

Elektronische Sprachkorpora in Wissenschaft und Unterricht DaF/DaZ – einige Vorschläge für die Nutzung der elektronischen Instrumente¹

Neue Technologien gewinnen immer mehr an Bedeutung – sogar in solchen Bereichen, die primär keine aufwändige Technik benötigen. Sie setzen sich selbst dort durch, wo man sie vor einigen Jahren noch nicht erwartet hätte. Ein solcher Bereich ist wohl auch die Fremdsprachenvermittlung.

In diesem Beitrag möchte ich auf elektronische Sprachkorpora fokussieren, einen Ausschnitt aus der Arbeit mit ihnen präsentieren und so ein Spektrum an Ideen für die Nutzung der Korpora im Bereich DaF/DaZ mit Schwerpunkt auf die Wortschatzarbeit eröffnen. Vor dem Ausflug zu den Korpora und in ihre Daten sollte zuerst kurz erwähnt werden, welche Bedeutung die Korpuslinguistik in der heutigen Zeit hat und welche Korpora es gibt, die für DaF/DaZ von Bedeutung sind.²

Die Nutzung der elektronischen Sprachkorpora ist in der Fremdsprachenforschung und im Unterricht ein relativ neues Element. Bis heute bedient man sich der traditionellen empirischen Instrumente (vgl. z.B. Muzikant 2007), die auch in Zukunft nicht an Bedeutung verlieren werden. In den letzten Jahren aber ist in vielen Sprachforschungsbereichen (vgl. Peloušková 2005) und auch im Sprachunterricht eine Orientierung auf die effektiven Korpusmethoden deutlich sichtbar (Kášová 2006). So basieren beispielsweise die in den letzten Jahren erschienenen Grammatiken und Wörterbücher weitestgehend auf Recherchen in elektronischen Korpora (Zifonun et al. 1997; Wernke et al. 2005; Kloosa et al. 2001 u.v.a.m.).

1. Korpuslinguistik

Die Korpuslinguistik ist jener Teil der Linguistik, der die Sprache systematisch anhand großer Mengen von realisierten sprachlichen Äußerungen (*Parole*) untersucht. Das Ziel des korpuslinguistischen Ansatzes in der modernen Linguistik ist es, die Strukturen und Funktionen der natürlichen Sprache besser und realitätsnäher zu beschreiben (vgl. z.B. Biber et al. 1998). Damit ist die Korpuslinguistik keine neue Theorie der linguistischen Untersuchung, sondern eher eine effektive Methode der Sprachforschung.³

aus: Theorie und Praxis
österreichische Beiträge zu DeuBl
als Fremdsprache 14/2007,
Schwerpunkt: Wortschatz,
S. 123 – 136.

Zu Nutznießern der Korpora zählen PhilologInnen im weiteren Sinne (auch LiteraturwissenschaftlerInnen können elektronische Korpora gut nutzen – z.B. für eine Recherche über typische Kollokationen eines/einer Autors/Autorin), ÜbersetzerInnen und Dolmetscher und DolmetscherInnen, SprachlehrerInnen, Schüler und SchülerInnen, Studenten und Studentinnen und eigentlich alle, die Interesse an dem Phänomen Sprache haben, denn die meisten elektronischen Korpora sind leicht zugänglich und auch für nicht Computergewandte einfach zu nutzen.

1.1. Korpusrecherche vs. traditionelle Datenerhebung

Die Vorteile der Arbeit mit elektronischen Korpora sind spätestens dann klar, wenn jemand sich eine (linguistische oder linguistisch bezogene) Frage stellt und die Antwort mit konkreten Beispielen belegen möchte. „Sagen Sie *das E-Mail* oder *die E-Mail*? Und wie schreiben Sie *das E-Mail* oder *E-mail*?“ Statt einer Umfrage bei MuttersprachlerInnen – was sich für Deutschlernende im Ausland besonders schwierig darstellt – oder statt einer langwierigen Recherche in diversen gedruckten Texten, bietet bereits eines der Korpora eine überwältigende Anzahl von Belegen: 18.090-mal *E-Mail* und 5.374-mal *E-mail* hat der Korpusmanager in über 4 Millionen unterschiedlichen Texten (über eine Billion Wörter)⁴ binnen 15 Sekunden im Mannheimer Korpus gefunden. Und mit diesen Daten kann man weiter analysieren – z.B. in welchen Texten *das E-Mail* oder *die E-Mail* vorkommt. Wie lange würde eine manuelle Recherche für dasselbe Ergebnis dauern?

