv weder speziell für DaF-Lerner noch für die Phraseologismensuche entworfen worden ist. Die Beispiele stammen alle aus der

² Aktuelle Informationen über *Phraseo-Lex* findet man unter der folgenden Internetadresse: <http://www2.informatik.uni-erlangen.de/IMMD-II/Research/Activities/CL>.

Pressesprache, d.h. es handelt sich um Schriftsprache, womit die gesprochene Sprache, in der besonders viele Redewendung vorkommen, nicht berücksichtigt wird. Auch Fachsprache wird nur begrenzt ins Auge gefasst. Sie kommt regelmäßig bedingt im Wirtschaftsteil als Wirtschaftssprache bzw. als Verteilersprache vor. Darüber hinaus muss der DaF-Lerner die Wendung bereits kennen, um ihren Gebrauch zu überprüfen. Bei Phraseologismen, die dem DaF-Lerner unbekannt sind, hilft die Datenbank nicht weiter.

Im Folgenden möchte ich einige Beispiele präsentieren, die ich willkürlich ausgewählt habe, um die Arbeit mit der Datenbank zu veranschaulichen:

Bei der Suche nach Phraseologismen im engeren Sinne fällt auf, dass sich das erfolgreiche Auffinden nach der Komplexität des Phraseologismus richtet. Phraseologismen, die aus wenigen Wörtern bestehen sind einfacher zu finden als komplexere Einheiten. Damit auch Letztere gefunden würden, müsste die Datenbank, wie bereits schon erwähnt, auf phraseologische Suchfunktionen spezialisiert werden. Beispiele:

- (1) die Ärmel auf- / hochkrepeln: Der Münchner SPD-Chef Franz Maget sagte, er wolle trotz des Affronts „keinen innerparteilichen Machtkampf anzetteln, sondern die Ärmel hochkrepeln³, damit wir die Direktmandate holen“ (SZ, 11.12.01, S. 41).
- (2) Jmdn. aufs Abstellgleis schieben: In Andreas Herzog wurde ein Star aufs Abstellgleis geschoben, der als Spielmacher der alten Schule gilt (SZ, 15.12.01, S. 47).
- (3) Jmdm. ein Dorn im Auge sein: So soll die gegenseitige Verflechtung der deutschen Wirtschaft beendet werden, die vor allem angelsächsischen Investoren ein Dorn im Auge ist (SZ, 18.12.01, S. 23).
- (4) Farbe bekennen: Am kommenden Mittwoch will Ulrich Brixner, Chef der frisch fusionierten DZ Bank, erstmals öffentlich Farbe bekennen (SZ, 11.12.01, S. 24).
- (5) Den Gürtel enger schnallen: Sollen etwa 50.000 Arbeitnehmer, die mit viel Fleiß und Einsatz Jahr für Jahr geschmackvolle Präsente produzieren, den Gürtel enger schnallen, nur weil manche Controller kleine Aufmerksamkeiten zu Weihnachten für verzichtbar halten (SZ, 21.12.01, S.21)?
- (6) Das Handtuch werfen: darf sich nicht wundern, wenn die Leistungen unserer Schüler sinken und mehr und mehr Lehrer „ausgebrannt“ das Handtuch werfen, weil sie sich nicht in der Lage sehen, neben interessantem Unterricht sechs Stunden am Tag Entertainer zu spielen und so die Schüler bei Laune zu halten (SZ, 18.12.01, S.39).
- (7) Etwas unter einen Hut bringen: Der Taktiker Arafat wollte unter einen Hut bringen, was nicht unter einen Hut zu bringen ist: überwiegend friedlichen Widerstand und bewaffneten Kampf (SZ, 17.12.01, S.4).
- (8) Bei jmdm. einen Stein im Brett haben: Die Firma Remington hat bei mir eigentlich einen Stein im Brett: Der Name erinnert mich an meinen Papa, der hatte immer einen Rasierapparat von Remington (SZ, 23.11.01, S. 28).
- (9) Gegen den Strom schwimmen: Anderen fehlte die Einsicht in die eigenen Möglichkeiten. Die meinten, man muss durchhalten, gegen den Strom schwimmen (SZ, 21.11.01, S.26).
- (10) An den Mann bringen: Wer vor dem Währungswechsel seinen Gebrauchten noch „von Privat“ an den Mann bringen und Fahrzeugschein gegen Bargeld tauschen will, sollte – so warnt der Automobilclub von Deutschland (AvD) – dringend einige Vorsichtsmaßnahmen beachten (SZ, 15.12.01, S.1).

