

# **Zum aktuellen Thema: Die Korpuslinguistik im deutsch-tschechischen Sprachvergleich**

*Iva KRATOCHVÍLOVÁ*

## **1. Textkorpora im Forschungsbereich der kontrastiven Linguistik**

Korpora im heutigen Sinne werden als strukturierte repräsentative Sammlungen von Texten verstanden, die in maschinenlesbarer Form aufbereitet werden und qualitativ eine umfangreiche Datenbasis für unterschiedliche Arten linguistischer Untersuchungen darstellen. Diese Korpora im Allgemeinen vermitteln eine objektive Untersuchungsmethode bei großen Textmengen und können auch im Endeffekt unsere Sprachreflexion beeinflussen.

Die kontrastive Sprachwissenschaft spezialisiert sich auf den Vergleich von unterschiedlichen Sprachsystemen. Aus dieser Perspektive gesehen stellen die bi- bzw. multilingualen Korpora in allen ihren Erscheinungsformen eine effiziente Vergleichsbasis für die Konfrontation von zwei oder mehreren Sprachsystemen dar. Die konkreten Ergebnisse werden von der mehrsprachigen Lexikographie oder in der Übersetzungswissenschaft angewandt, da die korpuslinguistische Sprachbetrachtung den authentischen Sprachgebrauch beobachtet. Zur Feststellung der anderssprachigen Äquivalenz, sowohl grammatikalischen als auch lexikalischen, ist es notwendig, das jeweilige Sprachelement im aktuellen Kontext, d. h. im Korpus, zu betrachten. Denn wie Teubert (1999, S. 302) zutreffend bemerkt, „dass die Bedeutung der Wörter Textbelegen zu entnehmen ist, war seit Adelung und Campe das Grundprinzip der klassischen Lexikographie, das von der Korpuslinguistik übernommen wird.“ Im zwischen-sprachlichen Kontrast muss zugleich ein wichtiger Aspekt berücksichtigt werden: Das Wort in der Ausgangssprache muss nicht unbedingt einer äquivalenten Struktur der Zielsprache entsprechen. Das Einzelwort ist demnach nicht die einzige Sinneinheit. Es ist daher die Aufgabe der bilingualen Korpora, nach äquivalenten Sinneinheiten, seien es Einzelwörter, Komposita, Mehrwortverbindungen oder Kollokationen zu suchen helfen. Die folgende Übersicht zeigt, welche Typen von Korpora für diese Aufgabe verwendet werden können:

Abb. 1 Relevante Typen der bilingualen elektronischen Korpora geschriebener Sprache



Erfahrungsgemäß können wir konstatieren, dass zu jedem von den dargestellten Korpusarten objektiv festgestellte Pros und Kontra aufgezählt werden können: Parallelkorpora, i.e. Übersetzungskorpora z. B. bringen den Vorteil einer kontrastbildenden Sichtweise. Weniger vorteilhaft jedoch ist bei parallel aufgestellten Textkorpora die Tatsache, dass Übersetzungen zumeist die Spuren der Originalsprache behalten. Sinclair (1997) spricht in diesem Kontext zutreffend von „translationese“. Diesen Einwand können wir aus unserer Praxis bestätigen, sehr wichtig bei der Korpuserstellung ist die Qualität der Übersetzung, die viel Aufwand bei der linguistischen Analyse erspart, je besser sie ist. Das Problem der „translationese“ ist jedoch erheblich größeres im Vergleich zu der mangelnden Repräsentativität, die bei kleineren bi- oder multidirektionalen Korpora ebenfalls als Störfaktor festgestellt werden kann. Die angeführten Problemkreise, deren Bestehen sich bei unserer Erfahrung mit Parallelkorpora bestätigt hat, führten uns bei unseren Überlegungen über die unmittelbaren kontextuellen Relationen zu dem Konzept des Kontrastkorpus, das in beiden Sprachen authentische, durch die Übersetzung unverzerrte, Daten liefern würde.

