

Zur Bearbeitung von 200 Lemmata

für das Große Deutsch-Tschechische akademische Wörterbuch

Ondřej Dušek

Schriftliche Arbeit in Lexikologie

Leitung: PhDr. Marie Vachková, Ph.D.

2009

Inhaltsverzeichnis

Inhaltsverzeichnis.....	2
1 Einleitung.....	3
1.1 Aufgabe	3
1.2 Analyse	3
2 Lemmaübersicht	4
2.1 Wortliste	4
2.2 Morphologische Charakteristik.....	5
2.3 Gruppierung der Wörter.....	5
2.4 Stilistische Markierung.....	6
2.5 Wortbildung.....	7
3 Arbeit mit den Wörterbüchern	8
3.1 Die verwendeten Lexika und andere Hilfsmittel.....	8
3.2 Probleme bei Äquivalentsuche	9
3.3 Probleme mit den Wörterbuchdefinitionen.....	11
4 Die Korpusanalyse.....	12
4.1 Theoretische Vorüberlegungen.....	12
4.2 Konfrontation des Wörterbuchmaterials mit den Texten im Korpus	13
5 Schlussbemerkungen.....	15
Quellenverzeichnis	16
Verwendete Lexika, Korpora und andere Internetquellen.....	17
Liste der bearbeiteten Lemmata.....	18
Beispiele der bearbeiteten Lemmata	19

1 Einleitung

1.1 Aufgabe

Das Projekt des *Großen Deutsch-Tschechischen Akademischen Wörterbuchs* (GDTAW) läuft schon seit 2000 an der Philosophischen Fakultät der Karls-Universität Prag (mehr dazu z.B. Vachková 2007). Das Ziel ist ein ca. 130.000 Stichwörter umfassendes zweisprachiges Wörterbuch zu entwickeln, wobei sowohl die bestehenden Wörterbücher, als auch empirische Korpusanalysen des berücksichtigt werden. Es wird dabei hauptsächlich mit dem *Deutschen Referenz-Korpus* (DeReKo) des IDS-Mannheim gearbeitet. Die Stichwortliste stützt sich auf das von dem Duden-Verlag herausgegebene *Deutsche Universalwörterbuch* (DUW, s. dazu Vachková 2007:33f.).

Im Rahmen dieses Projekts habe ich insgesamt 211 Lemmata bearbeitet. Es handelte sich um eine Auswahl der Substantive mit den Anfangsbuchstaben T und W-Z. Meine Aufgabe war, die Rechtschreibung, stilistische Merkmale und Morphologie des Worts nach dem DUW zu beschreiben, dann aufgrund der Analyse im Korpus (DeReKo) nach nützlichen Kollokationen, Fügungen oder Beispielsätzen zu suchen und die möglichst passenden Übersetzungen ins Tschechische für die einzelnen Wortbedeutungen zu finden. Alle auf diese Weise erworbenen Informationen wurden dann in einen nummerierten Dateninputraster eingegeben, der sowohl verschiedene Rubriken zur Rechtschreibung, Morphologie, Stilistik, Übersetzung und zu Kollokationen der einzelnen Wortbedeutungen, als auch einige Eingabefelder für Textbelege und Anmerkungen enthält, und nimmt somit die Struktur des künftigen Wörterbuchartikels vorweg (vgl. Zelená 2008). Es handelt sich allerdings lediglich um die erste Erstellung der Datenbankeinträge für die einzelnen Wörter; diese unterliegen noch weiterer Kontrolle und im Falle der Fachtermini auch der Begutachtung des Experten.

1.2 Analyse

In dieser Arbeit möchte ich zuerst eine Beschreibung der mir zugeteilten Lemmata erstellen und dann mein Vorgehen bei der Bearbeitung ihrer Datenbankeinträge anhand einiger theoretischer Voraussetzungen kommentieren.

Es wird zunächst eine Statistik verschiedener Eigenschaften der Wörter folgen, dann die Aufzählung einiger Gruppen mit gleichen typischen Merkmalen und ihre kurze Besprechung aus der Sicht der Morphologie, Semantik, Stilistik und Wortbildung.

Dann werde ich den Verlauf der Suche nach den tschechischen Übersetzungen der einzelnen Bedeutungen behandeln, bzw. auch die dabei angetroffenen Probleme, die sich sowohl aus dem Kontrast der zwei Sprachen und deren Aufbau des Wortschatzes, als auch aus den Wörterbuchdefinitionen selbst ergeben.

Ich werde dann einige theoretische Vorüberlegungen zur Korpusanalyse, bzw. Kookkurrenzanalyse vorstellen. Dann möchte ich auch die praktischen Ergebnisse der Analyse mit der Beschreibung im Wörterbuch vergleichen, und auch die Vor- und Nachteile der Korpusforschung und ihr Beitrag für die Arbeit an den Lemmata besprechen.

2 Lemmaübersicht

2.1 Wortliste

Die Liste der Stichwörter, die ich zur Bearbeitung bekommen habe, war nur der alphabetischen Position nach gewählt worden. Deshalb haben die Wörter im Allgemeinen nichts anderes gemeinsam – in manchen Aspekten weist also die Wortliste Merkmale einer Zufallsauswahl auf, man muss aber immer Rücksicht darauf nehmen, dass sie alle für das Wörterbuch bestimmt sind.

Dies gilt z. B. für die Trefferanzahl der einzelnen Wörter im DeReKo. Bei einer ganz zufälligen Wortmenge würden die absteigend geordneten Trefferzahlen einer Zipfschen Verteilung (Manning & Schütze 1999:23) entsprechen¹. Es zeigt sich, dass dies hier nicht der Fall ist: die Substantive mit größerer Anzahl der Korpusbelege sind stärker vertreten – was auch dem Zweck des Wörterbuchs gut entspricht. Das häufigste Wort in der Menge ist *Tag* mit über 1 Million Vorkommen in dem 3-Milliarden-Korpus, etwa zehn Wörter sind in DeReKo gar nicht zu finden. Der Durchschnitt der Trefferzahlen liegt zwar über 7.000, aber der Median ist 186 – bei den meisten Wörter arbeitete ich mit einigen Zehnten bis Hunderten von Korpusbelegen.

