

Silvie Cinková

Belegsuche bei der lexikographischen Bearbeitung von selten gebrauchtem Wortschatz

1. Einleitung

Mit der Diskrepanz zwischen Wörterbuchangaben und Resultaten von korpusbelegten lexikalischen Studien haben sich schon viele Wissenschaftler befasst, so z.B. SINCLAIR (1991). Die hier dargestellten Analysen zu diesem Thema entstanden bei der Arbeit an Wörterbuchartikeln für das große deutsch-tschechische Wörterbuch, das am Institut für germanische Studien der Karlsuniversität zu Prag verfasst wird. Das Wörterbuchprojekt ist auf neun Jahre angelegt und wird von der „Grant Agency of the Czech Republic – Czech Science Foundation“ unterstützt. Das entstehende Wörterbuch soll 130.000 Lemmata enthalten und sowohl in gedruckter als auch in elektronischer Form erscheinen. Es richtet sich vor allem an tschechischsprachige Wissenschaftler, Experten, Studierende und Übersetzer. Der Hauptgedanke bei der Lemmataauswahl und -gestaltung ist:

Das Wörterbuch soll nicht nur den lebendigen deutschen Sprachgebrauch, nicht nur den Wortschatzkern der wichtigsten wissenschaftlichen Bereiche abdecken, sondern auch die Wörter auflisten, die für diejenigen unentbehrlich sind, die sich mit der deutschen Literatur und Geschichte befassen. [...] Das oberste Gliederungsprinzip bilden dabei die Wortarten. Akzentuiert werden besonders der gegenwärtige Usus, die Umgangssprache, die Pragmatik und die Varietäten (stilistische Markierung). Produktive deutsche Wortbildungsmorpheme werden inventarisiert und an Beispielen dargestellt. Die Wörterbuchartikel werden zudem durch kontrastiven Sprachvergleich angereichert. (VACHKOVÁ 2002)

Die Grundlage für die Arbeit an diesem Wörterbuch bilden neben den Textkorpora COSMAS (IDS Mannheim) und SYN2000 (Institute of the Czech National Corpus Prag) zum Teil bekannte deutsche einsprachige Wörterbücher, vor allem das Duden-Universalwörterbuch auf CD-ROM und der Duden in zehn Bänden, außerdem wurden der Schülerduden in zwölf Bänden und Langenscheidts Großwörterbuch DaF ebenso wie tschechische Nachschlagewerke, etwa tschechische Bedeutungswörterbücher und das deutsch-tschechische sowie das tschechisch-deutsche Wörterbuch von SIEBENSCHNEIDER herangezogen. (VACHKOVÁ 2002). Das geplante Glossar orientiert sich vorläufig an demjenigen des Dudens auf CD-ROM.

Beim Verfassen der Wörterbuchartikel wird korpusorientiert vorgegangen. Korpusdaten werden stets mit bereits existierenden Nachschlagewerken verglichen, um die Übersicht über die einzelnen Bedeutungen zu behalten. Korpusbelege werden jedoch für ausschlaggebend gehalten, daher werden Angaben aus bestehenden Nachschlagewerken nicht selten ergänzt, umorganisiert oder revidiert. Das tschechische Korpus SYN2000 wird benutzt, um möglichst gute, kollokationsgerechte Übersetzungen der deutschen Belege zu sichern; die meisten Informationen werden jedoch dem deutschen Korpus COSMAS I entnommen, der „weltweit größten Sammlung von deutschsprachigen Textkorpora als empirischer Basis für die linguistische Forschung“¹ mit fast 2 Milliarden Textwörtern. Der ungeheure Umfang dieses „Archivs der geschriebenen Sprache“ macht es möglich, allgemein verbreitete Wörter in einer Vielfalt von Kontexten zu analysieren, was besonders bei der Ermittlung von Valenzstrukturen bei Verben und Adjektiven hilfreich ist. Selten gebrauchte und veraltete Wörter können dank des Korpus in der Regel zumindest einmal belegt werden.