Kontrastkorpora<sup>1</sup> basieren prinzipiell auf der Zusammenstellung von Texten in mehr als einer Sprache oder Varietät, wobei hervorgehoben werden muss, dass der Anwendungsbereich dieser Korpora in ähnlichen Kommunikationsbedingungen liegt, insbesondere bei der vergleichenden Sprachbetrachtung. Obwohl wir uns hier Sinclairs Feststellung anschließen, dass im wissenschaftlichen Diskurs immer noch keine Übereinstimmung bezüglich der Kriterien für die Vergleichbarkeit (vgl. Sinclair 1997, S. 126) herrscht, versuchten wir das vorliegende Korpus nach Themenbezogenheit und Textsortenspezifität zu gestalten. Die Vorteile dieses Korpusartus liegen vor allem in der durch die Übersetzung nicht verzerrten textuellen Authentizität des zu vergleichenden Materials.

Diese Darstellung der Korpusartologie könnte für die Vollständigkeit noch mit einem Beispiel ergänzt werden: Lauridsen (1996) führt in ihrer Korpusartologie eine

<sup>1</sup> Sinclair bezeichnet diesen Korpusartus als *vergleichbares Korpus* (siehe mehr bei Sinclair 1997, S. 126) Als Beispiel wird das *International Corpus of English – ICE* vorgestellt: ICE stellt eine Zusammenrückung von Korpora in den entsprechenden Varietäten des Englischen in der gleichen Zeitspanne und in entsprechenden Textgenres des gleichen Umfangs dar.

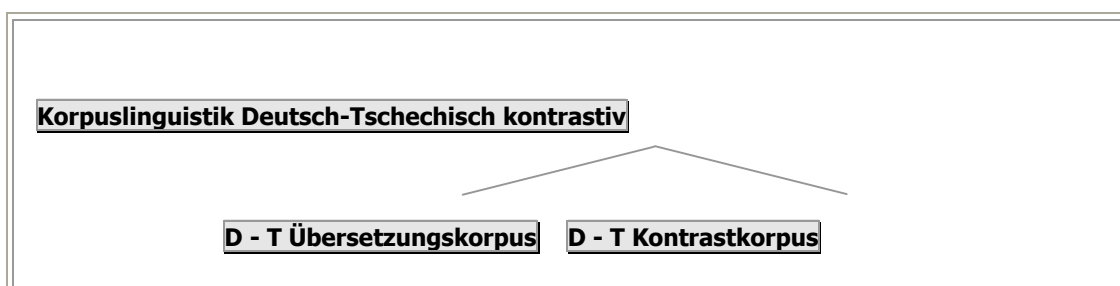
weitere Klasse der Lerner-Korpora<sup>2</sup> ein; die Texte werden von Lernern einer Fremdsprache geschaffen – entweder bei freier Sprachproduktion oder durch Übersetzungen. Die erste Subklasse muss nicht notwendig bilingual sein, bei der monolingualen Ausrichtung dient dieser Korpustyp entweder zur Überprüfung des aktuellen Sprachniveaus oder bei längerer Textaufnahme dient das Korpus zur Feststellung der Entfaltung der Sprachkompetenz bei Lernern.

## 2. Bilinguale Korpora im deutsch-tschechischen Sprachvergleich

Den gemeinsamen Stützpunkt der bereits laufenden Projekte des deutsch-tschechischen Sprachvergleichs stellen die Prinzipien der kontrastiven Linguistik dar, die sich traditionell in der Auslandsgermanistik großem Interesse erfreut. Dies ist unter anderem auch dadurch zu begründen, dass sprachvergleichende Ansätze einen großen Wert für die Vermittlung von Deutsch als Fremdsprache besitzen. Insbesondere im Bereich der Fehlerlinguistik, die zu einem bedeutenden Bestandteil der germanistischen Forschungsausrichtung der Auslandsgermanistik gehört, kann dies gezeigt werden<sup>3</sup>. Die Beschäftigung mit Schwierigkeiten bei der Textproduktion, die durch den Einfluss der Muttersprache zustande gekommen sind, stellt einen wichtigen Impuls zum Vergleich von Sprachsystemen dar und führt zu interessanten Fragestellungen in diesem Forschungsbereich.