2. Korpustypen

Für einen erfolgreichen Datengewinn ist zunächst die **Korpuswahl** sehr wichtig. Die Überlegung, welche Recherche ich unternehmen will, mündet logischerweise in die Wahl des geeigneten Instruments. Erwähnt werden hier nur für DaF/DaZ relevante Korpora.

2.1. Diachrone und synchrone Korpora

Für die historischen Untersuchungen der Sprache gibt es bereits einige Korpora – z.B. die *Mittelhochdeutsche Begriffswortbank (MHDBB)* an der Salzburger Universität oder *Das digitale Wörterbuch der deutschen Sprache des 20. Jahrhunderts (DWDS)*, das zwar ausschließlich das Neuhochdeutsche erfasst, sich jedoch wunderbar als Instrument zur Erfassung der (Ver-)Änderungen des Deutschen im vorigen Jahrhundert eignet (Beispiele siehe Kap. 4.1.2.).

Zur Betrachtung der heutigen deutschen Sprache sind schon einige Korpora für jeden per Internet zugänglich: z.B. *Das Deutsche Referenzkorpus (DeReKo)* am IDS-Mannheim, das *Korpus Wortschatz* erstellt von der Informatik der Universität Leipzig oder das schon erwähnte *DWDS*, um nur einige zu nennen.

2.2. Korpora der gesprochenen und geschriebenen Sprache

Alle bisher erwähnten Korpora erfassen nur die *geschriebene Sprache*. Es gibt allerdings einige Korpora der *gesprochenen Sprache*, z.B. das *Deutsche Spracharchiv (DSAr)* und die *Datenbank Gesprochenes Deutsch (DGD)* – beide am IDS-Mannheim.

2.3. Einsprachige und parallele Korpora

Parallele (mehrsprachige) Korpora sind viel seltener als *einsprachige*, weil ihr Aufbau wesentlich komplizierter ist, da die entsprechenden Passagen einander zugeordnet (aligned) werden müssen. (Káha/Peloušková 2006b)

2.4. Relevante Korpora für DaF/DaZ

Es gibt also mehrere Korpora des geschriebenen und gesprochenen Deutsch, wo die Sprache in ihrer natürlichen Form, also in authentischen Texten, erscheint. Zur Demonstration der Arbeit mit diesen Instrumenten in diesem Beitrag wählte ich die folgenden Korpora:

- **Das Deutsche Referenzkorpus (DeReKo)** am IDS-Mannheim – „die weltweit größte Sammlung elektronischer Korpora mit geschriebenen deutschsprachigen Texten aus der Gegenwart und der neueren Vergangenheit“⁵
- **Das digitale Wörterbuch der deutschen Sprache des 20. Jahrhunderts (DWDS)**⁶ an der Berlin-Brandenburgischen Akademie der Wissenschaften
- **Wortschatz** am Institut für Informatik der Universität Leipzig⁷
- **Das tschechisch-deutsche Parallelkorpus (CNPK)** an der Masaryk-Universität Brünn – das größte Parallelkorpus für Tschechisch und Deutsch⁸ – erwähne ich hier als einen Vertreter der Parallelkorpora.

Technische Angaben, sowie die Syntax der Korpusabfrage sind den jeweiligen Dokumentationen (Internetseiten) der Korpora zu entnehmen.

3. Möglichkeiten der Nutzung von Korpora

Elektronische Korpora kann man für die Untersuchung aller Ebenen der Sprache nutzen, wenn auch mit unterschiedlichem Aufwand. Text- und Satzsemanantik können aus den bisher selten und nur schwer zugänglichen „geparsten“ (syntaktisch annotierten) Korpora schöpfen. Für andere Bereiche eignen sich hingegen viele relativ gut zugängliche und einfach zu bedienende Korpora. Im Phonetikunterricht können Korpora der gesprochenen Sprache von großer Hilfe sein, weil ihre Daten oft auch akustisch abrufbar sind (DSAV und DGD aber auch andere, wie z.B. Bayerisches Archiv für Sprachsignale). Grammatik, Orthographie und der Wortschatz lassen sich am besten an Korpora der geschriebenen Sprache erlernen. Diese sind nämlich viel größer als die Korpora der gesprochenen Sprache.