³ Die Hervorhebungen wurden von mir veranlasst.

Bei den gefundenen Phraseologismen fällt auf, dass viele, z.B. 'etwas an den Mann bringen', 'etwas unter einen Hut bringen', 'das Handtuch werfen', etc. in den Bereich der Umgangssprache gehören. Dies ist etwas erstaunlich, da es sich um Pressesprache handelt, was aber auch wiederum zeigt, dass die gesprochene Sprache immer mehr in die geschriebene Sprache eindringt.

Als nächstes gehe ich auf einige Beispiele ein, die sich auf die Phraseologie im weiteren Sinne beziehen. Bei den folgenden Wendungen handelt es sich um Kollokationen, die wir vor allem im Bereich der Wirtschaftssprache antreffen und die für DaF-Lerner mit gewissen Problemen verbunden sind:

- (11) Konkurs anmelden: Würden Sie verreisen, müsste Ihre Stammkneipe Konkurs anmelden (SZ, 30.11.01, S. 6).
- (12) Kosten übernehmen: Ein Junge studiert Übersetzen an der Universität, die Kosten hat für fünf Jahre ein Sponsor übernommen (SZ, 05.12.01, S. 11).
- (13) Die Kosten tragen: ... ob die Stadtverwaltung oder die Hotellerie die wegen zusätzlicher Sicherheitsvorkehrungen gestiegenen Kosten tragen muss (SZ, 30.11.01, S. 37).
- (14) Die Aktie notieren: Die Aktie notiert derzeit unter einem Euro (Schlusskurs am Freitag 0,37 Euro) (SZ, 08.12.01, S. 32).
- (15) Rote Zahlen: An diesem Dienstag aber wird der Vorstandschef der Tochter, Ulrich Schumacher, rote Zahlen präsentieren müssen (SZ, 13.11.01, S. 25).
- (16) Rote Zahlen schreiben: Wegen der hohen Investitionen in die UMTS-Lizenz von mehr als acht Milliarden Euro im Jahr 2000 und der Vorlaufkosten wird Mobilcom auf Jahre hinaus rote Zahlen schreiben und rechnet erst im Jahr 2007 mit dem Breakeven (SZ, 13.11.01, S. 28).

Bei dieser Kollokation ist auffällig, dass die Suchmaschine sie genau in dieser Kombination findet und nicht alle Wendungen angibt, in denen das Adjektiv rot und das Substantiv Zahlen vorkommen.

Das Gleiche geschieht auch mit der nächsten Wendung, die zwar umgangssprachlich ist, aber durchaus in der Wirtschaftssprache vorkommt:

- (17) Etwas auf die hohe Kante legen: Durch Ansparungen für die private Rente werden allein in Deutschland ab dem Jahr 2008 jährlich rund 70 Milliarden DM zusätzlich auf die hohe Kante gelegt (SZ, 20.11.01, S. 24).

Als problemlos stellt sich die Suche nach den so genannten *Einwortphraseologismen* heraus:

- (18) Schmiergeld: Die Differenz ist nach Feststellungen der Staatsanwaltschaft als Schmiergeld an die Briefkastenfirmen Linsur, Ovessim, Great Aziz und ATG geflossen (SZ, 17.11.01, S.5)
- (19) Kurssturz: Jetzt sind derartige Visionen obsolet, erklären Analysten den Kurssturz (SZ, 21.11.01, S.28).

Ebenfalls problemlos ist die Suche nach Phraseologismen, die eine unikale Komponente erhalten, d.h. ein Lexem, das im deutschen Wortschatz nur noch in dieser Redewendung vorkommt. Die Suchmaschine braucht also nur nach diesem einen Lexem zu suchen:

- (20) klipp und klar: „Wir sagen klipp und klar, wo keine Gegenleistung erfolgt oder zuviel bezahlt wird, fordern wir das Geld von der Bahn zurück“, erklärte Michael Zirpel, Pressesprecher des Ministerium (SZ, 05.12.01, S. 24).

- (21) keinen Hehl aus etwas machen: EU-Agrarkommissar Franz Fischler, aber auch die spanische Verkehrs- und Energiekommissarin Loyola de Palacio machen aus ihrer Kritik keinen Hehl (SZ, 11.12.01, S. 21).
- (22) ins Fettnäpfchen treten: Was Christine Kaufmanns Größe ausmacht, ist dieses Übermaß an authentischer Unschuld, mit der sie ins Fettnäpfchen tritt... (SZ 07.12.01, S. 18).