Wie aus der schematischen Darstellung resultiert, konzentriert sich die Forschungsarbeit auf die folgenden wichtigen Projekte:

Schema 1: Deutsch-tschechische Vergleichskorpora im Forschungsbereich  
*Korpuslinguistik Deutsch-Tschechisch kontrastiv*



1) *Das deutsch-tschechische Übersetzungskorpus* ist als eine auf Übersetzungen von Texten unterschiedlicher Genres basierende Untersuchungsgrundlage konzipiert. Die Texte tschechischer und deutscher Autoren werden in der Ausgabeform synoptisiert, durch die Zeilennummerierung können sie auf der Zeilenebene verglichen werden. Die Suchprozesse können sowohl auf sprachsystematische kontrastive Analysen als auch auf

<sup>2</sup> Dazu vgl. auch Lauridsen (1996, S. 64f.)

<sup>3</sup> Vgl. Kratochvílová u. Rinas (2005, S. 277f.)

bestimmte translatorische Aufgaben ausgerichtet werden. Aufgrund der bisherigen Projektpraxis können wir resümieren: Die Extraktion von Wortübersetzungen aus Korpora wurde bislang meist als Zuordnungsproblem betrachtet: Ausgehend von parallelen Korpora wurde ermittelt, welche Sätze bzw. Wörter des Ausgangstextes mit welchen Sätzen bzw. Wörtern der zugehörigen Übersetzung korrespondieren. Solche Zuordnungsalgorithmen funktionieren in der Praxis sehr gut vor allem bei systembezogenen Fragestellungen<sup>4</sup>. Bei lexikalischen oder übersetzerischen Forschungsvorhaben besteht jedoch häufig das Problem, dass für ein gewünschtes Sprachenpaar und vor allem für eine gewünschte Textsorte oder für themenbezogene Texte oder Genres nur selten ein ausreichend großes paralleles Korpus zur Verfügung steht. Deswegen werden die Möglichkeiten der Korpuslinguistik bei solchen Forschungsvorhaben auf die Errichtung eines anderen Korpusstyps – des Kontrastkorpus – übertragen.

2) Das deutsch-tschechische Kontrastkorpus kann typologisch als eine Datengrundlage von zwei Originalsprachen charakterisiert werden (potentiell wären auch mehrere denkbar), die sowohl für kontrastive lexikologische, textlinguistische bzw. soziokulturelle und übersetzerische Untersuchungen geeignet ist. Die Kontrastkorpora erfassen den natürlichen Sprachgebrauch innerhalb des Genrespektrums, aus dem sie zusammengestellt werden. Ihr Hauptvorteil besteht darin, dass sie unbetroffen von unterschiedlichen durch die Translationsprozesse zu begründenden Verzerrungen eine natürliche Textbasis für den Sprachvergleich darstellen. Kontrastkorpora, in Analogie zu dem englischen Terminus *comparable corpora*, bestehen aus Originaltexten in der jeweiligen Vergleichssprache, die nach Textsorte, Thema oder kommunikativer Funktion synoptisiert bzw. aligniert werden. Sie können auch auf eine spezifische Domäne ausgerichtet sein, in unserem Fall auf die Ereignisse des Kriegs im Irak, wie sie in Nachrichtentexten und Kommentaren beider Vergleichssprachen reflektiert wurden. In diesem Beitrag wird ein Beispiel des konkreten analytischen Verfahrens präsentiert, das konkret auf die textspezifisch rekurrenten Mehrwortverbindungen fokussiert.