1 Das Produkt der Position in der Reihe und der Trefferzahl würde etwa konstant bleiben.
Schriftliche Arbeit in Lexikologie – Ondřej Dušek

Alle Lemmata sind im DUW aufgenommen, mit der Ausnahme des Worts *Tapp*, das nur in den älteren Versionen des Wörterbuchs enthalten ist. Bei den meisten Wörtern von der besprochenen Menge gibt DUW nur eine Bedeutung an. Eine kleinere Anzahl ist nach DUW zweideutig. Eine höhere Anzahl der postulierten Bedeutungen ist sehr selten, die Höchste (5) weisen *Tag* und *Wolke* auf.

2.2 Morphologische Charakteristik

Die Verteilung der Lemmata nach deren Genus ergab ein sehr interessantes Resultat: mehr als die Hälfte der Wörter sind Feminina (121). Dies hängt z.T. mit deren historischen Entwicklung (u. a. mit deren Wortbildung) zusammen (siehe 2.5). Die Vertretung der Maskulina und Neutra war etwa gleich (41, bzw. 35), den Rest bildeten Substantive mit zwei möglichen Genera – substantivierte Adjektive und einige spezielle Fälle (*der/das Zmorge/Zmorgen* usw.).

2.3 Gruppierung der Wörter

Obwohl die zu beschreibenden Wörter nach der alphabetischen Position gewählt wurden, lassen sich unter ihnen einige spezifische Gruppen finden, die ähnliche Merkmale aus dem Gesichtspunkt der Semantik, Wortbildung oder Verwendungsart aufweisen.

Die Mehrheit der Liste bilden Komposita, die jedoch im Allgemeinen keine weiteren Besonderheiten aufweisen. Es ist nur zu erwähnen, dass viele von ihnen aus der Fachsprache² stammen, wie es im Deutschen öfters passiert. Dank der alphabetischen Nähe bekam ich auch einige Gruppen der Komposita zur Bearbeitung, die mit demselben Bestimmungsglied beginnen, z. B. *Tarifangestellte*, *-abschluss*, *-auseinandersetzung*, *-bezirk*, *-gebiet*, *-streitigkeit*, *-vereinbarung* oder *Tatbeweis*, *-mensch*, *-motiv*, *-verdacht*, *-verdächtige*. In den meisten Fällen bedeutete dies einen Vorteil, weil sich die Wörter – was ihren Bezug auf eine Verwendungsdomäne angeht – ganz nahe waren (wie die Wörter auf *Tarif*-).

Die am kompliziertesten zu behandelnde Gruppe der Wörter bildeten einige ganz allgemeine Elemente aus dem Grundwortschatz, wie *Taube*, *Wolke*, *Würze*, *Wüste* oder *Wunde*. Bei ihnen begegnet man sehr häufig der Polysemie und metaphorischen

² Die hohe Anzahl der Fachausdrücke im Allgemeinen ist teilweise auch der alphabetischen Anordnung zuzurechnen – unter den Buchstaben X, Y und Z sind zahlreiche Lehnwörter zu erwarten.

Verwendungsweisen. Ihre Vertretung im Korpus ist in der Regel sehr stark, was zwar sehr viele Informationen, aber auch Probleme bereitet (siehe 4.2).

Am anderen Ende der Häufigkeitsskala stehen andere einfache Wörter, die aber nicht ganz allgemein verwendet werden. Meistens handelt es sich um einigermaßen stilistisch gefärbte Ausdrücke, sei es veraltete, gehobene oder landschaftliche und scherzhafte Wörter. Beispiele dazu sind *Woge*, *Wirre*, *Wruke*, *Zmorge*, *Zofe*, *Zosse* oder *Zote*. Gerade weil sie nicht zu der allgemeinen Standardsprache gehören, begegnet man ihnen im Korpus nur selten oder überhaupt nicht.

Unter den Lemmata finden sich auch einige Fachausdrücke, die nicht zusammengesetzt sind. Es handelt sich meistens um Benennungen der Instrumente, wie *Zwinge*, *Zange* oder *Winde*. Sie sind im Korpus in der Regel stark vertreten.

Eine ganz spezifische Gruppe bilden die Wörter griechischer oder lateinischer Herkunft. Hier geht es ohne Ausnahme um wissenschaftliche Benennungen, oft aus dem Bereich der Biologie und Medizin, wie *Tabes*, *Xylophage*, *Zygote* oder *Zystoskopie*. Die Anzahl ihrer Korpusbelege ist sehr unterschiedlich. Zu einigen solchen Wörtern auf Z- (wie *Ziliate*) begegnet man auch der Schreibvariante auf C-. Nur ein Lemma (*Tabula gratulatoria*) hat seine ursprüngliche lateinische Pluralendung bewahrt.

Andere Lehnwörter unterschieden sich in einigen Aspekten von den ursprünglich griechischen und lateinischen Ausdrücken. Es handelt sich meist um Entlehnungen neueren Datums oder ganz seltene ältere Benennungen, deshalb sind es oft Neutra mit der Pluralendung -s. Sehr häufig waren Wörter aus der Sprachen des Fernost, wie *Tai-Chi*, *Taekwondo* oder *Tamagotchi*, andere Herkunftsgebiete kamen jedoch auch vor: *World Wide Web*, *Tartufo*, *Woiwod*.

Zu erwähnen bleiben noch einige sehr kleine, aber ganz kompakte Gruppen. Erstens sind es vier naturwissenschaftliche Gattungsnamen, wie *Wildkatze* oder *Wolfsspinne*. Dann auch geographische Bezeichnungen (insgesamt zwei) – *Tabor* und *Zaire*. Die letzte Gruppe bilden zwei klangmalerische Ausdrücke – *Tatütata* und *Tapp*.

2.4 Stilistische Markierung

Relativ viele Wörter waren mit einer stilistischen Markierung im DUW gekennzeichnet. Nur leicht über die Hälfte der Ausdrücke ist allgemein neutral, bei manchen von ihnen ist

nur eine der Bedeutungen stilistisch gefärbt. Die einzelnen Markierungen sind aber sehr unterschiedlich – die Differenzierung im DUW ist sehr fein. Am häufigsten sind also die umgangssprachlichen Ausdrücke vertreten (8 Wörter von 91 stilistisch markierten), dann Fachausdrücke der Mathematik³ (6), Kochkunst (5), Zoologie, Autotechnik und Medizin (4). Bei den stilistischen Färbungen der einzelnen Bedeutungen überwiegen dagegen die Bezeichnungen für außerordentliche Sprachschichten, wie *umgangssprachlich*, *salopp*, und *veraltet* (jeweils 2 von insgesamt 23 Bedeutungsmarkierungen).