3.1. Grenzen der Möglichkeiten

Die Vorteile der Korpusarbeit sind unumstritten. Es ist allerdings auch darauf hinzuweisen, dass die Korpora in einigen Bereichen keine Hilfe sein können. Zu diesen zählen unter anderem:

- kontrastive Forschung der gesprochenen Sprache
- Untersuchung der im Korpus nicht vorhandenen Texte (z.B. Aufschriften)
- literaturwissenschaftliche Arbeiten im engeren Sinne (Interpretationen, Komparatistik)
- Aufstellung absoluter Urteile über die Sprache

Darüber hinaus ist festzuhalten, dass jedes Korpus durch seine Größe bzw. Ausgewogenheit der Texte begrenzt ist. Nichtsdestotrotz kann der Beitrag der elektronischen Korpora für die Spracharbeit äußerst ergiebig sein. Das möchte ich am Beispiel der Wortschatzarbeit demonstrieren.

4. Wortschatzarbeit mit den Korpora

Die Wortschatzarbeit mit dem Korpus kennt fast keine Grenzen, da auch der Wortschatz einer Sprache nahezu unbegrenzt ist. Aufgrund der einfachsten Fragen im Korpusmanager (Wortform, Wortkombination, Lemma und Tag = morphologische Kategorie⁹) lassen sich die wichtigsten Informationen über das abgefragte Phänomen gewinnen. Im Folgenden möchte ich zeigen, wie zwei Konkurrenzformen in verschiedenen Korpora vertreten sind (4.1.), also wie die Wörter im Wortschatz des 20. Jahrhunderts eingebettet

sind. Anhand einiger Übungen möchte ich vorführen, welche Möglichkeiten die elektronischen Korpora für die Vermittlung des Deutschen im DaF/DaZ-Unterricht bieten (4.2.). Beide Kapitel setzen sich zum Ziel, das Spektrum der in den Korpora enthaltenen Informationen vorzustellen, die in der vorelektronischen Zeit kaum zu gewinnen waren.

4.1. Fallbeispiel der Korpusarbeit: Tschechien und/oder Tschechei

Die Tschechen, vor allem die Generationen, die den größeren Teil ihres Lebens noch in der Tschechoslowakei erlebt haben, weigern sich oft sich mit der „neuen“ Bezeichnung ihres Landes („Česko“) zu identifizieren. Diese klingt für tschechische Ohren etwa wie „Tschecho“ im Deutschen. Paradoxerweise wissen aber viele, wie das Land auf Deutsch heißen soll: nämlich *Tschechien*.¹⁰ Ja auf keinen Fall die *Tschechei*! Aus historischen Gründen, selbstverständlich.

Fragestellung: Wie, wann und wo werden diese zwei Benennungen in deutschen Texten verwendet? Und wird auf Tschechisch differenziert, wenn im entsprechenden deutschen Text Tschechien oder Tschechei steht?

4.1.1. Synchrone Distribution

Den ersten Eindruck über die Häufigkeit der Verwendung liefert das IDS-Korpus DeReKo: Hier überwiegt eindeutig die Form *Tschechien* (38.500-mal) vor *Tschechei* (nicht einmal 1.000-mal). Die Anzeige der Treffer in den einzelnen Dokumenten ergibt, dass das Wort *Tschechei* nur in öffentlich zugänglichen Zeitungstexten und fast ausschließlich in österreichischen, seltener in schweizerischen, jedoch kaum in bundesdeutschen Zeitungen, vorkommt. Die Textortendistribution ist im IDS-Korpus schwer zu erschließen, denn der öffentliche Teil des Korpus besteht fast ausschließlich aus Zeitungs- und Zeitschriftentexten.

Die zeitliche Distribution von beiden Wörtern aus der synchronen Sicht ist ungefähr gleich. Beide Wörter kommen am häufigsten zwischen den Jahren 1996 und 1998 vor.

