

Korpuslinguistik und Unterricht Deutsch als Fremdsprache

Willi LANGE

1. Einleitung

Meine Damen und Herren, liebe Kolleginnen und Kollegen¹⁾, die hier Versammelten haben sehr unterschiedliche wissenschaftliche Schwerpunkte. Einen kleinen Ausschnitt davon konnten wir heute Nachmittag kennen lernen. Aber bei aller unterschiedlichen Schwerpunktsetzung in der Forschung haben wir doch zwei wichtige Gemeinsamkeiten: Da ist zum einen die deutsche Sprache und zum anderen die Vermittlungssituation. Wir unterrichten Deutsch als Fremdsprache. Das ist mein Ausgangspunkt.

Auch für erfahrene Lehrerinnen und Lehrer gibt es im Unterricht Situationen, wo man in Schwierigkeiten gerät, wo man Herausforderungen zu bestehen hat. Ich möchte hier drei Formen der Herausforderung unterscheiden. Da gibt es zunächst die **pädagogische Herausforderung**. Was machen Sie mit einem Studenten, der im Unterricht schläft? Oder was machen Sie mit einer Studentin, die sich weigert, an Gruppenarbeit teilzunehmen? Das sind wichtige Fragen, doch ich möchte heute nicht darauf eingehen. Eine zweite Form der Herausforderung ist die **didaktische Herausforderung**²⁾. Wie führen Sie den Akkusativ im Anfängerunterricht ein? Oder wie gestalten Sie eine längere Übungssequenz? Ich möchte hier zunächst über eine dritte Form der Herausforderung sprechen - die **sprachliche Herausforderung**. Sowohl für sogenannte Muttersprachler als auch für die nicht muttersprachlichen Lehrer gibt es Fragen von Lernern, wo wir „ins Schleudern kommen“. Im Folgenden möchte ich zunächst über zwei sprachliche Herausforderungen sprechen, die ich am eigenen Leib erlebt habe.

1) Dieser Text ist die überarbeitete Fassung eines Vortrags, der am 25.9.2004 im Rahmen eines Treffens der germanistischen Gesellschaft der Waseda Universität gehalten wurde (早稲田大学ドイツ語学・文学学会 研究発表会) gehalten wurde. Die Textsorte wurde bewusst beibehalten.

2) Ich verwende hier den Begriff *didaktisch* in einem sehr weiten Sinne. Vgl. Kron (1994, 42f.).

2. *Sushi* vs. *Sushis*

Ein Student kommt nach dem Unterricht unvermittelt zu mir und meint: „Sie haben gesagt, dass jedes deutsche Nomen maskulin, feminin oder neutrum ist. Wie ist das bei *Sushi*? Heißt es *der Sushi*, *die Sushi* oder *das Sushi*? Und was ist der Plural von *Sushi*? *Die Sushi*? *Die Sushis*? Oder etwas anderes?“ Wie hätten Sie an meiner Stelle reagiert?

Es gibt natürlich eine ganze Anzahl an Lösungsstrategien. Ich habe jedenfalls geantwortet: „Ich bin sicher, dass es *das Sushi* ist. Bei dem Plural bin ich nicht so sicher, wahrscheinlich ist es *die Sushi*. Aber ich werde nachsehen.“³⁾

2.1. Große Duden-*Sushis*

Die Recherche im Duden (1999), dem 10-bändigen großen Wörterbuch der deutschen Sprache, ergibt:

Sushi: [zu:ʃi], das, -s, -s(jap.): *aus rohem Fisch (Fleisch, Krustentieren, Gemüse, Pilzen u.a.) auf einer Unterlage aus Reis bestehendes Gericht*.: S., der Imbiss aus klebrigem Essigreis und rohem Fisch, gilt als trendy (Hörzu 52, 1998, 15) Statt US-Fastfood ist S. Favorit (Focus 3, 1993, 128)

Sprechen wir jetzt nicht über die schwierige Phonetik mit dem stimmhaften *s* und dem langen Vokal. Ich bringe das fast nicht über die Lippen. Thematisieren wir auch nicht die etwas pampige Bedeutungsbeschreibung, bei der *Nigiri-Zushi* und *Domburi* sich zu einer Masse klumpen. Wichtig sind in unserem Zusammenhang die grammatischen Angaben: *Sushi* ist ein Neutrum, der Genitiv Singular ist (*des*) *Sushis* und der Plural ist (*die*) *Sushis*.

Die Suche in anderen Nachschlagewerken bringt, wenn überhaupt, auch kein anderes Ergebnis. Können wir uns damit zufrieden geben? Bis vor wenigen Jahren hätten wir uns damit wohl zufrieden geben müssen. Heute jedoch sind die Recherchemöglichkeiten dank der Sprachdatenbanken viel breiter und tiefer geworden.