Das Suchen von Paar- bzw. Zwillingsformeln bereitet der Suchmaschine, wie die folgenden Beispiele zeigen, auch keine großen Schwierigkeiten:

- (23) fix und fertig: Die Wohnung sieht aus, als sei sie in den siebziger Jahren vom Möbelhaus fix und fertig eingerichtet und seither nicht mehr verändert worden (SZ, 14.12.01, S. 3).
- (24) Nur sechs der 30 Gemeinderäte votierten für das bereits fix und fertig geplante Projekt, darunter Bürgermeister Toni Neidlinger als einziger von der Mehrheitsfraktion CSU (SZ, 08.12.01, S. 8).
- (25) mit jmdm. durch dick und dünn gehen: Durch dick und dünn geht die ARD mit ihren Schmonzetten: Hier Ottfried Fischer und Reinhard Fendrich in Eine Insel zum Träumen (SZ, 30.11.01, S. 43).
- (26) Schulter an Schulter: Denn bei der Bekämpfung der Terror Finanzierung stehen die führenden Industriestaaten, anders als beim Thema Steuer-Dumping, Schulter an Schulter zusammen (SZ, 19.11.01, S. 23).

Die Suche nach komparativen Phraseologismen ist etwas mühsamer. Sie tauchen meist nach der Eingabe des im Vergleich genannten Substantivs auf. Komparative Phraseologismen sind jedoch in der Pressesprache nicht sehr häufig und viele von ihnen sind autorenspezifisch, d.h. das Grundwort ist unbekannt, wodurch sie im Korpus nicht auffindbar sind. Im *SZ-Archiv* habe ich lediglich ein Beispiel gefunden:

- (27) flink wie ein Wiesel: Flink wie ein Wiesel flitzt er los, schnell und sicher wie ein Hase schlägt er Haken und unbeirrt rennt er geradeaus (SZ, 01.12.01, S. 2).

Schnell und sicher wie ein Hase ist zwar auch ein phraseologischer Vergleich, der allgemein verständlich ist, jedoch ist es eine Art Ad-hoc-Bildung, die nicht in phraseologischen Wörterbüchern verzeichnet ist.

Kinegramme werden von der Suchmaschine schnell gefunden, da die in den Kinegrammen vorkommenden Verben meist nur in den Kinegrammen benutzt werden, so dass nach der Eingabe des Verbs der ganze Ausdruck angezeigt wird:

- (28) die Nase rümpfen: Dennoch rümpfen heute viele in Luxemburg nur noch die Nase, wenn der Name Lussi fällt (SZ, 06.12.01, S. 26).
- (29) (mit dem Kopf) nicken: Ein Mann im Hintergrund, es ist der Telefonbesitzer, identifiziert den Dieb mit einem Nicken (SZ, 11.12.01, S. 3).
- (30) Das reichte, um die Liebhaber von Gerüchten schwer und stumm nicken zu lassen (SZ, 10.12.01, S. 25).

Traditionelle Geflügelte Worte findet die Suchmaschine, wie die folgenden Beispiele zeigen, ohne große Probleme, jedoch muss man bedenken, dass Geflügelte Worte nicht zu den Eigenschaften der Pressesprache zählen:

- (31) Sein oder Nichtsein, das ist hier die Frage (SZ, 23.11.01, S.15).

Erstaunlicherweise lassen sich mit diesem Archiv auch Sprichwörter ausfindig machen. Hierzu möchte ich aber nur ein Beispiel angeben, da Sprichwörter nicht im Mittelpunkt dieses Beitrags stehen:

- (32) Turin – Als echte Piemontesen sind die Vertreter der Familie Agnelli keine redseligen Menschen. Reden ist Silber, Schweigen ist Gold, ohnehin reichen ein halbes Lächeln des *Avvocato* Gianni Agnelli oder eine hochgezogene Augenbraue von Bruder Umberto, um Heerscharen von Journalisten zu waghalsigen Interpretationen zu animieren (SZ, 21.11.01, S. 37).