4.1.2. Diachrone Distribution (20. Jahrhundert)

Eine wenig überraschende graphische Statistik über den Gebrauch des Wortes *Tschechei* im ganzen 20. Jahrhundert liefert das DWDS. Das Vorkommen der Allonyme in den Korpus-texten (in absoluten Zahlen) stellt das Instrument nach Dekaden automatisch dar:

4.1.3. Tschechische Entsprechungen

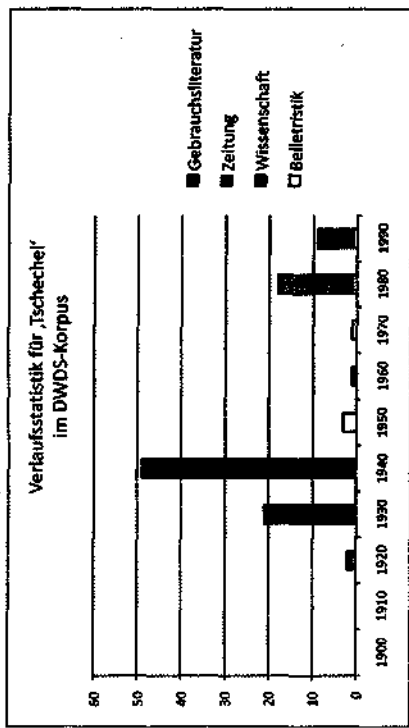
Da wir es hier mit einer lexikalischen Divergenz zu tun haben (Deutsch: *Tschechien* und *Tschechei* – zwei Simplexwörter; Tschechisch *Česko* – ein Simplexwort), bietet sich die Frage an, wie die deutsche Dichotomie im Tschechischen gelöst wird. Die Antwort kann das Tschechisch-deutsche Parallelkorpus andeuten:

- (1a) Meine Mutter war ja da noch in der Resttschechoslowakei, *Tschechei*, denn die Slowakei hat sich ja auch selbständig gemacht.
- (1b) Maminka potom byla ještě ve „zbytku Československa“, tedy v *Česku*, protože Slovensko vyhlásilo samostatnost.
- (2a) Oder sie haben sich von daheim von irgendwas die Etiketten rausgetrennt und haben sie in einen neuen Mantel reingenäht, weil in der *Tschechei* alles billiger war ...
- (2b) Nebo doma z něčeho odpárali štítky a našili je na nový kabát, protože v *Česku* bylo všechno lacinější ...
- (3a) Sind wir wirklich bereit, dem Ski-Tourismus einen der landschaftlich schönsten Winkel in der *Tschechei* zu opfern?
- (3b) Jsme připraveni neomezenému využívání obětovat ochranu jednoho z nejvzácnějších koutů české přírody?

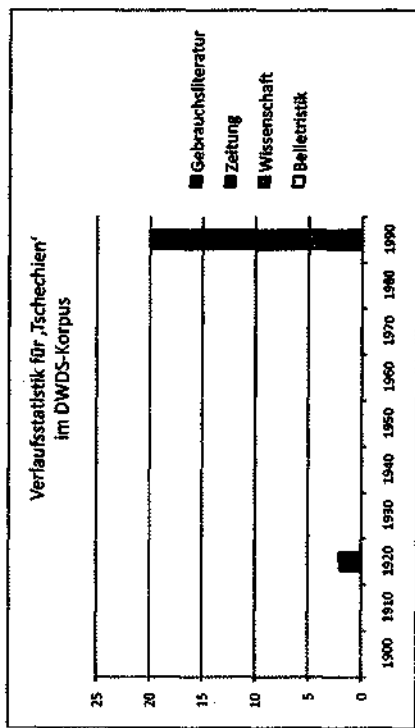
Der Form *Tschechei* (im Korpus allerdings nur 22-mal) entspricht im ČNPK in den meisten Fällen das „neue“ Wort *Česko*, egal ob es sich um das Land zwischen 1938 und 1945 (1), um eine umgangssprachliche Kürzung der Tschechoslowakischen sozialistischen Republik (2) oder um die Tschechische Republik nach 1993 (3) handelt. (Das Adjektiv *český* (3b) ist allerdings vom Substantiv *Čechy* [Böhmen] abgeleitet, kann jedoch auch als eine Derivation von *Česko* betrachtet werden.)

Der Eigennamen *Tschechien* (mit 100 Treffern) kommt in den Paralleltextrn häufig als *Česká republika* (der offizielle Name des Landes) (4) oder *ČR* (5), oft auch als eine politisch korrekte, inoffizielle Bezeichnung *české země* („Tschechische Länder“, also Böhmen und Mähren) (6), vor.