Bei der Erarbeitung der Verbartikel (Buchstaben T und W) ist man in dem vorläufigen Glossar jedoch häufiger auf Verben gestoßen, die im „Archiv der geschriebenen Sprache“ nicht oder nur sehr wenig belegt waren. Auch in bereits bestehenden Wörterbüchern waren diese Lexeme nicht zu finden, von einer systematischen Beschreibung ihrer Valenz und ihres Kollokationspotentials ganz zu schweigen. Des Weiteren gab es Lexeme, bei denen zumindest eines der genannten Lexika einen

¹ Ausführliche Informationen sind unter <http://www.ids-mannheim.de/kt/corpora.shtml> zugänglich.

Beispielsatz bot, allerdings waren die Belege im Korpus mit dem Wörterbuchbeispiel auf keinen gemeinsamen Nenner zu bringen. Die ungenügende Belegbarkeit dieser Verben im Korpus und die Ungewissheit bei der Äquivalentsuche machten es nötig, zusätzliche Quellen hinzuzuziehen. Das Internet schien eine ergiebige, effektiv durchsuchbare Quelle zu sein. Bei der Suche wurde die Suchmaschine Google (www.google.com) benutzt.

2. Selten belegte Verben

Von den bisher erarbeiteten Verben kommen im „Archiv der geschriebenen Sprache“ u.a. folgende nicht vor: *tentieren*, *testieren*, *theologisieren*, *titschen*, *trappen*, *werweißen*, *wuscheln*, *wuzeln*. Hier wurden Belegstellen im Internet gesucht. Es war zu vermuten, dass die im Duden als „veraltet“ und „landschaftlich“ markierten Verben wie *tentieren*, *trappen* oder *werweißen* auch im Web kaum zu finden sind. Dies ist allerdings nicht richtig. Bei einem Verb wie *wuscheln* ist hingegen erstaunlich, dass es im Korpus nicht belegt ist, wenn das Web davon – erwartungsgemäß – wimmelt. Die Suche im Web ergab folgende Beobachtungen, die hier im Vergleich mit entsprechenden Wörterbuchartikeln des Duden dargestellt werden.

2.1. tentieren:

COSMAS: keine Belege

Duden: Im Duden waren keine Belege für *tentieren* vorhanden. Es wurden jedoch zwei Bedeutungen angeführt:

1. (veraltet, noch landsch.) a) *untersuchen*, *prüfen*; b) *versuchen*, *unternehmen*.

2. (österr., ugs.) *beabsichtigen*

Im Fremdwörterduden war der Erklärungsteil fast identisch, nur noch um die Verben *betreiben*, *arbeiten* bereichert, und es wurde auf das lateinische *tentare* und das deutsche *Tentakel* hingewiesen.

Google: Im Web erscheint – laut www.google.com vom 10.02.2003 – die Form *tentieren* etwa 40-mal (eigentlich 67-mal, davon sind

jedoch 22 Konkordanzen unbrauchbar, weil es sich um Suchfehler wie *pa-tentieren* usw. handelt.) In manchen Konkordanzen ist das gesuchte Wort nicht zu sehen, ohne das jeweilige Dokument zu öffnen. Es blieben die Suchresultate:

- *tentierten*: 10-mal (mindestens 22 aus 69 unbrauchbar),
- *tentierst*: einmal,
- *tentiert*: 30-mal (mindestens 32 aus 67 unbrauchbar),
- *tentierte*: keine Treffer (aus 86 wohl alles unbrauchbar),
- *tentiere*: 8-mal (aus 11 mindestens 3 unbrauchbar).

Insgesamt sind bei Google etwa 90 Konkordanzen belegt. Die Hauptquelle waren Texte von Schriftstellern des 19. Jahrhunderts wie Hugo von Hofmannsthal, Willibald Alexis, Johann Nestroy und Fanny Lewald.² Des Weiteren fand sich das Wort in verschiedenen historischen Quellen³ wie dem Mirakelbuch Richter⁴, in einer lokalen Chronik von Weberberg⁵ und in einem Ausschnitt des – noch älteren – Deutschen Rechtswörterbuchs⁶.