Aufgrund der bisherigen Projektpraxis können wir kurz resümieren: Die Extraktion von Wortübersetzungen aus Korpora wurde bislang meist als Zuordnungsproblem betrachtet: Ausgehend von parallelen Korpora wurde ermittelt, welche Sätze bzw. Wörter des Ausgangstextes mit welchen Sätzen bzw. Wörtern der zugehörigen Übersetzung korrespondieren. Obwohl solche Zuordnungsalgorithmen relativ gut funktionieren, besteht in der Praxis das Problem, dass für ein gewünschtes Sprachenpaar und eine gewünschte Textsorte nur selten ein ausreichend großes paralleles Korpus zur Verfügung steht. Deswegen versuchen wir, die Möglichkeiten der Korpuslinguistik auf die

---

<sup>4</sup> Hiermit wird auf den Beitrag von V. Kotůlková verwiesen, die sich im Rahmen ihrer Forschung anhand des deutsch-tschechischen Übersetzungskorpus extensiv mit variablen Äquivalenzmöglichkeiten der deutschen Substantivkomposita beschäftigt.

Errichtung eines Kontrastkorpus zu übertragen, auf den wir in den weiteren Ausführungen die Aufmerksamkeit richten.

An dieser Stelle sei aber auch angemerkt, dass die neueren Trends in der korpuslinguistischen Forschung auf die gegenwärtige technologische Entwicklung flexibel reagieren. Dank neuer leistungsfähiger Technologien muss die Korpuslinguistik nicht mehr eine Herausforderung für spezialisierte Arbeitsstätten sein. Außer der extensiven Korpusarbeit, die sich mit dem Korpusaufbau und mit der Untersuchung der Sprachstrukturen anhand von umfangreichen Datenmengen beschäftigt, konzentriert sich die Forschung ebenfalls auf die pädagogische Auswirkung der Korpuslinguistik, i.e. aufgebaute Korpora sollen entweder direkt im Unterricht eingesetzt werden, da das Korpus die Sprache in vielfältigen Zusammenhängen reflektiert, oder sie können als Grundlage für die Konzipierung von Unterrichtsmaterialien dienen. In der pädagogischen Ausrichtung der Korpuslinguistik geht der Weg über kleinere projektspezifische Korpora, die zum unmittelbaren Einsatz bestimmt sind oder von denen die Grundlagenforschung im Rahmen kleinerer Projekte brauchbare Ergebnisse gewinnt. Korpora sollen deswegen nicht zum Selbstzweck aufgebaut werden, ihr Einsatz ist sinnvoll, wenn die introspektive Sicht des Linguisten hinzukommt.

### **3. Das deutsch-tschechische Kontrastkorpus und die kontrastive Kollokationsanalyse**

Kontrastkorpora, wie bereits erwähnt, erfassen den natürlichen Sprachgebrauch innerhalb des Genrespektrums, aus dem sie zusammengestellt wurden. Sie können ebenfalls auf eine spezifische Domäne ausgerichtet sein, z. B. EU-Legislative oder Medizin. Konkret können z. B. die EU-Judikatur oder Gebrauchsanweisungen für international verwendete Medikamente eine interessante Untersuchungsgrundlage bilden. Einen Spezialtypus der Vergleichskorpora bilden umfangreiche Korpora, die aus einer proportionalisierten Anzahl von repräsentativen Textsorten zusammengestellt werden (vgl. Altenberg/Granger 2001, S. 8ff) und zu unterschiedlichen sprachanalytischen Aufgaben benutzt werden.

Im Mittelpunkt einer kontrastiven Kollokationsanalyse steht ein aktuelles nicht subjektiv konstruiertes Beweismaterial im Sinne der modernen Korpuslinguistik (vgl. Wolf 2004, S. 135f.). Hierfür stellen elektronische, in unserem Fall bilinguale Korpora eine wichtige Voraussetzung dar. Unser Projekt der kontrastiven Kollokationsanalyse geht spezifisch vor und basiert auf einer neuen Herangehensweise an Untersuchungsquellen: auf einem Kontrastkorpus, das nicht mit Übersetzungen arbeitet, wie Parallelkorpora, sondern mit thematisch gebundenen und die Textsortenspezifik reflektierenden Textsammlungen in der jeweiligen Sprache. Man hat also nun für jede Einzelsprache verschiedene Korpora, die jedoch zueinander in Beziehung stehen – durch den gemeinsamen kommunikativen Bereich – oder durch Textreferenten.