2.5 Wortbildung

An der Menge der bearbeiteten Wörter lassen sich verschiedene Wortbildungsphänomene sehr gut beobachten. Es geht vor allem um die Komposition, aber auch um die Suffigierung und den damit zusammenhängenden Wortartwechsel. Viele von diesen Substantiven haben ihre Herkunft in anderen Wortarten – dieser Tatsache entspricht auch der hohe Anteil der Feminina: viele durch Suffigierung entstandene deverbative und deadjektivische Substantive haben eben diesen Genus (Fleischer & Barz 1992:146, 158, 172).

Wie schon gesagt, die Komposita bilden eine bedeutende Gruppe in der Wortliste. Die meisten sind aus zwei Teilwörtern zusammengesetzt, es finden sich aber auch dreiteilige Fügungen (*Zahnpastatube*, *Taschengeldentzug*). Fast alle diese Verbindungen sind Determinativkomposita (Fleischer & Barz 1992:245f.), bei manchen lässt sich ihre interne Syntax durch Genitiv erklären (*Tabakschnupfen*, *Tabuverletzung*, *Tachonadel*), andere bestimmen nur näher den Typ des Grundworts (*Tangahöschen*, *Taftbindung*), viele sind aber komplizierter oder nicht eindeutig konstruiert (*Tapezierbürste*, *Tangenssatz*). In der Menge befindet sich auch ein Kopulativkompositum – *Taubblindheit*.

Auch die Derivation durch Suffigierung betrifft sehr viele Wörter. Es geht um einige deadjektivische Abstrakta mit dem Suffix *-keit/-heit*, wie *Taubblindheit*, *Tätigkeit* u.a., weiter andere deverbative Substantive mit dem Suffix *-ung*, wie *Tätowierung* oder *Talentförderung*, sehr häufig taucht auch das Suffix *-e* auf. Bei ihm handelt es sich jedoch um zwei verschiedene Wortbildungssituationen – erstens die Deadjektive, meist Abstrakta, die entsprechende Eigenschaft ausdrücken (*Wohndichte*, *Wohnfläche*), und zweitens die Deverbativa, Benennungen der Tätigkeit wie *Zusage* oder der Instrumente

³ Dies ist auch den Anfangsbuchstaben *X* zu verdanken.
Schriftliche Arbeit in Lexikologie – Ondřej Dušek

wie *Winde* oder *Tapezierbürste* (vgl. Fleischer & Barz 1992:146-148). Im Bereich der Derivation sind noch die Diminutiva zu nennen – man begegnet hier sowohl den beiden standarddeutschen Diminutivsuffixen *-chen*, *-lein* (*Tätzchen*, *Tätzlein*), als auch dem süddeutschen umgangssprachlichen Suffix *-erl* (*Tascherl*).

Auch andere Wortbildungsarten kann man an der besprochenen Wortmenge beobachten: die Konversion an den substantivierten Adjektiven wie *Wohnsitzlose* oder *Wilde* oder an Substantiven, die aus Interjektionen (*Tatütata*) oder Phrasen (*Zuhause*, *Zmorge*) entstanden sind, und die implizite Derivation am letzten Glied des Kompositums *Taschengeldentzug*.

Eine von der Wortbildung her besonders interessante Gruppe bilden die Wörter *Takel*, *Takelage*, *Takelung* und *Takelwerk*. Sie sind alle fast synonymisch (mindestens jeweils in einer Bedeutung), werden aber mit verschiedenen Suffixen gebildet. Dies könnte im Falle der Suffixe *-ung* und *-age*⁴ konnotativ das Ergebnis des Auftakelns ausdrücken, während das Suffix *-werk* eher die Gesamtheit der Konstruktion betont.

3 Arbeit mit den Wörterbüchern

3.1 Die verwendeten Lexika und andere Hilfsmittel

Weil die Stichwortliste und auch die Einteilung der einzelnen Bedeutungen vom DUW ausgehen, richtete ich mich bei der Arbeit immer zuerst nach diesem Standardwerk. Es diente mir auch als Basis-Quelle der morphologischen und stilistischen Daten.

Erst wenn mir eine Wortbedeutung von der Definition im DUW nicht klar war, benutzte ich andere Lexika. Dabei bevorzugte ich immer die elektronischen – die Suche in ihnen ist viel schneller und obwohl sich vor allem die deutsch-tschechischen elektronischen Wörterbücher von der Qualität der Bearbeitung her mit dem größten gedruckten *Wörterbuch von Hugo Siebenschein* (SIEB) gar nicht messen können, sind sie aktueller und beinhalten oft auch sehr seltene Ausdrücke. Die zweisprachigen Lexika dienten mir allerdings hauptsächlich zur Orientierung – obgleich sie in manchen Fällen die einzelnen Bedeutungen gar nicht differenzieren, war die ganze Situation nach dem Vergleich mit dem DUW meistens klar.

4 Der erste Suffix ist einheimisch, der zweite französischer Herkunft.
Schriftliche Arbeit in Lexikologie – Ondřej Dušek

Von den einsprachigen deutschen Quellen habe ich mehrmals das *Digitale Wörterbuch der deutschen Sprache* (DWDS) der Brandenburgischen Akademie der Wissenschaften benutzt, um die Definition des Worts zu überprüfen. Für weitere Informationen über die morphologische Charakteristik (hauptsächlich bei den Komposita, weil DUW da nur selten detaillierten Bericht gibt) diente mir die Internetgrammatik *Canoo* (CAN). Ich verwendete in einigen Fällen auch das *Deutsche Wortschatz-Portal* (DWP) der Leipziger Universität, meist um die Liste der Synonyme zu erweitern.

Unter den zweisprachigen Lexika benutzte ich häufig neben dem *Lingea Lexikon* (LL) auch das Wörterbuch *Seznam Slovník* (SS) und in manchen Fällen – wenn die deutsch-tschechischen Quellen keine Auskunft boten – auch deutsch-englische Lexika, vor allem um den Sinn des Wortes besser zu verstehen – *Leo Deutsch-Englisches Wörterbuch* (LEO) und *Dict.cc – German-English Dictionary* (CC). SIEB verwendete ich nur selten.