In manchen dieser Konkordanzen tritt die Bedeutung *zu etwas neigen, sich versucht fühlen, etwas zu tun* ganz deutlich hervor, die auf das lateinische Wort *tentatio/temptatio (Versuchung)* zurückgeht.⁷ Diese Bedeutung wird im Duden doch nur bei *Tentation* und *Tentator* angeführt.

Unter den 15 am besten nutzbaren Konkordanzen tauchten überraschenderweise auch zeitgenössische Quellen auf, doch hatten die Belege hier die Bedeutung *zu etwas neigen* (also abweichend von der Duden-Bedeutung *beabsichtigen*). Um ein Beispiel zu nennen: *Das*

² www.nestroy.at/stuecke_materialien/52_eisenbahnheiraten/eisenbahn_screen.pdf. <http://www.gutenberg2000.de>, www.nestroy.at/stuecke_materialien/58_unbedeutende/58text.html, [zit. 2003-02-10].

³ Liselotte von der Pfalz: <http://www.physiologus.de/sprache.htm>. [zit. 2003-02-10].

⁴ Mirakelbuch Richter 1738, S. 306, Nr. 22: <http://members.aol.com/masteinba/texte/fuehrer/fresk4.htm>. [zit. 2003-02-10].

⁵ www.weberberg.de/weberberg/alltag.html. [zit. 2003-02-10].

⁶ Ausschnitt aus dem Deutschen Rechtswörterbuch. Kantor (A II 2 e): www.rzuser.uni-heidelberg.de/~cd2/drw/a/K51.htm. [zit. 2003-02-10].

⁷ z.B. bei Alexis: „Die Person hat nun mal was, daß man nicht weiß, was es ist [...]“ / „Grad das könnte mich tentieren“, rief der Kornett. „s ist nur, sie ist nicht nach meinem Gout.“ / „Wolfskehl liebt nur das Bornierte. Da oben sitzt die neuste, die er auf dem Zug hat.“ / „Man schaute nach der Loge im zweiten Range, nicht aber mit Diskretion, wo Walter van Asten hinter seiner Cousine stand.“ <http://www.gutenberg2000.de/alexis/ruhe/ruhe314.htm>. [zit. 2003-02-10].

*Graugouti ist genetisch gesehen ein Schwarz-Creme-Agouti. Der Cremeton soll allerdings leicht zu Buff tentieren.*⁸

Hier kann es sich um eine Interferenz zwischen den Lexemen *tentieren* und *tendieren* handeln. Diese war mehrmals zu beobachten und findet sich sogar in einem Rechtschreibglossar online⁹: *tendieren (zu etwas neigen) vgl. tentieren; [...] tentieren (beabsichtigen) vgl. tendieren*. Durch diese Interferenz kann dann auch solch ein isolierter Gebrauch zustande kommen: *Bin gerade am überlegen welchen Mixer ich mir hole. Ich tentiere zwischen Sk-one und sk-six.*¹⁰

2.2. testieren

COSMAS: keine Belege

Duden: [lat. testari = bezeugen, Zeuge sein, zu: testis = Zeuge]
 1a) (Hochschulw. früher) *ein Testat für etwas geben*: eine Vorlesung im Studienbuch testieren lassen;
 b) (bildungsspr.) *attestieren* (1): jmdm Leistungen testieren.
 2. (Rechtsspr.): *ein Testament errichten*.

Google:
 • *testieren*: 2130-mal,
 • *testierten*: 1170-mal,
 • *testierte*: 1750-mal,
 • *testiere*: 42-mal.

Alle im Duden angeführten Bedeutungen waren in den Konkordanzen vertreten. Um die Richtigkeit der Behauptung „früher“ bei (1a) zu testen, wurde noch eine Suche mit den Suchwörtern *testieren* und *Kurse* unternommen. Es wurden 462 Okkurrenzen gefunden, lauter typische

⁸ Die Farben der Meerschweinchen. www.beepworld.de/members21/rodentia/farben.htm. [zit. 2003-02-10]. Ähnlich z.B.: „Die Entscheidung wäre mir sehr schwer gefallen, aber tentiert hätte ich zum 'Quadrat'“. www.dogge-atlantis.at/tb_02.11.htm. [zit. 2003-02-10]. Weiter auch: „Die Mehrheit tentiert zum Chor und Richtung Band, modern und volkstümlich. Es wird am 9. Okt. ein Treffen mit den Musikern geben...“ [online, PDF] 6. Sitzungsprotokoll des Arbeitskreises Zukunftswerkstatt. www.teugn.de/5Agenda21/3AK-Zukunft/AKZ006.pdf. [zit. 2003-02-10].