- (4a) Gerade in der Gegenwart rücken die osteuropäischen Staaten wie Polen, Ungarn, *Tschechien*, die Slowakei, Slowenien, Kroatien und Bosnien in das Gesichtsfeld Mitteleuropas.
- (4b) Právě dnes se státy jako Polsko, Maďarsko, *Česká republika*, Slovensko, Slovinsko, Chorvatsko či Bosna stávají středem zájmu.
- (5a) In *Tschechien* ist es ein einzigartiges Beispiel eines Herrensitzes, der durch französische Architektur aus der Zeit des Königs Louis XVI. inspiriert war.



Grafik 1: Tschechien: Das Wort *Tschechien* kommt laut Korpusdaten am häufigsten in den 40er und 50er Jahren des 20. Jh. in „Gebrauchstexten“ (hellgrau) vor.



Grafik 2: Tschechien: Die Form *Tschechien* ist erst im dem (meist) publizistischen (dunkelgrau) Texten aus den 90er Jahren vertreten. ¹⁾

(5b) V ČR ojedinělý příklad panského sídla inspirovaného francouzskou architekturou z doby Ludvíka XVI.

(6a) Die Teilung der Tschechoslowakischen Föderativen Republik zum 1.1.1993 brachte außer dem Ende des Bestehens der ČSD auch die Entstehung der eigenständigen Tschechischen Bahnen (ČD) und der Eisenbahn der Slowakischen Republik (ŽSR), als auch die weitere Schwächung des Eisenbahnverkehrs zwischen Tschechien und der Slowakei.

(6b) Rozdělení Československé federativní republiky k 1.1.1993 přineslo kromě zániku ČSD a vzniku samostatných Českých drah (ČD) a Železnice Slovenské republiky (ŽSR) také další oslabení železniční přepravy mezi českými zeměmi a Slovenskem.

4.1.4. Typische Kollokationen

Alle Korpora können auch die typischen oder häufigsten Kookkurrenzen (Kollokationen) zu den gesuchten Wörtern aufzeichnen. Für *Tschechien* und *Tschechei* sind es die Präpositionen *nach* bzw. *in*, aber auch *gegen* (das Land wird oft in den Sportnachrichten erwähnt). Als autosemantische Kollokatoren kommen (fast ausschließlich) andere Länder vor – zum Beispiel in den Nachrichtentexten mit Auflistungen der EU-Bertrittsländer oder eben auch als Gegner Tschechiens in verschiedenen Sportarten.

4.1.5. Fazit

Die Recherchen in drei Korpora deuten darauf hin, dass die Benennung *Tschechien* sich im Deutschen in den letzten Jahren des 20. Jahrhunderts durchgesetzt hat, die allonyme Form *Tschechei* ist aber damit lange noch nicht ausgestorben. *Tschechei* bleibt auch in einigen belletristischen Texten lebendig (z.B. im ČNPK handelt es sich um eine Übersetzung von Jáchym Topol's „Sestra/Die Schwester“).

Der Form *Tschechien* entspricht im Tschechischen in etwa die offizielle Bezeichnung *Česká Republika* oder die Umschreibung *české země*. Dem oft negativ konnotierten Wort *Tschechei* steht im Tschechischen auch die Form *Česko* gegenüber, die übrigens eher negativ (umgangssprachlich, salopp) bewertet wird.

Die Ergebnisse zeigen, dass die bestehenden Korpora interessantes Material zu einer tiefer gehenden Untersuchung liefern können. Dabei verliert man keine Zeit mit der Datenerhebung, denn die Abfrage im ČNPK¹² dauert wenige Sekunden. Eine vergleichbare manuelle Recherche in parallelen Texten würde mehrere Jahre dauern.

4.2. Wortschatzübungen mit den Korpora

Die elektronischen Korpora stellen auch ein blitzschnelles und relativ zuverlässiges Hilfsmittel für Lehrer und Lehrerinnen für die Erstellung von Wortschatzübungen dar (4.2.1.). Das bedeutet das Aus für künstlich ausgedachte Übungssätze, die erstens die AutorInnen viel Energie kosten und zweitens nicht immer ganz realitätsnah klingen. Mit einem Internet- und Korpuszugang können aber auch die Lernenden selbst vieles über die Sprache erfahren (4.2.2.).

4.2.1. Übungsbeispiele

Mit Hilfe der einsprachigen Korpora lässt sich ein breites Spektrum an Ergänzungsübungen erstellen. Diese können z.B. so aussehen:

Ergänzen Sie *ziehen* oder *schleppen* in der richtigen Form!¹³

Nein, so was wird man nicht los, das	ich weiter mit mir rum, solange ich lebe.
Der Versuchsleiter	die Karten der Reihenfolge nach und ...
Er packte mich am Ärmel und	mich durch den Gang zwischen den Sitzen.
Die Baronin	inzwischen die Geschenke aus dem Korb.