2.2. COSMAS-*Sushi*

Korpora sind große maschinenlesbare Textsammlungen, die unter verschiedenen sprachlichen Gesichtspunkten durchsucht werden können. Es gibt private Korpora, die mit bestimmten wissenschaftlichen Zielsetzungen erstellt wurden. Es gibt

3) Die folgende Analyse ist ein Teilergebnis einer viel breiteren Untersuchung zum Thema *Sushi*. Vgl. Lange (2005).

öffentliche Korpora, die auf CD-ROM erhältlich sind. Hierzu zählen auch die Jahrgangsausgaben von deutschsprachigen Zeitungen wie *Frankfurter Allgemeine Zeitung*. Schließlich gibt es noch Korpora, die direkt über das Internet zugänglich sind. Diese Korpora bezeichne ich als **Sprachdatenbanken**.

Die größte deutsche Sprachdatenbank ist COSMAS (*Corpus Search, Management and Analysis System*) am Institut für Deutsche Sprache (IDS) in Mannheim. Ca. 2 Milliarden laufende Wörter Text geschriebener Sprache können hier durchsucht werden. Etwa die Hälfte des Korpus ist öffentlich, das heißt, die erhaltenen Suchergebnisse lassen sich auch exportieren und weiter bearbeiten.

Das Ergebnis der Recherche mit dem Suchwort *Sushis* zeigt Abbildung 1.

Diese Form der Darstellung wird als KWIC (**Key Word In Center**) bezeichnet. Neben der zusätzlichen Möglichkeit der Sortierung z.B. nach dem unmittelbaren Vorgänger kann man auch die Breite der unmittelbaren Umgebung variieren. Eine andere Option ist die Ausgabe im Volltext.

Betrachten wir das Gesamtergebnis, so fällt zunächst auf, dass es überhaupt nur 35 Belege gibt. Bei etwa 1 Milliarde Wörter laufender Text ist das nicht viel. Die Analyse der Einzelbeispiele ergibt, dass es keinen einzigen Beleg für den Genitiv Singular (*des Sushis*) gibt. Die grammatische Beschreibung im Duden (1999) steht auf schwachen Füßen. Im nächsten Schritt wird das Suchwort *Sushi* eingegeben. Eine zusammenfassende Auswertung ergibt:

Suchwort: *Sushis* 35 Treffer

| | |
|---------------|----|
| Singular | 0 |
| Plural | 35 |
| Zweifelsfälle | 0 |

Suchwort: *Sushi* 738 Treffer

| | |
|---------------|-----|
| Singular | 112 |
| Plural | 52 |
| Zweifelsfälle | 574 |

Sushi ist also etwa 20 Mal so häufig wie *Sushis*. Allerdings kann *Sushi* sowohl Singular als auch Plural sein. Die Einzelanalyse der 738 Treffer *Sushi* wird so vorgenommen, dass nur über die grammatische Umgebung entschieden wird, ob es sich um Singular, Plural oder um einen Zweifelsfall handelt. Beispiele wie *ein japanisches Sushi* oder *Sushi ist cool* sind eindeutig Singular, *japanische Sushi* eindeutig Plural. *Er isst Sushi* wird ebenso als Zweifelsfall kategorisiert wie *Sushi und Sashimi sind...* Die Tendenz zum Plural *Sushi* ist offensichtlich, und man kann auch sehen, dass die weitaus häufigste Verwendung die mit Nullartikel ist. Wir haben also über COSMAS Informationen bekommen, die weit über die Wörterbuchangaben hinausgehen.

Abbildung 1: Ergebnis der Anfrage *Sushis* in COSMAS

Datum : Samstag, den 2. Oktober 2004, 9:08:28
 Korpus : public — alle öffentlichen Korpora geschriebener Sprache
 Suchanfrage :
Sushis

KWIC-Übersicht (original/unsortiert)

Anz. Treffer = 35

Angezeigter Kontext: 5 Wörter links, 5 Wörter rechts.