Das Interessensfeld bilden die in diesen Texten vorkommenden Wortverbindungen, die als „vorgefertigte“ Textbausteine eine wichtige textbildende Funktion in beiden Sprachen haben. Den Ausgangspunkt der kontrastiven Untersuchung der usuellen Mehrwortverbindungen im Kontrast zweier Sprachen bildet das onomasiologische Herangehen – d. h. man stellt zunächst das zu differenzierende Inventar betreffender lexikalischen Einheiten in den beiden Sprachen auf. Eine wichtige Frage, die hierzu gestellt werden muss, bezieht sich auf die Methode der Datengewinnung, i.e. wie können die uns interessierenden Wortverbindungen, die wir nach unserem typologischen Vorschlag als diskursive Kollokationen bezeichnen, im nicht parallelen Textmaterial, aus dem das Kontrastkorpus kompiliert wurde, festgestellt und extrahiert werden. Es ist verständlich, dass mit der Analyse der benutzten lexikalischen Mittel, die bei größeren Textmengen ohne korpuslinguistische Stütze nicht auskommen kann, etliche, auch nicht präsupponierte Zusammenhänge erschlossen werden können, zumal gerade in der Analyse des spezifischen Sprachgebrauchs in der jeweiligen Sprache auch Antworten auf andere Fragen implizit miteinbezogen sind.

Die Textquellen des Kontrastkorpus entstammen aus Zeitungstexten der deutschen und tschechischen überregionalen Tageszeitungen, der *Süddeutschen Zeitung* und der *Právo*: Sie sind thematisch gebunden, i.e. referieren auf die Ereignisse im Irak 2003. Im Vordergrund für die Auswahl der Texte war vor allem die Vergleichbarkeit, der referentiell analogen i.e. nicht parallel übersetzten Texte. Das öffentliche Interesse und das Bedürfnis nach Informationen zum Thema 'Der Krieg im Irak' ist für beide Presselandschaften Deutschlands und Tschechiens von so großer Bedeutung, dass sich die Inhalte der wichtigsten Nachrichten auf der Titelseite sowie weiterer Artikel im Innern der Zeitungen, Leitartikel und Kommentare mit diesem Thema beinahe in jeder Ausgabe beschäftigen. In diesem Sinne können wir diese referentielle Bindung nicht nur als 'Thema' bezeichnen, sondern, über diese begrifflichen Grenzen hinausgehend, möchten wir solche thematische Zirkulation der sich wiederholenden sprachlichen Mittel, rekurrenten Phrasen, konventionellen aber auch sich erst konventionalisierenden Nominationen, durch die die Realität des Krieges versprachlicht wird, als Irak-Diskurs verstehen.

Eine konkrete Aufteilung der Untersuchungstexte konzentriert sich texttypologisch auf die informativen und meinungsbildenden Texte, d. h. Nachrichten und Kommentare, die einen gemeinsamen Textreferenten haben. Alle für die Untersuchung relevanten Texte in diesen Subkorpora beziehen sich zum Thema Krieg im Irak. Die Texte des Kontrastkorpus wurden selektiert, dokumentationstechnisch aufbereitet und für die weitere Untersuchung in zwei Suchmodule geteilt, jeweils mit einer anderen differenzierten Fokussierung.

Die folgende schematische Abbildung illustriert die unterschiedliche Ausrichtung beider Suchmodule des Kontrastkorpus:

Abb. 3 Strukturierung der Suchprozesse im Kontrastkorpus

<b>Modul I</b>		<b>Modul II</b>	
<b>Strukturierungskriterium:</b>	Textsorten des gesamten Jahrgangs in beiden Vergleichssprachen	<b>Strukturierungskriterium:</b>	Referenzielle Analogie für die Vergleichssprachen
<b>Textquellen:</b>	Nachrichten D, Cz Kommentare D, Cz	<b>Textquellen:</b>	Nachrichten D+Cz
<b>Format:</b>	Access-Database D, Cz	<b>Format (Oberfläche):</b>	Html, alligniert D+Cz
<b>Suchprozesse</b>	Signifikante Bigramme im Kontext statistische Daten Assoziationsmaße	<b>Suchprozesse</b>	Full-Text-Suche KWIC Schlagzeilen