Was fehlt hier?¹⁴

Wetterdach, hing ein alter	an diesem Wegkreuz, das zu untergebracht, die
Material ist in einem mobilen	ist, so Mauro Ferrati, dass
Hängartner. Das Delikate am	des großen Kreuzes aus der
werden. Eben bekommt auch der	aufbewahrt werden können.
neu nummeriert, damit diese im	von 40 Werken.
Von Tobey hat der Sammler ein	Christi stammt aus Rom.
führt – mitverarbeitet. Der	

... wie sich einige Länder hie und da irgendein vergessenes historisches aufwärmten.

Sei kein, lass Dir nichts gefallen.

Ich werde nicht die beleidigte spielen, mich auch auf diesem Posten voll einsetzen. Aber ich will meinen alten Job zurückhaben.

Ich will dir keinen ums Maul schmierem. Aber ich habe mich gewundert, wie du gespielt hast. Hut ab vor deiner Leistung!

Ergänzen Sie die Pluralformen.¹⁶

RIO DE JANEIRO. Verheerende Regen..... haben in den brasilianischen Bundesstaaten Alagoas und Pernambuco bisher 56 Menschen das Leben gekostet.

Schließlich soll für jeden Geschmack und für alle Alters..... etwas dabei sein. Mit der Konzeption hatte die Stadt die Sound of Frankfurt Veranstaltungs-GmbH beauftragt.

Die Umfrage des Verbandes zeigte: das Modell hat Mängel. Der Vorstand errechnet, dass es bei einer Realisierung unter den Gemeinden „mehr Pechvögel als Glücks.....“ gäbe. Dem Modell wird angelastet: Es ist unausgereift.

Echte Äpfel türmen sich zu bunten Bergen, Kohl..... in allen Variationen liegen in den Steigen. Kartoffeln mit echten Erdresten sind in Säcke abgefüllt.

Parallele Korpora ermöglichen den kontrastiven Blick auf die Strukturen der Sprachen durch simple Übungen:

Setzen Sie passende Diminutivformen ein, falls dies möglich ist.¹⁷

Wir sind doch schon fertig, Danny. Oder war das (Albert)?

My už jsme hotoví, Danny. Nebo to byl Bertík?

Hier, mein (Tochter), fühle ich mich am besten. Sogar sterben möchte ich hier.

Tady já, cérečko, mám sa najlepšie. Aj umíft bych tu chřel.

Ehe ich aber darauf zu sprechen komme, möchte ich die verehrte Frau Bednarek bitten, mir ein (Glas) mit Gespritztem (öster.) zu bringen, weißem natürlich.

Ale než se do toho pustím, požádal bych ctěnou Frau Bednarek, aby mi přinesla sklenku vinného střiku, bílého samozřejmě.

Geruheten Sie gut zu (schlafen)?

Ráčil jste spinkat, že?

4.2.2. Aufgaben zum Selbststudium

Einige Korpora ähneln bereits einem Thesaurus: Einen schnellen Überblick über die Beziehungen eines Wortes verschafft das aus „öffentlich zugänglichen Internettexten automatisch aufgebaute Korpus“¹⁸ Wortschatz an der Leipziger Universität. Eine einfache Abfrage zeigt dem Benutzer neben den grammatikalischen Angaben (alle Konjugationsformen, Partizipbildung etc.) auch paradigmatische Beziehungen des Wortes (Wortfeld, Synonyme, Antonyme) sowie auch syntagmatische Beziehungen – rechte und linke Nachbarn, typische Kollokatoren. Beispiele in Form einiger Konkordanzzeilen fehlen auch nicht. So stellen die elektronischen Korpora auch ein einzigartiges Instrument fürs autonome Lernen dar.¹⁹

- Was sind die typischen Kollokatoren vom Verb *ziehen*?
- Wann hat der *Hit* den *Schlager* abgelöst?
- Sind die Wörter *Tropp*, *Effeff*, *Pack* und *Kegel* Monokollokate?
- Was entspricht im Tschechischen (Englischen, Slowakischen, Französischen ...) dem österreichischen Wort *Buschenschank*?

