



Pedagogická  
fakulta  
Faculty  
of Education

Jihočeská univerzita  
v Českých Budějovicích  
University of South Bohemia  
in České Budějovice

Jihočeská univerzita v Českých Budějovicích

Pedagogická fakulta

Katedra germanistiky

Bakalářská práce

Genusvariation deutscher Substantive.  
Eine korpusgestützte Fallstudie am  
Beispiel von *Python*, *Samba*, *Rodel* und  
*Paprika*.

The grammatical gender variants of German substantives. A case study using the examples of *Python*, *Samba*, *Rodel* and *Paprika* corroborated by corpus.

Vypracovala: Lucie Kulichová

Vedoucí práce: Mgr. Jana Kusová, Ph.D.

České Budějovice 2014

## **Prohlášení**

Prohlašuji, že svoji bakalářskou práci jsem vypracovala samostatně pouze s použitím pramenů a literatury uvedených v seznamu citované literatury.

Prohlašuji, že v souladu s § 47b zákona č. 111/1998 Sb. v platném znění souhlasím se zveřejněním své bakalářské práce, a to v nezkrácené podobě pedagogickou fakultou elektronickou cestou ve veřejně přístupné části databáze STAG provozované Jihočeskou univerzitou v Českých Budějovicích na jejích internetových stránkách, a to se zachováním mého autorského práva k odevzdanému textu této kvalifikační práce. Souhlasím dále s tím, aby toutéž elektronickou cestou byly v souladu s uvedeným ustanovením zákona č.111/1998 Sb. zveřejněny posudky školitele a oponentů práce i záznam o průběhu a výsledku obhajoby kvalifikační práce. Rovněž souhlasím s porovnáním textu mé kvalifikační práce s databází kvalifikačních prací Theses.cz provozovanou Národním registrem vysokoškolských kvalifikačních prací a systémem na odhalování plagiátů.

V Českých Budějovicích dne 18. 8. 2014

.....

Lucie Kulichová

## **Poděkování**

Na úvod své práce bych chtěla velice poděkovat paní Mgr. Janě Kusové Ph.D. za odborné vedení mé práce, za cenné rady a připomínky a v neposlední řadě za její čas, trpělivost a ochotu, kterou mi vždy věnovala. Velice si vážím jejich rad, připomínek a korektur, díky nimž mohla být tato práce napsána.

## **Anotace**

**Rodová variace německých substantiv. Případová korpusová studie provedena na příkladu slov *Python*, *Samba*, *Rodel* a *Paprika*.**

Tato bakalářská práce se zabývá nejnovějšími tendencemi ve vývoji morfologického systému německých substantiv. Na příkladu čtyř substantiv *Python*, *Samba*, *Rodel* a *Paprika* bude provedena analýza kolísání rodu těchto substantiv, přičemž bude pozornost soustředěna především na sémantickou a diatopickou distribuci tohoto kolísání. Pro práci byl využit korpus *DeReKo* a jeho nástroje (COSMAS II).

V teoretické části práce bude nastíněn přínos metod korpusové lingvistiky pro výzkum současného jazyka. V praktické části pak bude provedena korpusová analýza výše uvedených substantiv.

## **Annotation**

**The grammatical gender variants of German substantives. A case study using the examples of *Python*, *Samba*, *Rodel* and *Paprika* corroborated by corpus.**

This bachelor's thesis deals with the newest trends in developments of morphological system of German nouns. The analysis of variation in gender of the nouns will be accomplished in four examples of nouns *Python*, *Samba*, *Rodel* and *Paprika*. Above all, the thesis will be focused on semantic and regional distribution of this variation. I have used the corpus *DeReKo* and its tools (COSMAS II) during writing the thesis.

In the theoretical part I further outline the benefits of methods in corpus linguistics for the research of contemporary language. In the practical part I made a corpus analysis of the nouns mentioned above.