3 Eurodicautom

Bei *Eurodicautom* handelt es sich um eine Terminologiedatenbank der Europäischen Union, die seit 1973 existiert und seit 1986 auch die spanische Sprache einschließt. Die Datenbank wird täglich weltweit von 120.000 Personen benutzt. Die gespeicherten Termini gehören zu den Fachgebieten der Landwirtschaft, Telekommunikation, Rechtswissenschaft, Wirtschaftswissenschaft, etc., d.h. in erster Linie erfasst die Datenbank den Bereich der Fachsprachen. Die meisten Termini sind alle in Originaltexten der Europäischen Union verzeichnet und können dann in alle EU-Sprachen übersetzt werden. Bei einigen Termini wird jedoch als Nachweis ein Wörterbucheintrag angegeben.⁴

Als nächstes überprüfe ich, ob die Datenbank in der Lage ist Phraseologismen zu identifizieren und sie ggf. adäquat ins Spanische zu übersetzen. Der Einfachheit halber benutze ich die Phraseologismen oder ähnliche, die ich bereits für das *SZ-Archiv* benutzt habe.

Phraseologismen im engeren Sinne, wie z.B. 'die Ärmel hochkrempeln', 'etwas unter einen Hut bringen', etc. sind in der Datenbank nicht verzeichnet bzw. durch die Suchmaschine nicht auffindbar. Dies lässt sich dadurch erklären, dass diese Terminologiedatenbank, wie bereits oben erwähnt, nur fachsprachliche Quellen und keine gemeinsprachlichen Quellen erfasst.

Anders sieht es bei Termini aus dem Bereich der Wirtschaftssprache aus. Da die Datenbank den Bereich der Wirtschaftssprache einschließt sind folglich Phraseologismen aus diesem Bereich anzutreffen, allerdings gehören die meisten Phraseologismen zu der Untergruppe der Kollokationen:

- (33) 'In Konkurs geraten' wird fälschlicherweise durch 'declaración de quiebra' übersetzt, was aber eher Konkurserklärung ist. Eine geeignete Übersetzung wäre *declararse en quiebra*.
- (34) 'Auf Kosten von' wird korrekt übersetzt und es werden verschiedene Ausdrücke, wie 'a costa de', 'a expensas de', 'por cuenta de', angezeigt.
- (35) 'Die Kosten zu tragen haben' wird durch 'correr con los gastos' korrekt wiedergegeben.
- (36) Auch der Fachterminus 'garantierte Aktie' wird durch 'acción con dividendo garantizado' richtig übersetzt, obwohl 'fondos garantizados' häufiger benutzt wird.

Selbst selten gebrauchte Ausdrücke, die in Wörterbüchern fehlen sind in der Datenbank verzeichnet und werden meist korrekt übersetzt. Ein Beispiel hierfür ist 'Zahlen auf das Komma ausrichten' vs. 'alinear cifras con la coma'.

Einwortphraseologismen sind auf Grund ihrer einfachen Struktur leicht von der Suchmaschine zu identifizieren und sind auch meist in der Datenbank verzeichnet. Hierzu nun einige Beispiele:

- (37) 'Kurssturz' wird durch 'caída de la clasificación crediticia del banco' übersetzt. Diese etwas umständlich wirkende Übersetzung kommt vermutlich durch den

⁴ Diese und andere Informationen zu dieser Datenbank findet man auf der folgenden Homepage: <http://europa.eu.int/eurodicautom/login.jsp>

gegebenen Kontext zu Stande. 'Schwarzmarkt' wird korrekt durch 'mercado negro' übersetzt.

Phraseologismen mit unikalen Komponenten sind in der Datenbank nicht verzeichnet. Dies lässt sich dadurch erklären, dass Phraseologismen wie 'ins Fettnäpfchen treten', 'keinen Hehl aus etwas machen', etc. nicht in den Bereich der Fachsprachen gehören.

Das Gleiche gilt auch für Paar- bzw. Zwillingsformeln, Kinogramme, komparative Phraseologismen und Sprichwörter. Die Datenbank hilft uns also im Bereich der Phraseologie nur bedingt und das auch nur, wenn es sich um den Bereich der Fachphraseologie handelt. Des Weiteren müsste die Suchmaschine noch spezieller für das Identifizieren von Phraseologismen konzipiert werden. Lobenswert ist aber die Übersetzungsfunktion gemeinsprachlicher Termini. Die Datenbank bietet uns zahlreiche Termini, die auf Grund ihrer Aktualität meist noch nicht in den Wörterbüchern verzeichnet sind, so z.B. 'Starterkit' vs. 'euromonedero'. Da dies jedoch schon in den Bereich der Terminologielehre und Terminologiearbeit gehört, möchte ich das an dieser Stelle nicht näher darlegen.