So könnten die Aufgaben zum Selbststudium lauten. In diese Kategorie gehört auch die unter 4.1. angeführte Recherche.

5. Zusammenfassung und Ausblick

Mit der Präsentation der Teilergebnisse einer linguistischen Untersuchung (Kap 4.1.) und mit einigen Beispielen von Übungen und Aufgaben, die auf Korpusdaten basieren, wollte ich in diesem Beitrag nicht nur die Breite des Spektrums der Korpusarbeit andeuten, sondern in erster Linie auf die Effektivität und Geschwindigkeit der Datenerhebung hinweisen.

Im Bereich der Sprachforschung wurden bereits einige Arbeiten von unterschiedlichem Umfang (von Seminararbeiten über Dissertationen, bis zu Wörter- und Grammatikbüchern) anhand der Korpusinstrumente verfasst. Viele Fragen der kontrastiven Sprachuntersuchung bleiben aber vorerst unbeantwortet, weil die entsprechenden Instrumente fehlen. Es gibt schon mehrere Parallelkorpora (MULTEXT, Europart u.a.), alle sind jedoch entweder zu klein oder zu spezifisch für eine allgemeine kontrastive Sprachuntersuchung (und dadurch auch für die allgemeine Fremdsprachenvermittlung). Als viel versprechend können die Arbeiten am Multikorpus *Intercorp*²⁰ betrachtet werden, einem Korpus, das seit 2001 in Prag entsteht, und fast alle europäischen und auch einige asiatische Sprachen beinhalten wird.

Wir haben nun einen kleinen Einblick in die vielfältigen Anwendungsmöglichkeiten von Korpora erhalten. Angesichts der Tatsache, dass Korpora eine ausgezeichnete Basis darstellen, um die Sprache realitätsgetreu und in ihrer Vielfalt zu vermitteln, erstaunt es, dass die LehrbuchautorInnen in ihren interaktiven Zusatzmaterialien bisher kaum auf allgemein zugängliche Korpora verweisen oder gar mit ihnen arbeiten. Ebenso gefordert wäre natürlich die Lehreraus- und -fortbildung, dieses Instrument den Lehrenden für den (Fremd-)Sprachenunterricht vertraut zu machen.

Anmerkungen

- 1 Dieser Artikel entstand im Rahmen des Forschungsvorhabens MSM 0021620823.
- 2 Da ich meinen Beitrag als Anregung für die Arbeit mit verschiedenen Korpora und Korpusinstrumenten sehe, wird hier auf Definitionen der Grundbegriffe und der Basisfunktionen, die jedes Korpus in der Dokumentation ausführlich beschreibt, verzichtet. Siehe dazu auch: Biber et al. 1998; Čermák/Blatná 2006; Lemnitzer/Zinsmeister 2006; Káňa 2006.
- 3 Korpora hat es ja auch früher gegeben, ihre Aufstellung und die Recherchen waren allerdings unständig.
- 4 IDS-Korpus – öffentliche Version.
- 5 <http://www.ids-mannheim.de/ld/projekte/korpora/>.
- 6 <http://www.dwds.de/>.
- 7 <http://wortschatz.uni-leipzig.de/>.
- 8 Aus urheberrechtlichen Gründen nur intern zugänglich. Eine Liste der parallelen Korpusanteile ist unter: <http://www.ped.muni.cz/wger/>, Icon „rtůzně“, „katalog čnpk“ abrufbar.
- 9 Káňa 2006.
- 10 Zu diesem Thema gab es z.T. heftige Diskussionen in den tschechischen, aber auch in deutschen und österreichischen Zeitungen nach 1993 (Jahr der Trennung der Tschechoslowakei). Siehe dazu auch das IDS-Korpus.
- 11 Tschechien ist jedoch keine Neuprägung der 90er Jahre. Es ist auch in zwei Dokumenten aus den 20er Jahren belegt.
- 12 Das ČNPK beinhaltet über 6 Millionen tschechische Wörter und über 6,5 Millionen Wörter in der deutschen Parallele in 200 unterschiedlichen Texten.
- 13 Quelle: das ČNPK.
- 14 Quelle: IDS-Korpus.
- 15 Quelle: IDS-Korpus.
- 16 Quelle: IDS-Korpus.
- 17 Quelle: Káňa 2005, 86–87.
- 18 <http://corpora.informatik.uni-leipzig.de/download.html> unter „Disclaimer“.
- 19 Káňová 2006.
- 20 Ein Teil dieses Multikorpus (<https://trnka.fi.cuni.cz/ucnk/intercorp/>) bildet auch das ČNPK, ein Korpus, das vielen tschechischen DaF-Studentinnen und DaF-LehrerInnen seit einigen Jahren zur Verfügung steht.