## **Inhalt**

1. Einleitung.....	7
2. Korpuslinguistik und Korpus – Eine Einführung .....	9
2.1. Korpuslinguistik.....	9
2.2. Historischer Überblick .....	10
2.3. Was ist ein Korpus? .....	12
2.4. Arten von Korpora .....	14
2.5. Institut für Deutsche Sprache (IDS) in Mannheim und Cosmas II .....	19
2.5.1. DeReKo .....	20
3. Substantive und ihre Charakteristika. Kontrast zwischen dem Deutschen und Tschechischen .....	22
4. Substantive und ihre Genera .....	23
4.1. Biologisches Genus .....	23
4.2. Grammatisches Genus.....	24
4.2.1. Genus bestimmende Suffixe.....	24
4.2.2. Genus nach semantischen Gruppen.....	25
4.2.3. Substantive mit zwei Genera .....	26
5. Korpusanalyse.....	27
5.1. Paprika (die, der) .....	27
5.2. Samba (die, der) .....	32
5.3. Python (die, der).....	36
5.4. Rodel (die, der).....	40
6. Schluss .....	45
7. Resumé.....	48
8. Literaturverzeichnis .....	50

# 1. Einleitung

Die vorliegende Bachelorarbeit beschäftigt sich mit dem Thema „*Genusvariation deutscher Substantive*“. Es handelt sich um eine Fallstudie, die aufgrund einer korpusgestützten Analyse der ausgewählten Wörter (*Python*, *Samba*, *Rodel* und *Paprika*) Beispiele der Genusvariation erforscht.

Im ersten Teil der Bachelorarbeit wurde die Aufmerksamkeit der Korpuslinguistik gewidmet. Die Korpuslinguistik ist eine moderne sprachwissenschaftliche Disziplin, die sich mit der Untersuchung der menschlichen Sprache beschäftigt. Im Vergleich zu den anderen linguistischen Disziplinen handelt es sich um ein ziemlich junges Fachgebiet der Linguistik, das sich in den 60er-Jahren des 20. Jahrhunderts entwickelte. Die moderne Informationschtechnologie, wie auch die sich schnell entwickelnden modernen Fortschritte, hängen mit der Entstehung der Korpuslinguistik zusammen. Das Ziel des theoretischen Teils dieser Bachelorarbeit ist, verschiedene Fragen, die im Zusammenhang mit der Korpuslinguistik stehen, zu beantworten. Was kann man sich unter dem Begriff „*Korpus*“ vorstellen? Welche Typen von Korpora stehen den Forschern zur Verfügung? Wie kann man mit solchen Korpora arbeiten?

Als nächstes wird das Institut für deutsche Sprache in Mannheim vorgestellt. Im Rahmen des Instituts wird das System COSMAS II. entwickelt, mit dem im praktischen Teil gearbeitet wird. Der nächste Teil der Arbeit stellt die Charakteristik der Wortart *Substantiv* dar, wie auch die grammatische Kategorie *Genus*. Hier werden die Unterschiede zwischen dem deutschen und tschechischen Genussystem vorgestellt und erklärt.

Anschließend kommt der praktische Teil. Die Korpusanalyse wurde mit Hilfe des Referenzkorpus durchgeführt. Aufgrund der Korpusrecherche werden vier ausgewählte Substantive, bei denen das Genus schwankt, analysiert. Bei der Auswertung wird die diatopische und semantische Ebene berücksichtigt.



## **2. Korpuslinguistik und Korpus – Eine Einführung**

### **2.1. Korpuslinguistik**

Die Korpuslinguistik ist eine moderne Disziplin, die sich mit der Untersuchung der menschlichen Sprache beschäftigt. Ein wichtiges Ziel liegt darin die Sprache und den Sprachgebrauch zu analysieren und beschreiben (Scherer, 2006, S. 1-2).

Die Disziplin entwickelt sich anhand von der großen technologischen Entwicklung und hängt mit dem Aufschwung von Computern zusammen. Deshalb gehört die Korpuslinguistik zu den jungen und sich schnell entwickelnden Disziplinen (Šulc, 1999, S. 10).

Die Korpuslinguistik findet auch in anderen Disziplinen Anwendung, die sich mit der Sprache beschäftigen und kooperiert mit vielen Hilfsdisziplinen (z.B. im Bereich der Stilistik oder der Literaturwissenschaft). Für die Disziplin ist noch die Zusammenarbeit mit vielen anderen Bereichen charakteristisch, wie z.B. die Zusammenarbeit mit der Textlinguistik, Computerlinguistik oder mathematischen Linguistik (Lenz, 2000, S. 2).