⁹ Neue deutsche Rechtschreibung per Mausclick: <http://www.tastschreiben.de/p0600280.htm>. [zit. 2003-02-10].

¹⁰ Forum: DJing: Producing/Equipment Forum. Thema: Sk-one or Sk-six. www.mzee.com/forum/archiv/17/2002/01/2/14070.html. [zit. 2003-02-10].

Kontexte wie: *Die Studierenden lassen sich alle besuchten Kurse, Projekte, Praktikas, Workshops und Unterrichtseinheiten testieren.*¹¹ Die Suchanfrage *bescheinigen* und *Kurse* ergab zwar 1630 Resultate, jedoch keine eindeutigen, da *Kurse* häufig in der Bedeutung *Valutakurse* auftreten. Bei der Einschränkung der Suchanfrage auf *bescheinigen* und *Kurse* und *Universität* wurde 493 Stellen gefunden, die ziemlich plausibel wirkten. Eine typische Konkordanz wäre z.B.: *Interessierte, die kein ganzes Studium leisten möchten, können sich auch einzelne Kurse bescheinigen lassen.*¹² Dieses Ergebnis stellt die Bezeichnung „veraltet“ des Duden in Frage, denn alle Seiten mit dieser Thematik sind in den letzten Jahren entstanden, obwohl *bescheinigen* insgesamt doch üblicher zu sein scheint.

2.3. theologisieren

COSMAS: keine Belege

Duden : etw. unter theologischem Aspekt erörtern.

Google: Die Form *theologisieren* kommt etwa 216-mal vor, dabei sind allerdings die meisten Okkurrenzen Substantive (*das Theologisieren*). Es überwiegt die Kollokation *Theologisieren mit Kindern und Jugendlichen*. In COSMAS erscheint die substantivierte Form auf *-en* gar nicht, wobei das Korpus zumindest fünf Belege zu *Theologisierung* enthält. Die Form *theologisierten* kommt bei Google 12-mal vor, allerdings meistens attributiv. Dasselbe gilt für die Form *theologisierte* (30-mal). Die Form *theologisiere* kommt 3-mal vor. Die Bedeutung entspricht etwa der im Duden angegebenen.

2.4. titschen

COSMAS: keine Belege

Duden: sw.V., hat [ablautende Nebenform von tatschen] (ostmd.): eintauchen, -tunken.

¹¹ Informationen - Musikpädagogischer Schwerpunkt - MWZ - HMT. www.hmtzh.ch/musik/studium/stg/mps.html. [zit. 2003-02-10].

¹² Beruf und Uni unter einem Hut. www.uni-hannover.de/en/aktuell/pressein/2000/2000presseinfo_78.htm. [zit. 2003-02-10].

- Google:
- *titschte*: 44-mal,
 - *titscht*: 96-mal,
 - *titschten*: 7-mal.