| | | | |
|-----|-----------------------------------|----------------|---------------------------------|
| A99 | voller Sandwiches, Salaten, | Sushis, | Würsten und Unmengen Wein und |
| A99 | Lust auf Capri-Hosen oder auf | Sushis | ab dem Förderband. Doch warum |
| R97 | erhältlich wie „klassische“ | Sushis | mit Lachs, Ei oder Makrele |
| R97 | Circle“ zu verdanken. Sie lernte | “Sushis | vom laufenden Band“ in London |
| R97 | Laufband frisch zubereitete | Sushis | herunter. Kein Warten mehr, bis |
| R97 | zusehen und verfolgen, wie die | Sushis | als Pärchen auf Tellern mit |
| R97 | leeren Teller. Die günstigsten | Sushis | flutschen auf blauen Tellern |
| R97 | — geruhsam zu. Obwohl: An | Sushis | scheiden sich die Geister. Den |
| R97 | und eingelegten Ingwer. Gute | Sushis | krümeln nicht, was man daran |
| R97 | Der Sushi-Circle (Opernplatz) mit | Sushis | vom laufenden Band kam |
| R99 | oder am Fließband sitzen und | Sushis | runterpicken. Mit den |
| R99 | werden, sondern der Belag. Gute | Sushis | krümeln nicht, sie lassen sich |
| E96 | zieht er es vor, die | Sushis | in den eigenen vier Wänden |
| E98 | herum und füllen den Gästen | Sushis | in den Mund, spülen mit |
| M95 | Mattjes und japanische | Sushis, | standen unter musikalischem |
| M95 | Matjes und japanische | Sushis, | standen unter mu- sikalischem |
| K98 | Der Klassiker unter den | Sushis | heißt Nigiri-Zushi, übersetzt |
| K98 | wir die wichtigsten Vertreter der | Sushis | auf:Bei Nigiri-Zushi wird roher |
| K98 | Couscous, die japanischen | Sushis, | die schmackhaften Tortellini |
| L99 | durchaus nicht zum Kinde. Selbst | Sushis | mit wechselndem Geschmack, so |
| N98 | bis hin zu japanischen | Sushis | allerlei „unbritische“ Artikel |
| O97 | verziehen und hier die berühmten | Sushis | des Hauses DO & CO |
| O99 | mit der Herstellung künstlicher | Sushis | und Sashimis beschäftigt — so |
| O99 | mit der Herstellung künstlicher | Sushis | und Sashimis beschäftigt — so |
| P92 | Oper. Und für jene, die | Sushis | und Sashimis nächst der Oper |
| P99 | machen auch wir die | Sushis | ständig frisch“, erklärt der |
| P00 | und geschmacklich feinsten | Sushis | und Sashimis, von der |
| X97 | mit den angeblich besten | Sushis | von ganz Linz ist seit |
| X98 | vom Fließband unbeschränkt mit | Sushis | bedienen. Im Kyoto darf man |
| X98 | sich nur am Abend mit | Sushis | sattessen, mittags werden pro |
| X98 | Stück serviert. Die Qualität der | Sushis | ist in beiden Restaurants recht |
| X98 | das „Running Sushi“ gefallen. | Sushis | und andere mehr oder weniger |
| X98 | Das ist speziell bei den | Sushis | wichtig. Denn diese kalten |
| X98 | Gästen werden jeweils nur sechs | Sushis | serviert, und die sind auch |
| X98 | groß und abwechslungsreich. | Sushis | werden auf dem Fließband mit |

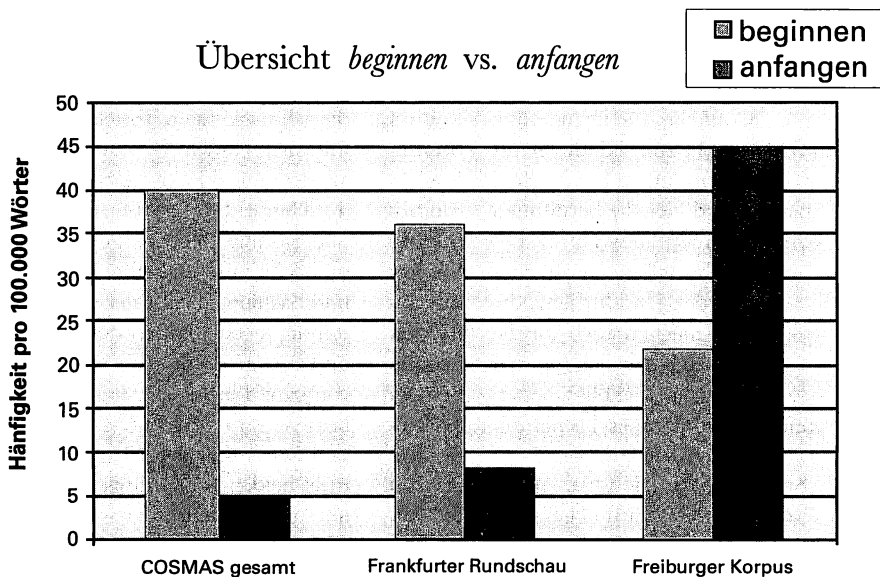
3. *anfangen* vs. *beginnen*

Die zweite sprachliche Herausforderung, über die ich Ihnen heute berichten möchte, liegt im Bereich der Lexik. Ich versuche in meinem Unterricht immer etwa 5 Minuten einzuplanen, die für offene (sprachliche) Fragen reserviert sind. Wenn Sie so wollen, provoziere ich meine sprachlichen Herausforderungen. Eine Frage, die dann einmal plötzlich auftauchte, war: „Was ist der Unterschied zwischen *anfangen* und *beginnen*?“ Wie hätten Sie an meiner Stelle reagiert?