**Modul I** (MI) liefert eine umfangreiche sortierte Datenbasis deutscher und tschechischer Nachrichten und Kommentare, die den ganzen Jahrgang abdeckt. Der Suchprozess konzentriert sich auf die Frage, bei welchen Zweierverbindungen von Wörtern in den untersuchten Textsorten eine statistische Signifikanz gezeigt werden kann, nach der diesen Bigrammen der Status 'Kollokationskandidaten' zugesagt werden könnte. Die gesamten Irak-bezogenen Texte wurden, jeweils nach Textsorten und Sprache, auf Wortverbindungen durchsucht. Diese Teilung ermöglichte es, die belegten Verbindungen von zwei aufeinander folgenden relevanten Wortformen, die im Weiteren als Bigramme bezeichnet werden, für jede Sprache und Textsorte separat zu extrahieren. Das Untersuchungsinteresse richtete sich, wie oben angedeutet wurde, vor allem auf den Vergleich von textsortenspezifischen Unterschieden in deutschen und tschechischen, informativen und meinungsbildenden Texten.

**Modul II** (MII) stellt in seiner Struktur das eigentliche Kontrastkorpus dar. Das auf der Textebene alignierte Kontrastkorpus, i.e. das Korpus bildet die sich thematisch und zeitlich entsprechenden Nachrichtentexte parallel ab, konzentriert sich explizit auf die sich analogisch entsprechenden Nachrichten mit der Thematik *Irak-Krieg*, die in der Süddeutschen Zeitung und in Právo im gesamten Jahrgang 2003 publiziert wurden. Die vom Modul I gewonnenen Suchergebnisse, bei denen die auf die Signifikanz des gemeinsamen Vorkommens hinweisende statistische Werte als differenzierendes Kriterium miteinbezogen wurden, bildeten einen wichtigen Ausgangspunkt für die weiteren Fragestellungen, die sich in der Gestaltung der Suchprozesse in diesem Kontrast-Korpus widerspiegeln. Zusammenfassend halten wir fest: Die Kollokationsforschung braucht bei domänenspezifischen Untersuchungen, wie z. B. im Falle einer fachsprachlichen Untersuchung, oder in unserem konkreten Fall, einer

themenbezogenen Studie, auch spezifische hochaktuelle kleinere Korpora, die einen spezifischen Teil der sonst kaum überschaubaren Menge der Kollokationen einer Sprache erfassen können. Deswegen wurden die Korpusarchitektur und die benutzten Werkzeuge der Anforderung an Zugänglichkeit und wenig komplizierter Handhabung angepasst. Der Grundgedanke war, dass die Methodologie als Ansatz für die Zusammenstellung neuer aktueller Textkorpora appliziert werden könnte, bei denen der eventuelle Einsatz im L2-Lernprozess denkbar wäre.

#### **4. Weitere Forschungsperspektive:**

##### ***Das Projekt Korpuslinguistik Deutsch-Tschechisch kontrastiv***

Den Leitfaden des dargestellten korpuslinguistischen Ansatzes bildete die Erfahrung aus den bereits laufenden Korpusprojekten, dass der Aufbau großer Korpora eine überaus anspruchsvolle Aufgabe ist, deren Effizienz von der ständigen Aktualisierung sowie der ununterbrochenen Fachadministration abhängt. Aus diesem Grunde zeigt sich der extensive Aufbau eines Forschungsteams für sehr sinnvoll. Die Zusammenarbeit zwischen dem Institut für deutsche Philologie und dem Kompetenzzentrum für die EDV-Philologie der Universität Würzburg und der germanistischen Arbeitsstätte in Opava gewann im Projekt *Korpuslinguistik Deutsch-Tschechisch kontrastiv* einen weiteren Kooperationspartner – die Germanistik an der Universität Ostrava. Auf diese Weise konzentriert sich die seit einigen Jahren sich entwickelnde Zusammenarbeit von Würzburg, Opava und Ostrava gemeinsam auf einen Forschungsbereich – auf die korpusbasierte kontrastive Erforschung der relevanten Textphänomene, die sich auf konkrete grammatisch und lexikalisch ausgerichtete Teiluntersuchungen stützt: Die Erforschung der Redewiedergabe in Texten, der Zeit- und Raumlinguistik oder der syntaktischen Funktionen von Infinitiven in den Vergleichssprachen können als konkrete Beispiele die vorgesehenen Ausrichtung dieses korpuslinguistischen Vorhabens angeführt werden.