Literatur

- Biber, Douglas/Conrad, Susan/Reppen, Randi (1998), *Corpus Linguistics. Investigating Language Structure and Use*, Cambridge: Cambridge University Press.
- Čermák, František (2006), „Korpusová lingvistika dnešní doby“, in: František Čermák/Renata Blatná (Hrsg.), *Studie z korpusové lingvistiky*, Praha, 9–18.
- Káňa, Tomáš (2005), *Wortbildung. Umriss der Theorie mit Übungen*, 1. Aufl., Brno.
- Káňa, Tomáš (2006), „Korpuslinguistik – eine übersehene Herausforderung für den Deutschunterricht“, in: Paul R. Portmann-Tselikas/Hans-Jürgen Krumm (Hrsg.), *Theorie und Praxis. Österreichische Beiträge zu Deutsch als Fremdsprache Band 9/2005. Schwerpunkt: Innovationen – Neue Wege im Deutschunterricht*, Innsbruck/Wien/Bozen: Studienverlag, 99–115.
- Káňa, Tomáš/Peloušková, Hana (2006a), „Elektronische Korpora in Tschechien und das tschechisch-deutsche Parallelkorpus“, in: Bernhard Kettemann/Georg Marko (Hrsg.), *Planning. Gluing and Painting Corpora. Inside the Applied Corpus Linguist's Workshop*, 1. Aufl., Frankfurt am Main: Lang, 27–46 (= Sprache im Kontext 24).
- Káňa, Tomáš/Peloušková, Hana (2006b), „Was kann das parallele Korpus? Funktionen, Möglichkeiten und Grenzen“, in: FF UCM & Gesellschaft für Sprache und Sprachen (Hrsg.), *Sprache und Sprachen im mitteleuropäischen Raum*, Trnava, 603–615.
- Káňová, Martina (2006), *E-Learning – využití internetových projektů při přípravě učitelův německého jazyka*, Prešov.
- Kloss, Annette et al., Hrsg. (2001), *DUDEN – Deutsches Universalwörterbuch*, 4. Aufl., Mannheim/Wien u.a.
- Lemnitzer, Lothar/Zinsmeister, Heike (2006), *Korpuslinguistik. Eine Einführung*, Tübingen: Narr.
- Muzikant, Mojmir (2007), *Deutsche Dialekte in Tschechien, ihre Gemeinsamkeiten und Besonderheiten im Vergleich zum Mutterboden*, Brno.
- Peloušková, Hana (2005), „Informační technologie jako podpora studia učitelství německého jazyka“, in: Věra Janíková/Tomáš Káňa (Hrsg.), *Inovativní tendence v kurikulu studijního oboru „Učitelství německého jazyka pro základní školy“*, Brno.
- Wernke, Mathias et al., Hrsg. (2005), *DUDEN – Die Grammatik*, 7. Aufl., Mannheim.
- Zifonun, Gisela/Hoffmann, Ludger/Strecke, Bruno u.a. (1997), *Grammatik der deutschen Sprache*, Berlin.

Korpora

Bayerisches Archiv für Sprachsignale:

<http://www.phonetik.uni-muenchen.de/Bas/BasHomedeu.html>

Das deutsche Referenzkorpus (DeReKo):

<http://www.ids-mannheim.de/kl/projekte/dereko/>

Das digitale Wörterbuch der deutschen Sprache des 20. Jahrhunderts (DWDS):

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

Deutsches Spracharchiv (DSAv) und Datenbank Gesprochenes Deutsch (DGD):

<http://dsav-oeff.ids-mannheim.de/>

InterCorp:

<https://trnka.fi.cuni.cz/ucnk/intercorp/>

Mittelhochdeutsche Begriffsdatenbank (MHDBDB):

<http://mhdbdb.sbg.ac.at:8000/>

Tschechisch-deutsches Parallelkorpus/Cesko-nemecký paralelní korpus (ČNPK):

momentan kein öffentlicher Zugang

Wortschatz – Universität Leipzig:

<http://wortschatz.uni-leipzig.de/>