Sie wissen natürlich, was jetzt kommt. Wir schauen bei COSMAS nach.⁴⁾ Bei der Suche nach *anfangen* oder *beginnen* wollen wir natürlich auch Formen wie *fängt...an*, *anzufangen*, *beginnst*, *beginnt*, etc. erfassen. Das ist sehr einfach über den „Lemma-Befehl“ erreichbar. Abbildung 2 zeigt die Ergebnisse etwas aufbereitet.

Der jeweils linke Balken repräsentiert *beginnen* und der rechte *anfangen*. In dieser Darstellung wird nicht mit der absoluten Zahl der Treffer gearbeitet, sondern die Treffer werden auf jeweils 100.000 Wörter laufenden Text umgerechnet. Mit diesem

Abbildung 2: Korpusrecherche *anfangen* vs. *beginnen* in COSMAS I und II



4) Ein Blick in Wörterbücher ist nicht sehr erhellend. Bei Klappenbach/Steinitz (1965ff.) wird zum Beispiel *anfangen* in der 1. Bedeutung mit *beginnen* erklärt.

Verfahren kann man auch die Ergebnisse von verschiedenen Korpora mit einander vergleichbar machen.

Ein Blick auf *COSMAS gesamt* zeigt, dass in der geschriebenen deutschen Gegenwartssprache *beginnen* etwa acht Mal so häufig ist wie *anfangen*. Die Überprüfung eines Teilkorpus von COSMAS, einige Jahrgänge der Zeitung *Frankfurter Rundschau*, bestätigen den deutlichen Frequenzunterschied. Das Freiburger Korpus ist eine Sammlung von transkribierten Texten gesprochener Sprache, die Ende der 60er/Anfang der 70er Jahre im Rahmen von Projekten zur Gesprächsanalyse veröffentlicht wurden⁵⁾. In COSMAS I, das im März 2003 vom Netz genommen wurde, war das Freiburger Korpus noch ein Teilkorpus. Die hier vorgelegten Ergebnisse für das Freiburger Korpus beruhen auf einer Recherche, die im November 2002 durchgeführt wurde. Die Daten des Freiburger Korpus zeigen, dass *anfangen* in der gesprochenen Sprache etwa doppelt so häufig ist wie *beginnen*. Unser erstes Ergebnis ist also, dass der Unterschied zwischen *anfangen* und *beginnen* ein Unterschied zwischen gesprochener und geschriebener Sprache ist⁶⁾. Wechseln wir hier kurz von der deskriptiven Ebene auf die präskriptive Ebene. Man könnte aus dem vorliegenden Ergebnis eine sehr einfache Empfehlung für Anfänger ableiten: Verwende in der 1. und 2. Person *anfangen* und in der 3. Person das Verb *beginnen*.

Doch wir können in der Analyse noch einen Schritt weiter gehen. In COSMAS ist es möglich, eine Kollokationsanalyse (Kookurrenzanalyse) durchzuführen. Dabei wird die statistisch (hoch)signifikante Umgebung des Suchausdrucks erfasst. Die Ausgabeform der Ergebnisse ist ein bisschen gewöhnungsbedürftig, und so habe ich die typischsten Verbindungen der Verben *anfangen* und *beginnen* in einer einfachen Liste zusammengefasst⁷⁾:

Sehr typische Kollokationen mit *beginnen*:

- um ... Uhr beginnen

5) Vgl. <http://www.ids-mannheim.de/pub/abgeschlossen/heutig/titel3.html> ; das Freiburger Korpus ist jetzt der Organisationseinheit „Korpusstelle für gesprochenes Deutsch“ des IDS in Mannheim zugeordnet. Über das Internet ist die DGD (Datenbank gesprochenes Deutsch) leicht zugänglich <http://www.idsmannheim.de/ksgd/>. Leider ist es zur Zeit noch nicht möglich, einfach nach Lemmata zu suchen. Das war auch der Grund, warum das Ergebnis einer Recherche in Cosmas I vom November 2002 in diese Übersicht einbezogen wurde.

6) Natürlich kann man auch anführen, dass *anfangen* ein trennbares Verb ist und *beginnen* nicht.

7) Zur Kollokationsanalyse, die bei COSMAS von Cyril Belica entwickelt wurde und jetzt Kookurrenzanalyse heißt, etwas ausführlicher <http://www.idsmannheim.de/kt/misc/tutorial.html> (Stand 22.4.2004). Ein Teil der Kookurrenzanalyse in COSMAS ist auch ohne Registrierung und Client-Server-System als Kookurrenzdatenbank zugänglich unter <http://corpora.ids-mannheim.de/ccdb/>.