Abschließend können wir zusammenfassen: Für die Forschungsoptik dieses korpuslinguistischen Projektes in aller Vielfalt seiner Teilbereiche sind die beiden genannten Typen der Sprachkorpora von Relevanz, i.e. sowohl die Parallelkorpora als auch die Kontrastkorpora. Sie sind eine begründbare Stütze für die kontrastiven Herangehensweise, insbesondere bei nicht verwandten Sprachsystemen, bei denen die Frage der Äquivalenz eine überragende Rolle spielt. Vergleichende Sprachbetrachtungen, sollen sie Vollständigkeit und Objektivität anstreben, können deshalb mit bilingualen Korpora als Untersuchungsbasis ihren Zielsetzungen sehr gut nachkommen.



## Literaturverzeichnis

- ČERMÁK, František (1995): Komputační lexikografie. In: Čermák, František; Blatná, Renata (Hg.): *Manuál lexikografie*. Praha.
- KRATOCHVÍLOVÁ, Iva (2004): Usuelle Wortverbindungen: Fokussierung des aktuellen Themas „Kollokabilität lexikalischer Einheiten“ auf den Bereich des deutsch-tschechischen Sprachkontrast. In: *Brünner Beiträge zur Germanistik und Nordistik*. Jg. 9, S. 97–107.
- KRATOCHVÍLOVÁ, Iva/RINAS, Karsten (2005): Das deutsch-tschechische Fehlerkorpus. Ein Projekt zur systematischen Erfassung der häufigsten Interferenzfehler tschechischer Sprecher. In: Schwitalla, Johannes; Wegstein, Werner (Hg.): *Korpuslinguistik deutsch: synchron-diachron-kontrastiv*. Tübingen.
- KRATOCHVÍLOVÁ, Iva (2006): Kollokationen: Mehr oder weniger feste Wortverbindungen in Lexikon und Text. Überlegungen zu einer begrifflichen Abgrenzung der Mehrwortlexik. In: *Brünner Beiträge zur Germanistik und Nordistik*. Jg. 11. (i.D.).
- LAURIDSEN, Karen. M. (1996): Text corpora and contrastive linguistics: Which type of corpus for which type of analysis? In: Aijmer, Karin; Altenberg, Bengt; Johansson, Mats (Hg.): *Languages in Contrast*. Papers from a Symposium on Text-based Cross-linguistic Studies Lund 4–5 March 1994. Lund
- LEMNITZER, Lothar/ZINSMEISTER, Heike (2006): *Korpuslinguistik*. Eine Einführung. Tübingen.
- ROTHKEGEL, Annely (1994): Kollokationssbildung und Textbildung. In: Sandig, Barbara (Hg.): *Europhras 92*. Tendenzen der Phraseologieforschung. Bochum, S. 499–523.
- SINCLAIR, John (1991): *Corpus, Concordance, Collocation*. Oxford.
- SINCLAIR, John (1997): Corpus evidence in language description. In: Wichmann, Anne; Fligelstone, Steve (Hg.): *Teaching and Language Corpora*. London, New York, S. 27–39.
- TEUBERT, Wolfgang (2002): The role of parallel corpora in translation and multilingual lexicography. In: Altenberg, Bengt; Granger, Sylviane (Hg.): *Lexis in Contrast*. Corpus based approaches. Amsterdam; Philadelphia.
- WOLF, Norbert R. (2004): Plädoyer für eine Korpuslinguistik. In: Kratochvílová, Iva; Vaňková, Lenka (Hg.): *Germanistik im Spiegel der Generationen*. Opava, Ostrava.