Es erwies sich als wenig ergiebig, nach der infiniten Form und der Form *titschen* zu suchen, da zu viele andere Belegstellen gefunden wurden (vor allem Eigennamen, denn Google unterscheidet nicht zwischen Groß- und Kleinschreibung). Die Form *getitscht* kam etwa 24-mal vor, dabei nur einmal in der im Duden angegebenen Bedeutung: *Wir blieben ungefähr 20 Minuten im Wasser. In der Zeit waren wir um die Wette getaucht, haben uns mit Wasser vollgespritzt oder gegenseitig getitscht. Zweimal kam ein Kontext mit Ostereiern vor: ... am Frühstückstisch schlägt man die bunten Ostersymbole vor dem Essen gegeneinander, sie werden je nach regionalem Sprachgebrauch getitscht oder gedotzt, um zu ermitteln, wessen Ei die härteste Schale hat.*¹³ Dort wird *titschen* eher synonym zu *dotzen* gebraucht. Sonst wird *dotzen* hauptsächlich zusammen mit *Ball* gebraucht, in der Bedeutung *aufprallen und hochspringen* (DUW). *Dotzen* und *titschen* kommen sich semantisch nahe in Wendungen wie *jemandem eine titschen* (zweimal in diesem Suchergebnis). *Titschen* kommt oft mit *Ball* vor, und zwar entweder kann man *den Ball titschen*, oder *der Ball titscht (auf)*: *Erst sieht man eine attraktive blonde Frau Volleyball spielen, dann steht sie vor einer Wand und titscht einen Ball immer wieder dagegen.*¹⁴ *Jedoch spielt man mit einem tennisballgroßen Ball, der nicht titscht, in ein dreieckiges Tor und nicht in ein Netz.*¹⁵ Darauf zurück geht wahrscheinlich die Wendung (*vor Ärger, vor Freude*) *im Dreieck titschen* (*sein* im Perfekt, insgesamt 11-mal in allen möglichen Formen vorgekommen): *Da war ich echt platt – ich hätte niemals im Leben damit gerechnet, dass jemand meine Musik rausbringt! Ich bin im Dreieck getitscht! Echt fett!!!*¹⁶ Die anderen untersuchten Formen bargen auch auffallend häufig Beispiele, die der Duden-Beschreibung nicht entsprachen.

¹³ private.freepage.de/cgi-bin/feets/freepage_ext/41030x030A/rewrite/gruenvogel/brauchtum/osterbraeu che.html. [zit. 2003-02-10].

¹⁴ www.hofter.de/Doppelte.htm. [zit. 2003-02-10].

¹⁵ Der Pädant – Lacrosse am Päda: www.paedagogium-godesberg.de/paedant/art0621.htm. [zit. 2003-02-10].

¹⁶ Electronic music info pool: www.ae-pool.de/images/archiv/fangkiebassbeton/fangkiebassbeton_interv1.html. [zit. 2003-02-10].

2.5. trappen

COSMAS: keine Belege

Duden: [aus dem Niederdt. < mnd. trappen, urspr. lautm.]: mit kurzen und hörbaren Schritten gehen.

Google:

- *trappten*: etwa 25-mal,
- *trappte*: 30-mal,
- *getrappt*: 97-mal.

Die Form *trappen* war für die Suchanfrage aufgrund der zahlreichen Eigennamen und anderssprachigen Treffer unbrauchbar. Google zeigte vor allem, dass das Verb für einen einmaligen Tritt (also nicht unbedingt für Gehen) stehen kann: *Trappt man drauf, dann berühren sich die beiden Metallplatten an der Innenseite und schließen den Stromkreis.*¹⁷

Aus den übersehbaren Funden ging hervor, dass *trappen* auch in anderen Bedeutungen in Anlehnung an das englische *to trap* (*in einer Falle fangen*) benutzt wird, und zwar in den Naturwissenschaften: *In einem anderen Modell wird angenommen, dass die durch die Photoionisation der Ge-Ge-Bindung freiwerdenden Elektronen in der Nähe getrappt werden.*¹⁸ In entsprechenden englischen (nicht in deutschsprachigen Ländern entstandenen) Texten kommt *to trap* vor: *High energy electrons have been trapped by a laser beam for the first time.*¹⁹ Es ist auch ein Terminus in der Informatik: *Das Programm trappt alle 32-Bit-Zugriffe auf I/O-Ports oberhalb von 400 Hz.*²⁰ Einmal tauchte *trappen* auch als Verb zu *Trapper* auf (ursprünglich Pelztierjäger in Nordamerika): *Mittlerweile trappt er über 30 Jahre lang und kennt sich draußen in der Natur weit besser aus als in Downtown Fairbanks. Für ihn ist das Trappen nicht mehr wegzudenken.*²¹

¹⁷ Die Funktionsweise des Elektrorads. www.members.aon.at/akis/rafunk.htm. Ähnlich auch: „Trappt der Elefant auf die Maus“. www.blinde-kuh.de/witze/tierwitze.html. [zit. 2003-02-10].