- am (+ Wochentag) beginnen
- heute/morgen/..... beginnen
- (da)mit beginnen

Sehr typische Kollokationen mit *anfangen*:

- (damit) viel/etwas/ wenig/ nichts anfangen können
- von vorne/bei Null/neu ... anfangen
- angefangen (hat alles mit)
- anfangen, zu + Infinitiv

Damit können Lerner sicher etwas anfangen⁸⁾.

Unter Forschungsgesichtspunkten könnte man noch einmal weiter gehen. Es gibt nämlich einen auffälligen Unterschied zwischen *anfangen* und *beginnen* bei der Verwendung des Partizips I. Während fast keine Belege für *anfangend** existieren, sind die Belege für *beginnend** mit mehr als 5% (innerhalb der Lemmaformen) recht häufig. Das ist ein Befund, der wohl nicht so einfach bei der Besinnung auf das Sprachgefühl auftaucht. Kurz: Es sind noch sehr viele Dinge zu entdecken.

Die zwei Fallbeispiele sollten zeigen, dass die Korpuslinguistik eine große Hilfe für den Unterricht Deutsch als Fremdsprache bietet. Wir müssen uns als Lehrer oder Wissenschaftler nicht mehr auf unsere Intuition oder auf Nachschlagewerke verlassen. Wir können uns anhand von großen Mengen authentischen Sprachmaterials unmittelbar vergewissern. Noch viel wichtiger scheint mir zu sein, dass wir einen ganz anderen Blick auf die Sprache bekommen. Wir leiden als Menschen, die sich mit Sprache beschäftigen, unter einer gewissen *déformation professionnelle*. Der ausgefallene Ausdruck, die überraschende sprachliche Fügung ist interessanter als die gewöhnliche Verbindung. Korpora weisen uns mit Macht auf die häufigsten und absolut normalsten sprachlichen Mittel hin. Und diesen Blick brauchen wir besonders für unsere Studenten, die anfangen, Deutsch zu lernen.

4. Hochfrequente unregelmäßige deutsche Verben

Es gibt viele didaktische Herausforderungen. Eine davon ist die Wort-

8) Eine Einschränkung ist allerdings notwendig. Die Kollokationsanalyse wurde in COSMAS II durchgeführt. Die sehr typischen Kollokationen für *anfangen* beziehen sich nur auf die **geschriebene** Sprache.

schatzvermittlung⁹⁾. Es sind damit eine ganze Menge grundlegender Fragen verbunden. Ich kann hier die zwei wichtigsten fachlichen Kontroversen nur kurz skizzieren.

Bei der einen Kontroverse geht es um die **Lerntheorie** im weiten Sinne. Wenn man im Anschluss an Krashen (1987) zwischen Spracherwerb und Sprachlernen unterscheidet und davon ausgeht, dass Spracherwerb in einer natürlichen Reihenfolge unbewusst abläuft, genügt es dann nicht, den gerade dem Lernniveau angemessenen Input zu geben? Konkreter gesagt: Ist nicht zum Beispiel extensive Lektüre zum Spaß viel wirksamer als die bewusste Auseinandersetzung mit sprachlichem Material in Form von Wortschatzlisten, mnemotechnischen Verfahren oder anderen Strategien zur Aneignung von Wortschatz?

Bei der anderen Kontroverse geht es um die **Auswahl des Wortschatzes** und seine Einbettung. Man kann die Wortschatzauswahl sprachimmanent treffen und dann spielen Frequenzgesichtspunkte die entscheidende Rolle. Man kann sich aber auch auf den Standpunkt stellen, dass dem sprachlichen Handeln in Situationen eine übergeordnete Rolle gegeben werden muss. Die Selektion geht dann von der Auswahl der relevanten Situationen über die relevanten Sprachhandlungen zu den relevanten sprachlichen Mitteln. Dieser Ansatz wurde zuerst in *Kontaktschwelle Deutsch als Fremdsprache* in Baldegger/u.a. (1980) vertreten und ist auch in der deutschen Ausarbeitung des europäischen Referenzrahmens *Profile*, Glaboniat/ u.a. (2002), realisiert.

Ich persönlich glaube, dass die Extrempositionen in den beiden skizzierten Kontroversen wichtige Säulen für den Unterricht Deutsch als Fremdsprache darstellen. Es gibt beim Sprachenlernen sowohl bewusste als auch unbewusste Prozesse, und eine Auswahl von Vokabular muss sowohl unter sprachimmanenten als auch unter situativen Gesichtspunkten vorgenommen werden. Allerdings, und das ist wichtig, sollten die Auswahlkriterien transparent und möglichst empirisch fundiert sein.

Eine wichtige Aufgabe im Anfängerunterricht Deutsch als Fremdsprache ist, hochfrequenten Wortschatz bewusst zu vermitteln¹⁰⁾. Entsprechend sollte die Auswahl des Lernmaterials gestaltet sein. Lernern sollte zudem ganz klar gemacht werden, dass Wort nicht gleich Wort ist, auch wenn zwei Wörter im Wörterbucheintrag etwa

9) Dass Fremdsprachenunterricht nicht auf Grammatik- und Wortschatzvermittlung reduziert werden kann, ist selbstverständlich. Die Vermittlung von kulturellem Wissen i. w. S., von sprachlichen Fertigkeiten und Strategien und die affektive Komponente im Unterricht sind ebenfalls zentral.