Auf der tschechischen Seite überprüfte ich oft mögliche Varianten der Übersetzung im Tschechischen Nationalkorpus (*Český národní korpus*, ČNK). Wenn es sich um Fragen der Rechtschreibung-Varianten handelte, benutzte ich das Internet-Handbuch *Jazyková příručka* (JP) der Tschechischen Akademie der Wissenschaften.

Sehr oft, meist bei den Fachausdrücken, half mir bei der Suche nach der Übersetzung die Internet-Enzyklopädie *Wikipedia* (WP). Wenn es in ihr einen Artikel über dasselbe Objekt im Deutschen und im Tschechischen gibt, sind die beiden miteinander verbunden und dadurch ist die Übersetzung möglich. Öfters, wenn ein Ausdruck in den Lexika nicht zu finden war, benutzte ich auch die *Google-Suche* (GOO). Man darf hier nicht nur nach dem betroffenen deutschen Ausdruck oder Übersetzungskandidaten suchen, sondern es auch versuchen, das deutsche Wort nur in den tschechisch geschriebenen Webseiten zu finden.

3.2 Probleme bei Äquivalentsuche

Das immer wiederkehrende Problem bei der Äquivalentwahl ist die Tatsache, dass es für eine Wortbedeutung nur sehr selten eine 1:1-Korrespondenz der Ausdrücke gibt. Nach Čermák (1995:238f.) ist dies eigentlich nur bei den fachsprachlichen Termini der Fall, was sich hier in der Regel bestätigt hat. Sehr oft musste ich zwei oder mehr tschechische Äquivalente, unter denen nur im Kontext zu wählen ist, für den deutschen Ausdruck angeben, z. B. *debata, beseda, rozprava, pokec* für *Talk*.

Ein weiteres Problem der Wortschatzunterschiede ist das Fehlen eines lexikalisierten tschechischen Ausdrucks. Dieser muss dann durch Beschreibung oder freie Wortverbindung ersetzt werden (Čermák 1995:ebd.). Einer solchen Situation bin ich z.B. bei den Wörter *Talisch* und *Wippe* begegnet. Der erste Ausdruck ist ein Lehnwort, das vielleicht im Tschechischen nach bloßer Anpassung der Orthographie verwendet werden könnte, ich konnte aber nur einen Beleg im Internet dafür finden; deshalb musste ich eine Beschreibung angeben. Im Falle von *Wippe* geht es darum, dass es im Tschechischen keinen Einwortausdruck (*houpačka* wäre zu allgemein) und auch keine feste Wortverbindung dafür gibt. Daher muss die Übersetzung *houpačka* noch eine weitere Bestimmungsangabe enthalten, nämlich *zespodu podepřená*.

Damit hängt eine andere Situation zusammen, die ich sehr oft angetroffen habe, und zwar die Übersetzung der Komposita. Komposition ist im Tschechischen sehr selten im Unterschied zum Deutschen, deshalb mussten in der Regel Mehrwortausdrücke verwendet werden – und diese sind im Tschechischen öfters nicht fest geregelt, es gibt mehrere Varianten und keine überwiegt eindeutig. Deshalb wurden bei vielen Komposita als tschechische Äquivalente mehrere synonymische Wortverbindungen angegeben, wie z.B. im Falle von *Zahnpastatube* - *tuba zubní pasty*, *tuba se zubní pastou*, *tuba na zubní pastu*, *tuba od zubní pasty*. Manche Verbindungen sind doch auch im Tschechischen stabilisiert und terminologisiert, es gibt keine häufiger verwendeten Alternativen, z.B. *Zahnfäule* ist auf Tschechisch eindeutig *zubní kaz*.

Bei den Lehnwörtern geht es dagegen häufig nur um eine Anpassung der Rechtschreibung, weil sie international unverändert verwendet werden, wie z.B. *Tamagotchi* oder *Tai-Chi*. Um ihre Äquivalente zu erschließen, benutzte ich oft WP, dann auch GOO und vor allem JP, um die richtige Rechtschreibung-Variante zu wählen. So habe ich für das erste Beispiel nur eine Möglichkeit, *tamagoči*, gefunden, aber für das Zweite sind im Tschechischen drei Schreibungen erlaubt: *tai-či*, *tchaj-t'i* und *taj-či*.

Damit hängen auch die Fachausdrücke zusammen, bei denen es sich z.T. auch um Lehnwörter handelt. Hier habe ich jeweils Übersetzungskandidaten gefunden, die von einem Experten noch geprüft werden müssen. WP mit weiterer Kontrolle in GOO oder ČNK hat sich hier gut bewährt; falls ein Ausdruck in WP nicht zu finden war⁵ (wie z.B.

5 Weil die deutsche WP viel mehr Stichwörter als die Tschechische enthält, bedeutet dies eher, dass nur *Schriftliche Arbeit in Lexikologie – Ondřej Dušek* 10

Talisch), musste ich ihn im tschechischen Web auf Deutsch suchen, bzw. hypothetische tschechische Rechtschreibungen der Suchmaschine zur Bestätigung vorlegen.

3.3 Probleme mit den Wörterbuchdefinitionen

Schon die semantische Beschreibung der einzelnen Einträge im einsprachigen Wörterbuch an sich ist eine sehr schwierige Aufgabe. Wie Hanks (2008a:233-235) berichtet, ist diese Arbeit noch durch die lange Aristotelische und Leibnizsche Tradition eher unpraktisch belastet. Die Beschreibungen im Wörterbuch sind auf die Substituierbarkeit mit dem definierten Wort gerichtet, und deshalb tendieren sie eher zur zwar genauen, aber wenig anwendbaren Paraphrasierung. Oft kann dabei aber von einer Definition im Sinne der Logik nur unter Vorbehalt oder gar nicht die Rede sein (Harras, Haß, & Strauß 1991:13-18). Über die Art der Verwendung des konkreten Wortes besagen die Beschreibungen oft auch nicht viel (vgl. Schippan 2002:58; Vachková 2007:107ff.).