¹⁸ Die Herstellung von Bragg-Gittern in Quarz-Fasern: www.ex.physik.uni-ulm.de/lehre/PhysikalischeElektronik/Phys_Elekt/node180.html. [zit. 2003-02-10].

¹⁹ PhysicsWeb - Electrons trapped by laser beam. physicsweb.org/article/news/3/12/4. [zit. 2003-02-10].

²⁰ Leserforum: www.heise.de/kiosk/archiv/ct/94/09/008/art.htm. [zit. 2003-02-10]. Dementsprechend auf English: „GE-100 instructions sequences were interpreted by the emulator micro-programs at the exception of a few instructions noticeably I/O that were trapped to the emulator support software.“ http://perso.club-internet.fr/febcm/english/section_iv_modes_and_emulators.htm. [zit. 2003-02-10].

²¹ E-mail aus Alaska. <http://www.wdr.de/online/alaska/tagebuch/20020116.phtml> [zit. 2003-02-10].

2.6. werweißen (werweissen)

COSMAS: keine Belege

Duden: [aus „wer weiß ... (ob, wann, wo)“] (schweiz.): *hin und her raten, sich überlegen*: alle haben gewerweißt, was zu tun sei. [...]

Google

- *werweissen*: 119-mal, allerdings meist als Substantiv (was selbst in COSMAS 38-mal belegt ist),
- *werweissten*: 3-mal,
- *werweisste*: 8-mal,
- *gewerweisst*: 8-mal,
- *werweisse*: 2-mal,
- *werweißten*: 5-mal,
- *werweißte*: 1-mal,
- *gewerweißte*: 1-mal,
- *werweiße*: 2-mal.

Die Bedeutung entspricht der im Duden angeführten.

2.7. wuscheln

COSMAS: keine Belege

Duden: [laut- und bewegungsnachahmend, vielleicht beeinflusst von wischen] (landsch.): *mit der Hand durch die vollen Haare fahren*: Sie wuschelt in seinem Haar.

Google: Die Form *wuscheln* kommt 957-mal vor, wobei die Substantivform nicht zu überwiegen scheint.

- *wuschelten*: 20-mal,
- *wuschelt*: 251-mal,
- *wuschele*: 49-mal,
- *wuschle*: 19-mal,
- *gewuschelt*: 57-mal.

Man wuschelt übrigens auch *durch jemandes Haar*, oder einfach transitiv gebraucht, *jemandes Haar/Kopf/Mähne wuscheln, jemandem das Haar/den Kopf/die Mähne wuscheln*.

2.8. wuzeln

COSMAS: keine Belege

Duden: [laut- u. bewegungsnachahmend] (bayr., österr., ugs.): a) *drehen, wickeln*: sich eine Zigarette w.; b) <w. + sich> *sich drängen*: er wuzelt sich durch die Menge.

Google:

- *gewuzelt*: 20-mal,
- *wuzelt*: 34-mal,
- *wuzelte*: 10-mal,
- *wuzelten*: 11-mal,
- *wuzele*: kein Fund brauchbar,
- *wuzle*: 2-mal (sonst unbrauchbar).

Bei *wuzeln* (490-mal) überwogen Substantive, in COSMAS 13 substantivische Belege mit entsprechender Bedeutung. Die beiden vom Duden angesprochenen Bedeutungen waren auch bei Google vertreten. Doch geht aus der Beschreibung im Duden nicht besonders deutlich hervor, was alles gewuzelt werden kann, z.B. *sich im Schnee/in einem Misthaufen usw. wuzeln, etwas in Semmelbrösel/Zucker usw. wuzeln* u.a. *Wuzeln* hat noch eine weitere Bedeutung, die sonst im COSMAS nur aus dem Substantiv hervorgegangen wäre, und zwar (intransitiv) *Tischfußball spielen: Nachher haben wir Darts gespielt und gewuzelt.*²² Manche Seiten (und teilweise auch die Beschreibung im Duden) suggerieren, dass es sich um einen Austriazismus handelt. Die Suchanfrage „*wuzeln + Tischfußball*, Sprache: Deutsch, Land: Österreich“ ergab 63 Seiten. Die gleiche Suchanfrage in der bundesdeutschen Domäne ergab 34 Seiten. Doch ist dies natürlich kein Beweis für die regionale Distribution des Wortes, denn man kann die „süddeutschen“ Seiten nicht von den übrigen bundesdeutschen Seiten isolieren. Außerdem werden die Zahlen von dem hinzugefügten Suchwort *Tischfußball* beeinflusst. Es werden so keine Seiten erfasst, die den Tischfußball betreffen, ohne das erklärende Wort *Tischfußball* zu gebrauchen.