10) Vgl. auch Nation (2001, 16).

gleichen Umfang aufweisen. Die Frequenzunterschiede können riesig sein.

In Deutschlehrwerken für Anfänger, die nur für den japanischen Markt hergestellt wurden, gibt es meistens eine Liste mit unregelmäßigen Verben. Die Auswahl der Verben ist recht willkürlich. Welche von den bei Helbig/Buscha (2001, 36ff.) verzeichneten 192 unregelmäßigen Verben im weiteren Sinne sollten aufgenommen werden? Hier spielen wohl Tradition, Intuition und Geschmack eine wichtige Rolle. Sie wissen schon, was kommt — der Blick ins Korpus. Die Liste in Abbildung 3 ist das modifizierte Ergebnis einer ausführlicheren Recherche.

Abb. 3: Vorschlag für eine frequenzbasierte Auswahlliste deutscher unregelmäßiger Verben

Die 50 wichtigsten unregelmäßigen deutschen Verben (W. Lange, Stand 24.5.04)

| Nr. | Infinitiv | Bedeutung | Präsens | Präteritum | Partizip II | Frequenz* |
|-----|-----------|-----------|---------|------------|-------------|-----------|
| 1 | sein | ある, ~である | ist | war | gewesen | 11091900 |
| 2 | werden | ~になる | wird | wurde | (ge)worden | 8908532 |
| 3 | haben | もっている | hat | hatte | gehabt | 6149877 |
| 4 | können | ~できる | kann | konnte | (können) | 2390689 |
| 5 | sollen | ~すべきである | soll | sollte | (sollen) | 1677580 |
| 6 | geben | 与える | gibt | gab | gegeben | 1505027 |
| 7 | wollen | 欲する | will | wollte | (wollen) | 1413715 |
| 8 | müssen | ~ねばならない | muss | musste | (müssen) | 1408339 |
| 5 | kommen | 来る | kommt | kam | gekommen | 1173951 |
| 10 | gehen | 行く | geht | ging | gegangen | 992598 |
| 11 | stehen | 立っている | steht | stand | gestanden | 936861 |
| 12 | lassen | ~させる | lässt | ließ | (lassen) | 778215 |
| 13 | finden | 見つける | findet | fand | gefunden | 686661 |
| 14 | sehen | 見る | sieht | sah | gesehen | 644295 |
| 15 | bleiben | とどまる | bleibt | blieb | geblieben | 597879 |
| 16 | bringen | もって来る | bringt | brachte | gebracht | 526196 |
| 17 | liegen | 横たわっている | liegt | lag | gelegen | 494563 |
| 18 | nehmen | 取る | nimmt | nahm | genommen | 489252 |
| 19 | dürfen | ~してもよい | darf | durfte | (dürfen) | 440455 |
| 20 | halten | 保つ, 止まる | hält | hielt | gehalten | 439860 |
| 21 | sprechen | 話す | spricht | sprach | gesprochen | 295600 |
| 22 | tun | する | tut | tat | getan | 288592 |
| 23 | bieten | 提供する | bietet | bot | geboten | 287296 |
| 24 | fallen | 落ちる | fällt | fiel | gefallen | 271785 |
| 25 | ziehen | 引く, 移動する | zieht | zog | gezogen | 265246 |

| Nr. | Infinitiv | Bedeutung | Präsens | Präteritum | Partizip II | Frequenz* |
|-----|-----------|-----------|-----------|------------|-------------|-----------|
| 26 | treffen | 当たる, 会う | trifft | traf | getroffen | 256929 |
| 27 | wissen | 知っている | weiß | wusste | gewusst | 207941 |
| 28 | laufen | 走る | läuft | lief | gelaufen | 206532 |
| 29 | fahren | (乗り物で) 行く | fährt | fuhr | gefahren | 204583 |
| 30 | steigen | 登る | steigt | stieg | gestiegen | 194922 |
| 31 | schreiben | 書く | schreibt | schrieb | geschrieben | 178946 |
| 32 | tragen | 運ぶ | trägt | trug | getragen | 175579 |
| 33 | schlagen | 打つ | schlägt | schlug | geschlagen | 173022 |
| 34 | schließen | 閉じる | schließt | schloss | geschlossen | 157883 |
| 35 | denken | 考える | denkt | dachte | gedacht | 148457 |
| 36 | helfen | 助ける | hilft | half | geholfen | 141897 |
| 37 | kennen | 知っている | kennt | kannte | gekannt | 121539 |
| 38 | sitzen | 座っている | sitzt | saß | gesessen | 120781 |
| 39 | sterben | 死ぬ | stirbt | starb | gestorben | 102637 |
| 40 | rufen | 呼ぶ | ruft | rief | gerufen | 91308 |
| 41 | lesen | 読む | liest | las | gelesen | 88957 |
| 42 | singen | 歌う | singt | sang | gesungen | 60311 |
| 43 | fliegen | 飛ぶ | fliegt | flog | geflogen | 56856 |
| 44 | essen | 食べる | isst | aß | gegessen | 29211 |
| 45 | fangen | 捕らえる | fängt | fang | gefangen | 28555 |
| 46 | trinken | 飲む | trinkt | trank | getrunken | 26622 |
| 47 | schneiden | 切る | schneidet | schnitt | geschnitten | 22196 |
| 48 | schlafen | 眠る | schläft | schief | geschlafen | 20595 |
| 49 | schwimmen | 泳ぐ | schwimmt | schwamm | geschwommen | 12658 |
| 50 | waschen | 洗う | wäscht | wusch | gewaschen | 3262 |