4 Textkorpora des Instituts für deutsche Sprache

Als nächstes möchte ich auf die Textkorpora des Instituts für deutsche Sprache eingehen. Die *IDS*-Textkorpora mit einem Gesamtumfang von nunmehr ca. 1736 Mill. Textwörter (das wären etwa 4.340.000 Buchseiten) bzw. 1440 Mill. Wortformen beinhalten deutschsprachige belletristische, populärwissenschaftliche und wissenschaftliche Texte, eine große Anzahl von Zeitungstexten sowie eine breite Palette weiterer Textarten aus Vergangenheit und Gegenwart. Sie werden von der Arbeitsgruppe für Korpustechnologie kontinuierlich weiterentwickelt, u.a. in Form von Monitor-Korpora für diachronische Untersuchungen. Die *IDS*-Textkorpora stellen ein unverzichtbares Instrumentarium nicht nur für die linguistische Forschung am *IDS*, sondern durch die zunehmende Zahl von Online-Nutzern für die nationale und internationale Germanistikforschung dar. Sie werden aber auch in wachsendem Maße für interdisziplinäre Untersuchungen genutzt, etwa in den Fachgebieten Psychologie, Neurologie, Kognitionswissenschaft, Sprachtherapie, Kommunikations- und Medienwissenschaft, Statistik. Aus urheberrechtlichen Gründen ist nur ein Teil der *COSMAS – I - Korpora* öffentlich zugänglich⁵. Hierzu gehören der *Mannheimer Morgen* (*mmm* / 1989, 1991, 1994-2001), das *Korpus Magazin Lufthansa Bordbuch* (*mlb* / 1995-1997) und die folgenden Korpora:

Grammatik-Korpus (gr1)

Bonner Zeitungskorpus (bzk)

Handbuchkorpora (hbk), gesondert nach Jahrgängen: *h85*,
h86, h87, h88

Wendekorpus (wk): *Wendekorpus West (wkb; Bundesrepublik
Deutschland)* *Wendekorpus Ost (wkd; DDR)*

LIMAS-Korpus (lim)

*Korpus-Kartei der Gesellschaft für deutsche Sprache, Wiesbaden
(GfS)*

Goethe-Korpus (goe)

GRIMM-Korpus (gri)

Marx-Engels-Korpora:

⁵ Diese Informationen findet man auf der folgenden Homepage: <http://www.ids-mannheim.de/kt/corpora.shtml>.

Marx-Engels-Gesamtausgabe -ausgewählte Texte (*meg*)

Marx-Engels-Werke - ausgewählte Texte (*mew*), Herausgeber-Anmerkungstexte zu *mew* (*mwa*)

Die genannten Korpora beziehen sich auf die geschriebene Sprache. Für die gesprochene Sprache (verschriftlichte Form) stehen drei Korpora zur Verfügung. An dieser Stelle sollte das *IDS* expliziter darauf hinweisen, dass die Korpora um 1960/70 entstanden sind und nur rund 700 Texte enthalten. Da die Korpora veraltet sind, findet man hier keine aktuellen Wendungen, wie 'nicht immer, aber immer öfter':

Freiburger Korpus (fko)

Dialogstrukturenkorpus (dsk)

PFEFFER-Korpus (pfe)

Die Suchmaschine ist, wie die folgenden Beispiele zeigen, in der Lage Phraseologismen im engeren Sinne zu erkennen. Da das Korpus viel umfangreicher ist als das *SZ-Archiv*, erhält der Benutzer logischerweise mehr Beispiele. Für 'die Ärmel hochkrempeln' erhalten wir 131 Treffer. Man hat die Möglichkeit mehrere Korpora abzufragen, so gibt es z.B. eine Suchfunktion, die alle geschriebene Korpora erfasst oder eine, die nur die Korpora der gesprochenen Sprache umfasst. Die folgenden Beispiele sollen die Arbeit mit der Suchmaschine veranschaulichen:

- (38) die Ärmel hochkrempeln: Bögler versprach, man werde die Ärmel hochkrempeln um die Probleme zu bewältigen und Schweiß vergießen (Mannheimer Morgen, 18.01.1989).
- (39) das Handtuch werfen: Trotzdem mußten zahlreiche Händler das Handtuch werfen (Computer Zeitung, 11.05.1995, S. 10).
- (40) Luftschlösser bauen: Wahrscheinlich muß man, um beweglich zu bleiben und Freiräume zu gewinnen, seine Zukunft auch herbeiträumen können und ebensolche Luftschlösser bauen wie der Junge in Martin Honerts Fata Morgana (Frankfurter Rundschau, 30.06.1998, S. 10).