²² Tagebuch Christoph & Lollo: www.christophundlollo.com/tb8.htm [zit. 2003-02-11].

3. Schlussfolgerung

Diese kleine Untersuchung weckt also Fragen wie: Kann man sich auf das umfangreiche Korpus COSMAS verlassen? Soll man lieber gleich auf das Internet zurückgreifen? Die Antwort auf die erste Frage lautet: meistens schon. Die Antwort auf die zweite Frage lautet: nein.

Allerdings gibt es kein ausreichend großes Korpus der gesprochenen deutschen Sprache. Selbst wenn die im COSMAS erfasste Zeitungssprache sich manchmal der gesprochenen nähert, deckt das Korpus die gesprochene Sprache nicht vollständig ab. Hier muss man sich nach wie vor auf unsystematische Stichproben verlassen. Dabei kann das Web behilflich sein. Im Internet verschmilzt die geschriebene Sprache mit der gesprochenen häufiger, als dies in den traditionellen Medien der Fall ist, und es werden Unmengen von interessensspezifischen Texten veröffentlicht, die in allgemein orientierten gedruckten Medien kaum zur Geltung kommen würden.

Bei der Arbeit mit dieser Quelle ist jedoch immer streng zu betrachten, wie viele User am jeweiligen Diskurs beteiligt sind und ob derselbe Sprachstil auch außerhalb des Internets verwendet wird. Webmaterialien sieht man ziemlich gut an, ob der Diskurs im Wesentlichen auf die virtuelle Welt beschränkt bleibt und von ihr geprägt wird (Fantasy- und Computerspiele, Herstellung und Austausch von Musik- und Videodateien, Hackerforen, Webpornographie, Chat-Einrichtungen u.ä.), oder ob die Webseiten das reale gesellschaftliche Leben widerspiegeln (elektronische Bulletins von Schulen, amtlichen Einrichtungen und Kirchen, Sport-, Interessen- und Kulturvereine, Vorträge, Aufsätze, Fachzeitschriften oder elektronische Konferenzen zu aktuellen lebensbezogenen Themen).

Beachtet man diese Regel, dann kann das Internet interessante Belege für den aktuellen Sprachgebrauch liefern, wie auch zwei weitere Beispiele zeigen: *taillieren* und *tariieren*. Die wörterbuchtypischen Kollokatoren für *taillieren* (meist im Partizip II) sind: *Kleid*, *Jacke* usw. Google ergibt etwa 840 Okkurrenzen für *tailliert* und *Kleid*, aber auch 550 Okkurrenzen für *tailliert* und *Ski*. Skifahren ist seit drei Generationen zum Massensport geworden. Skifahren heißt bei den meisten Menschen nicht Identifizierung mit einer relativ geschlossenen Gruppe wie z.B. Hacken, Rappen, Darking oder Graffiti. Es ist sehr wahrscheinlich, dass der Skifahrerwortschatz viele Gesellschaftsschichten erreicht. Sogar die statistische Kollokationsanalyse

in COSMAS teilt dem Wort *Ski* einen LLR-Wert²³ von 1568 zu, das Wort *Jacke* (das Kleidungsstück mit dem höchsten LLR-Wert) erreicht dagegen nur einen Wert von 399. (BELICA o.J.; STEYER 2000, 2002)

*Tarieren*²⁴ wird bei Google fast ausschließlich mit Tauchen verknüpft, wobei die Anzahl der Okkurrenzen ungeheuer ist (960 Treffer für *Tauchen* + *tarieren*), weil Tauchen Mode geworden ist und potentiell zum Massensport werden kann. Im Vergleich zu *taillieren* hat es im COSMAS noch keine Position errungen (bisher keine direkten Belege bei Verben, Wortformen wie *Tarierübungen* und *Tarierweste* sind jedoch reichlich vorhanden). Schließt man alle Wortformen von dem Stamm *-tarier-* mit in die Suchanfrage ein, beziehen sich viele Belege auf das Tauchen.