* in 935 Millionen Wörtern Zeitungstext; Quelle: COSMAS - Datenbank des Instituts für Deutsche Sprache

Ich bin mir über die Problematik einer solchen Liste sehr bewusst. Hier stichpunktartig einige Hinweise und Kommentare:

- Die Frequenzanalysen wurden in COSMAS II, also einem Korpus für **geschriebene Sprache**, durchgeführt. Nachdem es noch kein umfangreiches Korpus für gesprochene deutsche Sprache gibt, ist das eine Notlösung¹¹⁾.

11) In dem Korpus der Akademie Berlin Brandenburg, das gerade aufgebaut wird, gibt es ein Teilkorpus *gesprochene Sprache*, das im 1. Halbjahr 2005 ans Netz gehen soll. Doch sowohl die Auswahlkriterien (u.a. auch historische politische Reden) als auch der Umfang (im Sept. 2004 waren es nur ca. 3,5 Mio. laufende Wörter von geplanten 10 Mio. Wörtern) machen einen skeptisch, ob eine wirkliche Hilfe in Fragen der gesprochenen deutschen Gegenwartssprache zu erwarten ist.

- Die Liste ist nicht das Ergebnis einer reinen Frequenzanalyse. Es wurden nur Simplicia ausgewählt, die Wortbildungsprozessen zugänglich sind. So schieden zum Beispiel *gewinnen* und *verlieren* — obwohl hochfrequent — aus. In COSMAS sind Verben wie *schlafen* oder *waschen* recht selten. In die Liste wurden alle Verben aufgenommen, die in der deutschen Ausarbeitung des europäischen Referenzrahmens, Glaboniat/u.a. (2002), auf der Stufe A1 oder A2 **aktiv** beherrscht werden sollten.
- Die absoluten Zahlen beziehen sich auf die aufgeführten **Wortformen**. Beim Eintrag *fahren* beispielsweise ergibt die Summe der Wortformen (*fahren, fährt, fuhr, gefahren*) den Wert 204583. Die Wortform *fährt* in einem Ausdruck wie *fährt ...ab* ist dabei mitgezählt.
- Die **3. Person Singular Präsens** wurde immer aufgeführt, da sie innerhalb der Lemmaformen meist auf Rang 1 liegt.
- Das **Partizip II von Modalverben** (*gekonnt, gemusst, gedurft*) ist in COSMAS extrem selten. Deshalb wurde es nicht aufgenommen. Die wirklich häufige Form bei der Perfektbildung der Modalverben, der Infinitiv, wurde unter der Rubrik, *Partizip II* in Klammern gesetzt.
- Die Bedeutungsangabe im Japanischen ist wirklich rudimentär. Hier gibt es einen Konflikt zwischen Kürze und Genauigkeit. Das Problem ist auch deshalb besonders groß, weil mit höherer Frequenz eines Wortes die Zahl der verschiedenen Bedeutungen zunimmt. Die Auswahl der „Haupt-bedeutung“ wurde intuitiv vorgenommen und ist nicht über eine Korpusrecherche abgesichert.
- Nachdem Wortschatzlernen sowohl in die Breite als auch in die Tiefe ein kumulativer Prozess ist, sollte die Liste als ein erster Schritt für die Lerner betrachtet werden, auf den dann Differenzierungen folgen.
- Geplant ist die Liste als Hypertext, was erlaubt, auch korpusbasierte typische Verwendungsbeispiele einzubinden.

5. Die wichtigsten deutschen Sprachdatenbanken in Kürze

Es gibt neben COSMAS auch andere deutsche Sprachdatenbanken. Eine ausführliche Darstellung des Aufbaus, der Recherchemöglichkeiten und der Probleme würde hier zu weit führen. Ich möchte Ihnen in Abbildung 4 übersichtsartig nur die wichtigsten Informationen geben.