Sowohl fachsprachliche als auch nicht fachsprachliche Kollokationen werden von den Textkorpora problemlos angezeigt. Darüber hinaus hat der Benutzer auch die Möglichkeit, ein Kollokationsanalyse-Parameter zu verwenden.

- (41) Konkurs anmelden: Wenn ein öffentliches Gemeinwesen unter den Voraussetzungen Konkurs anmelden könnte, unter denen eine Aktiengesellschaft dazu verpflichtet ist, hätte sich Lafontaine sofort nach seinem Regierungsantritt zum zuständigen Amtsrichter in Saarbrücken begeben müssen (Die Zeit, 16.08.85, S. 04).
- (42) den Hund ausführen: Im Supermarkt einkaufen, den Hund ausführen, mit Freunden in die Kneipe gehen - alles kein Problem für den Niederländer: "Wer sich nicht mehr auf die Straße traut, ist ein überkandidelter Spinner" (Mannheimer Morgen, 19.10.2000).
- (43) Abhilfe schaffen: Auch eine häufigere Versetzung der Beschäftigten in gefährdeten Abteilungen wie Planung und Auftragsvergabe könnte Abhilfe schaffen (Mannheimer Morgen, 15.07.1995).

Einwortphraseologismen sind in der Datenbank zahlreich vorhanden und werden auf Grund ihrer Einfachheit von der Suchmaschine identifiziert:

- (44) Papierkrieg: "Der leidige Papierkrieg gerät mir langsam aus der Kontrolle", klagt ein Mann, der für Esso an der Zapfsäule steht und nach eigenem Bekunden "nie fertig wird" (Mannheimer Morgen, 03.05.1989).

- (45) Börsensturz: Überdies ist die Immobilien- und Bankenkrise, die zum erwähnten Börsensturz führte, noch immer nicht ausgestanden (Zürcher Tagesanzeiger, 19.03.1996, S. 39).

Das Gleiche gilt auch für Phraseologismen mit unikalenen Komponenten. In der Datenbank sind fast alle unikalenen Komponenten zu finden, was man von dem SZ-Archiv nicht behaupten kann:

- (46) Maulaffen feilhalten: Die Eingeborenen halten Maulaffen feil, hat man ja nicht alle Tage, so ein Schauspiel direkt vor der Haustür (Mannheimer Morgen, 13.11.1998).
- (47) etwas auf dem Kerbholz haben: Allerdings: Beide Richter hatten nur unvollständige Akten über die Delinquenten vorliegen - nach ihnen hatte Christian nur fünf Taten auf dem Kerbholz, Patrick lediglich sechs (Mannheimer Morgen, 03.07.1998).

Auch Paar- bzw. Zwillingformeln werden von der Suchmaschine leicht erkannt:

- (48) Hab und Gut: Nun aber, in dem Halbjahr 1688/1689, wurden Leib und Leben, Hab und Gut der Menschen in und um Kurpfalz systematisch vernichtet (Mannheimer Morgen, 21.03.1989).
- (49) frank und frei: Weiland habe die Masse des Volks ja auch Hitlers Buch "Mein Kampf" nicht gelesen, worin der Nazi-Führer seine kriminellen, später verwirklichten Absichten frank und frei dargelegt hatte (Mannheimer Morgen, 03.05.1989).

Die Suche nach komparativen Phraseologismen erweist sich auf Grund der genannten Gründen als schwierig, jedoch lassen sich Dank des umfangreichen Korpus mehrere Beispiele ausfindig machen:

- (50) frech wie Oskar: Auf dem Platz frech wie Oskar, bewiesen die Spieler erneut, daß sie eine Bereicherung für die Regionalliga Süd sind (Frankfurter Rundschau, 020.10.1997, S. 26).
- (51) dumm wie Bohnenstroh: Man bleibt nicht dumm wie Bohnenstroh, sondern erfährt bei einem Rundgang, daß Hülsenfrüchte zur großen Familie der Schmetterlingsblütler gehören, aber auch, daß nur ganz bestimmte Sorten genießbar sind (Mannheimer Morgen, 28.04.1989).