Viele im Web enthaltene Wörter, die im Korpus scheinbar nicht belegt sind, können durch solch eine erweiterte Rückfrage im Korpus doch noch verifiziert werden. Insgesamt stellt das Korpus den allgemeinen geschriebenen Sprachusus zwar zuverlässiger dar als das Web, weil im Web viel Abweichendes auftauchen kann. Jedoch ist das Web sehr bereichernd, und es wäre schade, es als unseriös zu verwerfen. Am besten ist es sicherlich, sich bei selten gebrauchtem Wortschatz vom Web inspirieren zu lassen, jedoch die Belege aus dem Web im Korpus zu verifizieren, um sicher zu gehen, dass der betreffende Ausdruck allgemein bekannt ist. Deshalb ist die Anwendung des Webs zumindest als ergänzende Quelle ausgesprochen empfehlenswert.

Literatur

Arbeitsgruppe für Korpustechnologie (2002): Die COSMAS-Kollokationsanalyse. <http://www.ids-mannheim.de/kt/kollok.html>, letzte Änderung: Okt. 2002.

Belica, Cyril (1995-2002): *Statistische Kollokationsanalyse und Clustering*. <http://corpora.ids-mannheim.de/cosmas>. Institut für Deutsche Sprache, Mannheim.

²³ LLR (Log-likelihood ratio) ist eine mathematisch-statistische Methode, mit deren Hilfe signifikante Kollokatoren von beliebigen Trefferobjekten in einem Korpus gefunden werden können.

²⁴ 1. die Tara (einer verpackten Ware) bestimmen. 2. (Physik) das Gewicht auf einer Waage durch Gegengewichte ausgleichen. (DUDEN 1999)

- COSMAS (o.J.): *Textkorpus COSMAS. IDS Mannheim*. <http://corpora.ids-mannheim.de/cosmas/>
- Duden (1999): *Das große Wörterbuch der deutschen Sprache in 10 Bänden*. 3. Aufl., Mannheim.
- Duden (2000): *Das große Fremdwörterbuch. Herkunft und Bedeutung der Fremdwörter*. 2. Aufl., Mannheim.
- Duden (2001): *Deutsches Universalwörterbuch*. 4. Aufl., Mannheim.
- Sinclair, John (1995³, 1991¹): *Corpus, Concordance, Collocation*. Oxford.
- Steyer, Kathrin (2000): „Usuelle Wortverbindungen des Deutschen. Linguistisches Konzept und lexikografische Möglichkeiten.“ In: *Deutsche Sprache 2*: 101-125.
- Steyer, Kathrin (2002): „Wenn der Schwanz mit dem Hund wedelt. Zum linguistischen Erklärungspotenzial der korpusbasierten Kookkurrenzanalyse.“ In: Haß-Zumkehr, Ulrike; Kallmeyer, Werner (Hrsg.): *Ansichten zur deutschen Sprache. Festschrift für Gerhard Stickel zum 65. Geburtstag* (= Studien zur deutschen Sprache). Tübingen: 215-236.
- Vachková, Marie (2002): „Das große deutsch-tschechische Wörterbuchprojekt im Kontext der zeitgenössischen deutsch-tschechischen Übersetzungslexikographie.“ In: Kunzmann-Müller, Barbara; Zielinski, Monika (Hrsg.): *Sprachwandel und Lexikographie. Beispiele aus slavischen Sprachen, dem Ungarischen und Albanischen* (= Berliner slawistische Arbeiten). Frankfurt /M. [im Druck].