Wenn solche einsprachige Beschreibungen für die Zwecke der zweisprachigen Lexikographie angewandt werden, wie die Einträge vom DUW im Falle der vorliegenden Arbeit, zeigen sich etliche Probleme ganz deutlich. Erstens, es kommt hier ganz oft gerade zur Paraphrasierung in den Definitionen, wie. z. B. *Wortfolge = Folge von Wörtern*, wo die Ambiguität, die man an den möglichen tschechischen Äquivalenten *slovořled* (Anordnung der Satzglieder) und *sled několika slov* (Reihe einiger Wörter) illustrieren kann, gar nicht ausgedrückt wird. Zweitens, oft begegnet man einer sehr feinen Bedeutungsgliederung, wobei die Übersetzung ins Tschechische stets dieselbe bleibt – es geht um eine metonymische Verwendung, die in beiden Sprachen gleich verläuft, wie z.B. bei *Tamagotchi* als das virtuelle Tier und als das Spielzeug selbst, materiell. Oft wird auch die sog. reguläre Mehrdeutigkeit (Schippan 2002:162) explizit beschrieben, vgl. *Tätowierung* als Tätigkeit und ihr Ergebnis. Für die Übertragung ins Tschechische nützt dies nicht viel. Und drittens habe ich auch teilweise oder völlig zirkuläre Definitionen (vgl. Schläfer 2002:99) während der Arbeit gefunden, wie *Tanzfest = Tanzvergnügen = vgl. Tanzerei (1) = (ugs.) kleines Tanzfest*. Solche Erklärung ist dann überhaupt nicht anwendbar.

Auf der anderen Seite darf nicht verschwiegen bleiben, dass die Definitionen meistens für die Findung der tschechischen Übersetzung einen guten Dienst geleistet haben. Die Verwendungsbeispiele boten dabei noch zusätzliche Hilfe, damit sie aber auch im GDTAW

der deutsche Ausdruck in WP vorhanden war und nicht sein tschechischer Gegenüber.
Schriftliche Arbeit in Lexikologie – Ondřej Dušek

angewandt werden können, müssen sie noch durch den wirklichen Gebrauch bestätigt werden (siehe 4.2).

4 Die Korpusanalyse

4.1 Theoretische Vorüberlegungen

Für die Bearbeitung der Artikel für ein zweisprachiges Wörterbuch sind aber nicht nur die einsprachigen Definitionen und Übersetzung-Äquivalente nötig, sondern auch ihre Überprüfung an den wirklichen Textdaten und Erweiterung um den üblichsten Kontext. Die verwendete Sammlung der Texte muss dabei groß genug sein, um ein mindestens annähernd dem Realgebrauch entsprechendes Bild zu erzeugen, was heute die elektronischen Korpora wie DeReKo ermöglichen. Die enorme Diskrepanz zwischen der Imagination der Wörterbuchverfasser und dem wirklichen Stand der Dinge ist schon seit der Entstehung erster elektronischer Korpora bekannt (Hanks 2000) und deshalb die Anwendung des Korpus auch für die Arbeit am GDTAW prägend (Vachková 2007:13ff.). Dabei werden nicht nur die Übersetzungen überprüft, sondern auch viele Informationen über die Verwendung der Wörter im Kontext erworben. In der Regel wird die Analyse der Korpusdaten gleichzeitig mit der Suche nach den Übersetzungen durchgeführt, weil es als sehr günstig erscheint, wenn sich diese zwei Tätigkeiten gegenseitig beeinflussen und bereichern.

DeReKo, mit dem im Rahmen des GDTAW-Projekts meist gearbeitet wird, bietet eine unübertroffene Anzahl der deutschsprachigen Texte und dazu eine Menge Instrumente, die man bei der Recherche anwenden kann. Neben bloßer Suche nach einer Wortform verfügt das Software über Lemmatisierung, d.h. alle möglichen Wortformen werden bei der Analyse miteinbezogen. Das wichtigste Hilfsmittel für die Untersuchung der Verwendungsweisen der Wörter ist jedoch die Kookkurrenzanalyse⁶. Bei diesem völlig automatischen Prozess werden alle Vorkommnisse des analysierten Wortes gesammelt und dann in ihren Kontexten nach den wahrscheinlichsten Kookkurrenzpartnern gesucht, d.h. Wörtern, die sehr oft in einer festen Wortverbindung mit dem analysierten Lemma auftreten. Die Methode stützt sich auf theoretische Konzepte der Statistik – es werden die Häufigkeiten aufgezählt, wie viele Male das analysierte Wort mit dem konkreten Kandidaten zusammen auftritt (kookkurriert) und nicht; davon wird das Log-

⁶ Siehe auch die genaue Beschreibung auf <http://www.ids-mannheim.de/kl/misc/tutorial.html>.
Schriftliche Arbeit in Lexikologie – Ondřej Dušek

Likelihood-Ratio (LLR-Wert)⁷ ausgerechnet (Manning & Schütze 1999:172-175), die als Maß der Verkettung der zwei Wörter dient: die Paare mit höchstem LLR-Wert sind die (nach diesem statistischen Modell⁸) wahrscheinlichsten Kookurrenten. Wie jede Statistik ist der LLR-Wert nur für die im Korpus häufiger vertretenen Wörter zuverlässig, in der Regel bilden aber seltene Wörter keine festeren Verbindungen⁹.

4.2 Konfrontation des Wörterbuchmaterials mit den Texten im Korpus

Ich habe für alle bearbeiteten Wörter zuerst alle Belege in DeReKo aufsuchen lassen und dann, wenn die Abfrage genug Treffer¹⁰ ergab, führte ich die Kookurrenzanalyse durch. Aus der Kontrolle der Ergebnisse ermittelte ich dann die festen Kollokationen der einzelnen Bedeutungen und auch die Verwendungsweisen der Bedeutungen selbst.

- | | |
|--|---|
| 1. schäumende W-n; | 1. die W-n glätten / versuchen, die W-n zu glätten |
| 2. die W-n schlugen über ihm zusammen; | 2. die W-n glätten sich / inzwischen, mittlerweile, längst geglättet |
| 3. von den W-n hin und her geworfen werden; | 3. die W-n gehen hoch |
| 4. Ü. die W-n der Begeisterung, Erregung gingen hoch; | 4. die W-n hochgehen lassen |
| 5. Ü. sie schwammen auf den W-n des Ruhms; | 5. W-n d. Empörung |
| 6. I. die W-n glätten | 6. W-n d. Begeisterung |
| 7. I. die W-n glätten sich | 7. die W-n schlagen hoch |
| | 8. auf e-r W. / schwimmen |
| | 9. W. der Erregung |

Abbildung 1: Beispiele für das Wort "Woge" aus DUW (links) und DeReKo (rechts)