Abb. 4: Vergleich zwischen den drei wichtigsten deutschen Sprachdatenbanken

Grober Vergleich der drei wichtigsten Korpora deutscher Sprache

(Stand September 2004)

| Korpusbezeichnung | COSMAS | Wortschatz Leipzig | DWDS (im Aufbau) |
|-------------------|---|---|---|
| Institution | Institut für deutsche Sprache (IDS) | Universität Leipzig Institut für Informatik | Akademie der Wissenschaften Berlin-Brandenburg |
| Adresse | http://www.idsmannheim.de/cosmas2/ | http://wortschatz.informatik.unileipzig.de/index_js.html | http://www.dwds.de/ |
| Umfang Wörter | gesamt ca. 2 Mrd. öffentlich ca. 1 Mrd | ca. 500 Mio. | Kern ca. 100 Mio., Ergänzung ca. 1 Mrd. |
| Aufbauprinzip | Begrenzung auf geschriebene Sprache, nicht repräsentativ | opportunistisch | Kern: repräsentativ, Ergänzungskorpus: opportunistisch |
| Zugänglichkeit | Registrierung notwendig, Client-Server-System | direkt | Registrierung teilweise notwendig |
| Stärken | sehr flexible linguistische Suche | schnelle, einfache Suche und vielseitige Ergebnisse | Repräsentativität des Kernkorpus |
| Schwächen | komplexe Suchsyntax | keine Lemmasuche | im Aufbau |

6. Unterlassungen und Ausblick

In den Ausführungen wurde vieles versäumt. Ich habe nichts zur Geschichte und zu den Grundlagen der Korpuslinguistik gesagt¹²⁾. Es fehlten auch die notwendigen kritischen Anmerkungen zu den korpuslinguistischen Verfahren bzw. zu deren Begrenzung. Schließlich habe ich kein Wort verloren über die vielen anderen Anwendungsmöglichkeiten, die die Korpuslinguistik in der Fremdsprachenvermittlung finden kann¹³⁾.

Ich wollte Ihnen nur veranschaulichen, dass wir als Lerner, Lehrer und als Wissenschaftler heute eine faszinierende Chance haben. Wir haben extrem schnellen,

12) Besonders hilfreich ist hier Biber (1998).

13) Einige Ideen gibt Mukherjee (2002).

breiten und feinen Zugriff auf die deutsche Sprache — so wie sie tatsächlich verwendet wird.

Bibliografie:

- Baldegger, Markus/ u.a. (1980): Kontaktschwelle Deutsch als Fremdsprache. Berlin (Langenscheidt)
- Biber, Douglas/ u.a. (1998): Corpus linguistics. Investigation language structure and use. Cambridge (University Press)
- Duden (1999). Das große Wörterbuch der deutschen Sprache in zehn Bänden. 3. Aufl., Mannheim (Dudenverlag)
- Europarat (Hg.) (2001): Gemeinsamer europäischer Referenzrahmen für Sprachen: lernen, lehren, beurteilen. Berlin (Langenscheidt)
- Glaboniat, Manuela/ u.a. (2002): Profile deutsch. Gemeinsamer europäischer Referenzrahmen. Berlin (Langenscheidt)
- Helbig, Gerhard/Buscha, Joachim (2001). Deutsche Grammatik. Ein Handbuch für den Ausländerunterricht. Berlin (Langenscheidt)
- Klappenbach, Ruth/ Steinitz, Wolfgang (1964ff.): Wörterbuch der deutschen Gegenwartssprache. 6 Bände, Berlin (Akademie-Verlag)
- Krashen, Stephen D. (1987): Language Acquisition and Language Learning. Englewood Cliffs (Prentice Hall)
- Kron, Friedrich, W. (1994): Grundwissen Didaktik. 2. Aufl., München (Reinhardt)
- Lange, Willi (2005): „Sushi am laufenden Band — in Enzyklopädien, Wörterbüchern und Sprachdatenbanken“ — in: 文化論集 (bunka ronshu) 26, Tokyo (早稲田商学同攻会) (Waseda Shougakubu Doukoukai), im Druck
- Mukherjee, Joybrato (2002): Korpuslinguistik und Englischunterricht. Eine Einführung. Frankfurt a. M. (Lang)
- Nation, Paul I. S. (2001): Learning Vocabulary in Another Language. Cambridge (University Press)
- Schmitt, Norbert (2000): Vocabulary in Language Teaching. Cambridge (University Press)

Internetadressen der 3 wichtigsten deutschen Sprachdatenbanken

COSMAS: <http://www.ids-mannheim.de/cosmas2/>

Wortschatz Leipzig: http://wortschatz.informatik.uni-leipzig.de/index_js.html

DWDS: <http://www.dwds.de/>