Die Suche nach Kinegrammen stellt sich, wie bereits vorhin erläutert als wesentlich einfacher heraus:

- (52) die Achseln zucken: Daß die Erwähnung seines Namens jenseits des Rheines zunehmend Stirnrünzeln auslöst, läßt den Belgier nur noch die Achseln zucken (Mannheimer Morgen, 31.01.1998).

Die Suchmaschine ist sogar im Stande Sprichwörter ausfindig zu machen, auch wenn die mich hier nicht weiter interessieren, da sie streng genommen in den Bereich der Parömiologie gehören, daher nur ein Beispiel:

- (53) Morgenstund hat Gold im Mund: Er hat auch versprochen dafür zu sorgen, dass den Frühaufstehern und Genießern - «Morgenstund hat Gold im Mund» - unter seinen Badegästen würziger Kaffee und frische Gipfeli bereitstehen (St. Galler Tagblatt, 24.06.1998, S. 17).

Die Suchmaschine bzw. die Textkorpora des *IDS* stellen eine große Hilfe bei dem Umgang mit Phraseologismen dar. Da es sich um ausführliche Textkorpora handelt gibt es nur wenige Phraseologismen, die nicht in der Datenbank verzeichnet sind. Dies ist auch ein Vorteil gegenüber dem *SZ-Archiv*, das sich nur auf die Pressesprache bezieht, nicht so umfangreich ist und nur auf den letzten Monat zurückgeht. Wünschenswert wäre, dass das *IDS* das Textkorpus der gesprochenen Sprache erweitern würde, da Phraseologismen ja vor allem dort zu finden sind. Der Benutzer muss genau wie beim *SZ-Archiv* die Phraseologismen kennen, um sich dann bestimmte Verwendungsweisen der Phraseologismen anzusehen und diese dann später beim Umgang mit der Fremdsprache anzuwenden.

Das Lexikondatenbanksystem *Phraseo-Lex* ist zur Zeit noch nicht über Internet zugänglich, so dass ich auf konkrete Fallbeispiele mit Hilfe dieses Systems verzichten muss.

5 Ausblick

Gut konzipierte Datenbanken können u.a. bei der Entwicklung phraseologischer Wörterbücher eine wichtige Rolle spielen. Diese Wörterbücher stünden dann wiederum dem Übersetzer zur Seite, der dann qualitativ bessere Informationen zu den einzelnen Phraseologismen bekäme. Des Weiteren könnten die Datenbanken, wie bereits angedeutet, von DaF-Lernern benutzt werden, die vor allem von den zahlreichen Beispielen profitieren können.

Dieser Beitrag ist sozusagen ein Blick in die Zukunft. Der Computer und das Internet stellen schon ein wichtiges Werkzeug der Sprachwissenschaft dar. Was die Phraseologie anbetrifft, so habe ich in diesem Artikel Dinge dargelegt, die noch vor gut zehn Jahren unvorstellbar gewesen wären, jedoch wünscht man sich für die Zukunft, dass die Anzahl guter Datenbanken und die Funktion dieser Datenbanken in den nächsten Jahren noch deutlich zunimmt.

Bibliographie

1 Korpus:

SZ-Archiv

Eurodicautom

Textkorpora des *Instituts für deutsche Sprache*

2 Wissenschaftliche Literatur:

Batori, I. et al. (1995): *Computational Linguistics. Computerlinguistik. Ein internationales Handbuch zur computergestützten Sprachforschung und ihrer Anwendungen*. Berlin: De Gruyter.

Burger, Harald et al. (1982): *Handbuch der Phraseologie*. Berlin: De Gruyter.

Delplanque, C. (1999): „Der Beitrag der Fachphrase zur Wissensrepräsentation.“ In: Fernández Bravo, N. / Behr, I. / Rozier, C. (Hg.): *Phraseme und typisierte Rede*. Tübingen: Stauffenburg Verlag, 81-97.

Dobrovol'skij, D. (1989): „Linguistische Grundlagen zur computergestützten Phraseographie.“ In: *Zeitschrift für Germanistische Linguistik* 10, 528-536.

Duhme, M. (1991): *Phraseologie der deutschen Wirtschaftssprache*. Essen: Verlag Die Blaue Eule.

Wotjak, B. (1992): *Phraseolexeme in System und Text*. Tübingen: Max Niemeyer Verlag.