Was die Kollokationen betrifft, oft stimmen sie (mindestens teilweise) sehr gut zu den Verwendungsbeispielen im DUW. Es finden sich aber öfters viele sehr häufig verwendete Fügungen, die im DUW gar nicht erfasst werden. Wie die Abbildung 1 zeigt, gibt DUW für das Wort *Woge* viele Beispiele an, die mit dem Befund von DeReKo übereinstimmen, von den ersten drei ist aber nur der erste als Kollokation zu belegen, und zwar mit einem relativ geringerem LLR-Wert. Dafür gibt es mehrere nützliche Verbindungen, die im DUW nicht enthalten sind. Die Liste der nach dem LLR-Wert sortierten Kookurrenten von DeReKo muss jedoch immer noch manuell gesiebt werden (in der Abbildung 1 werden nur die relevanten Ergebnisse angezeigt), weil es darin relativ viel Störung (Eigennamen,

⁷ Wie schon der Name andeutet, die zwei Wahrscheinlichkeiten werden logarithmiert und dann dividiert.

⁸ Es gibt auch andere Maße für die Ermittlung der wahrscheinlich verketteten Wörter (Hanks 2008b), in DeReKo wird der LLR-Wert verwendet.

⁹ Für die seltenen Wörter lassen sich auch alle Korpusbelege sehr einfach durchgehen, was mögliche Wortverbindungen entdecken würde.

¹⁰ Die IDS-Software selbst empfiehlt mindestens 100 Treffer, damit das Ergebnis zuverlässig ist.

Duplikate usw.) gibt¹¹. Eine spezielle Quelle der Störung für einige Wörter ist die Homonymie der Formen, z.B. die deadjektivischen Substantive konkurrieren mit den groß geschriebenen Adjektiven am Satzanfang¹².

Die Korpusanalyse besagt aber nicht nur über die Wortverbindungen, sondern auch über die Bedeutungen des konkreten Wortes selbst sehr viel. Wenn ich die einzelnen Kookurrenten dem Wort zuzuordnen versuchte, zeigte sich mehrmals, dass es noch eine weitere (und nicht marginale) Bedeutung des Wortes gibt, die im DUW nicht enthalten ist, wie z.B. *Wohnstätte* = *ubytovna* (*Wohnheim, Herberge*) oder *Tascherl* = *malá taška – zavazadlo* (*kleine Tasche – Gepäck*), wo diese unbeschriebene Bedeutung im Korpus sogar überwiegt.

An den Ergebnissen der Analyse wird auch sichtbar, in welchem Sinne das Wort häufiger verwendet wird. In manchen Fällen stimmt dies auch mit der Anordnung der Bedeutungen im DUW nicht überein, wie z.B. bei *Würze*, wo die Mehrheit der Treffer dem nach DUW zweiten Sinn *koření, aroma, chut'* (*würziger, aromatischer Geschmack od. Geruch*) zufällt. Ähnliches gilt auch für die Wörter mit zwei Schreibvarianten. Es hat sich beispielsweise gezeigt, dass beim Wort *Tabaksrauch/Tabakrauch* die letztere Schreibung viel häufiger ist.

Auf der anderen Seite kann eine allzu große Überzahl der einen Bedeutung auch Probleme mit sich bringen – die Treffer der anderen Bedeutungen gehen darin leicht verloren, wie bei *Tag*, wo die letzten drei Sinne *svátek* (*Feiertag*), *dnové, život* (*geh. Leben*) und *menstruace* (*Menstruation*) nicht gefunden werden können. Man darf aber nicht sagen, dass diese gar nicht existieren. Andere solche Probleme hängen damit, dass sich in DeReKo meistens Zeitungsartikel befinden, d.h. Fachsprache und Umgangssprache sind nicht so stark vertreten.

Ein weiteres angetroffenes Phänomen ist die Unschärfe der Bedeutungen. Wie schon gesagt (siehe 3.3), begegnet man im DUW öfters einer eher zu feinen Differenzierung. Dies hat sich im Korpus bestätigt – manchmal ist in den Belegen unklar, um welchen genauen Sinn aus der Liste im DUW es sich handelt (wie bei *Tätigkeit* und *činnost* (*das*

11 Bei manchen Wörtern (wie *Wolfsgrube*) herrschen die Eigennamen sogar vor.

12 Diese Störung betrifft die Wörter selbst, nicht nur die Kookkurrenzanalyse. Sie könnte mindestens teilweise durch morphologische Markierung der Wörter beseitigt werden, DeReKo bietet aber zur Zeit diese Markierung an vollem Umfang der Daten nicht.

Sichbeschäftigen) versus *zaměstnání* (*Arbeit*)). Dieses Problem hängt aber auch mit der Unschärfe der Sprachgebilde selbst (Lakoff 2008:221) zusammen. Nach Hanks (2008a) sollen deshalb vielmehr potenzielle Bedeutungen oder Muster in den Wörterbüchern angegeben werden, womit die Ergebnisse der Korpusanalyse gut übereinstimmen.

5 Schlussbemerkungen

Bei der Bearbeitung der Lemmata habe ich also nicht nur meinen Wortschatz erweitert, sondern auch viel Interessantes über die Wörter und das Verhalten ihrer Bedeutungen im Allgemeinen erfahren.

Die Arbeit mit den Wörterbüchern hat in der Praxis sichtbar gemacht, wie problematisch die Definitionen der einzelnen Bedeutungen sein können. Die Schwierigkeiten bei der Suche nach den Übersetzungen haben auch manche Unterschiede des Wortschatzaufbaus im Tschechischen und im Deutschen klar gestellt.

An den Ergebnissen der Korpusanalyse zeigte sich, wie alle vom Verfasser selbst ausgedachten Beispiele nie die Fülle des wirklichen Sprachgebrauchs erfassen können. Auf der anderen Seite ist deutlich, dass auch die Korpusarbeit allein nicht alle Verwendungsweisen des konkreten Wortes erfassen kann, weil das Korpus, obwohl immens groß, immer begrenzt ist. Der riesige Datenumfang verbirgt auch viele Randerscheinungen, die aber doch beschrieben werden sollen. Deshalb ist es am günstigsten, beide Quellen bei der Arbeit mit den Lemmata auszunutzen.

Quellenverzeichnis

- Čermák, František. 1995. Překladová lexikografie. In František Čermák & Renata Blatná (eds.), *Manuál lexikografie*, 230-248. 1st ed. Jinočany: H&H.
- Fleischer, Wolfgang & Irmhild Barz. 1992. *Wortbildung der deutschen Gegenwartssprache*. 3rd ed. Tübingen: M. Niemeyer.
- Hanks, Patrick. 2000. Immediate Context Analysis: Distinguishing Meanings By Studying Usage. In Chris Heffer & Helen Stauntson (eds.), *Words in Context: A Tribute to John Sinclair on his Retirement*. Birmingham: University of Birmingham.
- Hanks, Patrick. 2008a. Linguistic Norms and Pragmatic Exploitations. In Patrick Hanks (ed.), *Lexicology*, vol. 5, 233-255. 1st ed. London; New York: Routledge.
- Hanks, Patrick. 2008b. Word Association Norms, Mutual Information and Lexicography. In Patrick Hanks (ed.), *Lexicology*, vol. 6, 150-167. 1st ed. London; New York: Routledge.
- Harras, Gisela, Ulrike Haß & Gerhard Strauß. 1991. *Wortbedeutungen und ihre Darstellung im Wörterbuch*. Berlin; New York: W. de Gruyter.
- Lakoff, George. 2008. Hedges and meaning Criteria. In Patrick Hanks (ed.), *Lexicology*, vol. 5, 221-232. 1st ed. London; New York: Routledge.
- Manning, Christopher D. & Hinrich Schütze. 1999. *Foundations of statistical natural language processing*. Cambridge Massachusetts: MIT Press.
- Schippan, Thea. 2002. *Lexikologie der deutschen Gegenwartssprache*. 2nd ed. Tübingen: Niemeyer.
- Schlaefel, Michael. 2002. *Lexikologie und Lexikographie : eine Einführung am Beispiel deutscher Wörterbücher*. Berlin: Erich Schmidt.
- Vachková, Marie. 2007. *Kapitoly k německo-české metalexikografii*. Praha: Univerzita Karlova.
- Zelená, Lucie. 2008. Zu Bearbeitung der Substantive im GDTAW. In Marie Vachková (ed.), *Beiträge zur bilingualen Lexikographie*, 221-228. Praha: Univerzita Karlova.

Verwendete Lexika, Korpora und andere Internetquellen

- CAN - *Canoo – Deutsche Wörterbücher und Grammatik* (<http://www.canoo.net>).
- CC - *Dict.cc – English-German Dictionary* (<http://www.dict.cc/>).
- ČNK - *Český národní korpus* (<http://www.korpus.cz/>).
- DeReKo - *Deutsches Referenz-Korpus* (<http://www.ids-mannheim.de/cosmas2/>).
- DUW - *Duden Universalwörterbuch*. Version 3.0, 2001.
- DWDS - *Digitales Wörterbuch der deutschen Sprache* (<http://www.dwds.de>).
- DWP - *Deutscher Wortschatz-Portal* (<http://wortschatz.uni-leipzig.de/>).
- GOO - *Google Suche* (<http://www.google.com/>).
- JP - *Internetová jazyková příručka* (<http://prirucka.ujc.cas.cz/>).
- LL - *Lingea Lexicon – Deutsch-Tschechisches Großwörterbuch*. Version 4.0, 2002.
- LEO - *Leo Deutsch-Englisches Wörterbuch* (<http://dict.leo.org/>).
- SIEB - Siebenschein, H. et al. 1970. *Německo-český slovník*. Praha: SPN.
- SS - *Seznam slovník* (<http://slovník.seznam.cz>).
- WP - *Wikipedia* (<http://de.wikipedia.org>, <http://cs.wikipedia.org>).

Liste der bearbeiteten Lemmata

Wilde, Wildkatze, Wildschütz, Williams Christbirne, Windbö, Windböe, Winde, Windlade, Windrose, Wippe, Wirkungsstätte, Wirre, Wirtspflanze, Wirtsstube, Woge, Wohndichte, Wohnfläche, Wohnküche, Wohnsitzlose, Wohnstätte, Wohnstube, Wöhrde, Woiwod, Woiwode, Wolfsgrube, Wolfsspinne, Wolke, Wollmütze, World Wide Web, Wortfamilie, Wortfolge, Wortgruppe, Worthülse, Wortliste, Wruke, Wuchtbrumme, Wunde, Wundrose, Wunschliste, Würde, Wurstbrühe, Wurstküche, Wurstsorte, Wurstsuppe, Würze, Würzsoße, Wüste, x-Achse, x-fache, Xanthippe, Xylophage, Xylose, y-Achse, z-Achse, Zahnbürste, Zahnfäule, Zahnpastatube, Zahnspange, Zahnstange, Zähre, Zaire, Zange, Zapfsäule, Zarge, Zicke, Zickzacklinie, Zigarre, Zikade, Ziliate, Zilie, Zimtstange, Zimtzicke, Zimtziege, Zirkularnote, Zitze, Zivikourage, Zivilisationsstufe, Zivilsache, Zmorge, Zofe, Zosse, Zote, Zuchtbulle, Zuchthausstrafe, Zuchtrüde, Zuchtstätte, Zufahrtsrampe, Zufallsvariable, Zufluchtsstätte, Zugfolge, Zugstange, Zuhause, Zukünftige, Zündanlage, Zündfunke, Zündfunken, Zündhilfe, Zündspule, Zusage, Zusatzfrage, Zusprache, Zustandsgröße, Zuwachsrage, Zwangspause, Zwanzigfache, Zwanzigjährige, Zwicke, Zwinge, Zwölffache, Zwölfjährige, Zygote, Zymologie, Zyste, Zystoskopie, Zytode, Zytologie, T-Eisen, Tabakschnupfen, Tabaksrauch, Tabasco(r), Tabellenkalkulation, Tabes, Tablettenabhängige, Tablettenmissbrauch, Tablettensüchtige, Tabor, Tabula gratulatoria, Tabuthema, Tabuverletzung, Tachistoskop, Tacho, Tachograph, Tachonadel, Tachostand, Taekwondo, Tafellappen, Taferlklasse, Taftbindung, Tag¹, Tag², Tag³, Tai-Chi, Tai-Chi-Chuan, Takel, Takelage, Takelung, Takelwerk, Talentförderung, Talgkerze, Talisch, Talk², Talkerde, Talkshow, Tamagotchi, Tamil, Tangahörschen, Tangenssatz, Tangentenviereck, Tangram, Tanklastzug, Tankzug, Tannensterben, Tantra, Tanzfest, Tanzpaar, Tapergreis, Tapezierbürste, Tapezierung, Tapp¹, Tapsigkeit, Tarantella, Tarifabschluss, Tarifangestellte, Tarifaueinandersetzung, Tarifbezirk, Tarifgebiet, Tarifstreitigkeit, Tarifvereinbarung, Tartufo, Taschenfeitel, Taschengeldentzug, Taschenlexikon, Taschenl, Täschneri, Tastversuch, Tatami, Tatbeweis, Tätigkeit, Tätigkeitsbereich, Tätigkeitsbericht, Tätigkeitsbeschreibung, Tätigkeitsgebiet, Tätigkeitsprofil, Tätigkeit, Tatkraft, Tatmensch, Tatmotiv, Tätowierung, Tatsachenbericht, Tatterich, Tatterigkeit, Tatütata, Tatverdacht, Tatverdächtige, Tätzchen, Tätzlein, Tau¹, Tau², Tau³, Taubblindheit, Täubchen, Taube¹, Taube², Taubendreck, Taubenei

Beispiele der bearbeiteten Lemmata

Tabasco (r)

- 1) der
- 1a) die
- 2) Tabasco
- 2a) Tabascosoße
- 2b) Ta|b#as|co
- 2c) Ta|b#as|co|so|ße
- 3) -s, (jen sg.)
- 3a) -, (jen sg.)
- 6) subst.
- 7) tabasko
- 7) tabasco
- 7) tabasco omáčka
- 11) Nach Wunsch kann man auch einige Tropfen T. einrühren.
- 11b) Podle přání je možné přimíchat i několik kapek tabasco omáčky.
- 21) i varianta „Tabascosauce“
- 23) Ondřej Dušek

Talk

- 1) der
- 2) Talk
- 2b) T_alk
- 3) -s, -s
- 5) (žargon)
- 6) subst.
- 7) debata
- 7a) neutrální
- 7) беседа
- 7a) neutrální
- 7) rozprava
- 7a) neutrální
- 7) pokec
- 7a) (hovor.)
- 11) Er begann als Erster den täglichen T. im deutschen TV.
- 11b) Jako první začal s každodenními rozpravami v německé televizi.
- 16) die Plauderei
- 16) die Unterhaltung
- 16) das Gespräch
- 23) Ondřej Dušek

Tatütata

- 1) das
- 2) Tatütata
- 2b) Ta|t_ü|ta|t_a
- 3) -s, -s
- 5) (hovor.)
- 6) subst.

- 7) houkání
- 7a) neutrální
- 7) zvuk sirény
- 7a) neutrální
- 10a) např. hasičského vozu
- 11) Mit Blaulicht und T. raste der erste Löschzug zur regionalen Schule.
- 11b) S majáčkem a houkáním se k oblastní škole řítí první hasící četa.
- 23) Ondřej Dušek

Wippe

- 1) die
- 2) Wippe
- 2b) W#ip|pe
- 3) -, -n
- 6) subst.
- 7) houpačka
- 10a) zesponu podepřená, ne závěsná
- 7) vahadlová houpačka ?
- 11) auf der W. sitzen/schaukeln
- 11b) sedět/houpat se na houpačce
- 16) die Wippschaukel
- 18) wippen
- 21) EXPERT – vahadlová / kyvadlová houpačka ?
- 23) Ondřej Dušek

Wohnstätte

- 1) die
- 2) Wohnstätte
- 2b) W_ohn|stät|te
- 3) -, -n
- 6) subst.
- 7) A místo bydliště
- 7) A bydliště
- 7a) neutrální
- 9) vzeš.
- 11) W. von Mozart/Beethoven
- 11b) místo bydliště Mozarta/Beethovena
- 7) B ubytovna
- 7) B domov
- 11) W. für Behinderte
- 11b) domov pro postižené
- 14) dva
- 16) A Wohnort
- 16) A Wohnplatz
- 16) A Wohnsitz
- 16) B Wohnheim
- 21) význam B DUW vůbec neuvádí, ale v korpusu se vyskytuje
- 23) Ondřej Dušek

Wortfolge

- 1) die
- 2) Wortfolge
- 2b) W#ort|fol|ge
- 3) -, -n
- 6) subst.
- 7) **A** slovosled
- 7) **A** pořádek slov
- 9) (jaz.)
- 11) Es werden bestimmte standardisierte W-n wie Subjekt, Prädikat und Objekt eingehalten.
- 11b) Dodržují se určité standardizované pořádky slov jako podmět – přísudek – objekt.
- 7) **B** skupina slov
- 7) **B** slovní spojení ?
- 11) Ein Zungenbrecher ist eine W., deren schnelle Aussprache schwierig ist.
- 11b) Jazykolam je skupina slov, kterou je obtížné vyslovit.
- 11) In §14 entfällt die W. “den Generalsekretär”
- 11b) V §14 se vypouští slovní spojení “generálního sekretáře”
- 14) dva
- 21) EXPERT, DUW uvádí jen “Folge von Wörtern”, což se dá interpretovat oběma způsoby
- 23) Ondřej Dušek

Würze

- 1) die
- 2) Würze
- 2b) W#ür|ze
- 3) -, -n

- 6) subst.
- 7) **A** koření
- 7) **B** aroma
- 7) **B** koření
- 7) **B** chuť
- 11) eine besondere/die richtige W. (dem Gericht) verleihen/geben
- 11b) dodat (jídlu) správné aroma
- 11) j-m/etw. fehlt die W.
- 11a) přen.
- 11b) někomu/něčemu chybí šmrnc
- 11) die W. des Lebens
- 11a) přen.
- 11b) koření života
- 13) In der Kürze liegt die W.
- 13a) přen.
- 13c) Krátkost je duší vtipu. ?
- 13c) V jednoduchosti je krása. ?
- 7) **C** sladina
- 9) odbor. jaz.
- 14) tři
- 16) **A** das Gewürz
- 16) **A** das Würzmittel
- 16) **A** die Würzung
- 16) **B** das Aroma
- 16) **C** die Bierwürze
- 18) das Gewürz, würzen
- 20) das Würzmittel
- 20a) die Bierwürze
- 21) EXPERT pro C, naprostá většina dokladů je k významu B; často v přen. významu
- 23) Ondřej Dušek