In der Sprachwissenschaft oder Linguistik geht es darum, den Sprachgebrauch und die Sprache zu untersuchen. Die Sprachdaten werden empirisch untersucht und Scherer (2006, S. 2) stellt insgesamt drei Methoden dar, die der Korpuslinguistik zur Verfügung stehen. Die drei Methoden sind 1. Experiment, 2. Befragung und 3. Arbeit mit einem Korpus. Bei den Methoden Experiment und Befragung kommen aber Nachteile vor. Der erste Nachteil ist, dass eine große Menge von sprachlichen Daten gesammelt und bewertet wird, was sehr zeitaufwendig sein kann. Der zweite Nachteil ist, *„dass die Versuchspersonen vielleicht nicht die Antworten geben werden, die ihrem spontanem Sprachgebrauch entsprechen“*. Mit der dritten Methode können wir diese Nachteile vermeiden (Scherer, 2006, S. 2-3).

Lemnitzer/Zinsmeister (2010, S. 6-7) führen zwei „*Forschetypen*“ an: der erste Typ ist der Denker, der sich an die Sprachkompetenz orientiert und der zweite Typ ist der Beobachter, der empirisch arbeitet.

Der Beobachter sammelt authentische Sprachdaten und hat Interesse daran möglichst viele Daten zu erforschen. „*Je mehr Daten, desto besser*“ (Lemnitzer/Zinsmeister, 2010, S. 6). Die Hypothesen und Aussagen des Beobachters werden durch die Daten entweder bestätigt oder widerlegt. Dieser Forschertyp interessiert sich vor allem für die alltäglichen Phänomene der Sprache.

Im Gegensatz dazu sammelt der Denker sammelt jahrelang alle wichtigen Daten in seinem Kopf und beschäftigt sich nicht mit den sprachlichen Phänomenen, die im Alltag vorkommen (Lemnitzer/Zinsmeister, 2010, S. 6-7).

Lemnitzer/Zinsmeister (2010, S. 7) kamen zum folgenden Schluss: „*eine enge Zusammenarbeit zwischen Denkern und Beobachtern, also zwischen theoretischen Linguisten und empirisch arbeitenden Linguisten, erscheint uns fruchtbar für beide Seiten*“ (Lemnitzer/Zinsmeister, 2010, S. 7).

## **2.2. Historischer Überblick**

Die Korpuslinguistik, wie man sie heute kennt, ist eine ziemlich junge und moderne Disziplin. Ihre Wurzeln sind in den sechziger Jahren des 20. Jahrhunderts zu finden. Ihre Entstehung ist mit dem schnellen technischen Fortschritt verbunden und zwar mit dem Gebrauch von Computertechnologien (Lenz, 2000, S. 6).

Das bedeutet aber nicht, dass sich niemand früher mit dem Sprachgebrauch befasst hätte.

In der Vergangenheit hat sich schon eine Generation der Linguisten für die Sprache und den authentischen Sprachgebrauch interessiert. Die Linguisten haben sich mit der Sprache beschäftigt, wobei sie Kartei (ein früherer Begriff für die Textsammlung) verwendeten. Mit Karteien haben im 17. Jahrhundert die Lexikographen gearbeitet (Šulc, 1999, S. 28).

Henry Kučera a Nelson Francis haben im Jahr 1964 das erste elektronische Korpus zusammengestellt. „*Das Brown University Standard Corpus of Present-Day Edited American English*“ (kurz *Brown Corpus*) enthält ungefähr eine Million Wörtern. Das Brown Corpus hatte eine repräsentative Funktion, und zwar für das amerikanische Englisch und galt als wichtige Vorlage für andere Korpora (Šulc, 1999, S. 30).

Lenz (2000, S. 6) führt an, dass die Möglichkeit, eine große Menge von Texten systematisch analysieren zu können, mit der Verwendung von Computern in Zusammenhang steht. Aufgrund dieser Tatsache wurden viele Korpora gebildet (z.B. „*das LIMAS-Korpus*“ oder „*die Mannheimer Korpora*“) (Lenz, 2000, S. 6).

Eine wichtige Rolle spielt in der Korpuslinguistik die Entwicklung des Internets. Dank des Internets konnte sich Korpuslinguistik schnell durchsetzen. Das World Wide Web bietet den Linguisten eine reiche Datenmenge für ihre Untersuchungen. Trotzdem stoßen wir bei der Verwendung des Internets auf Schwierigkeiten, und zwar ist es bei einigen Texten ist schwer Daten über „*die Herkunft, den Entstehungszeitpunkt oder die Autorschaft*“ herauszufinden (Lemnitzer/Zinsmeister, 2010, S. 43).

### 2.3. Was ist ein Korpus?

Das Wort „*Korpus*“ (*corpus, -oris*, Körper) stammt ursprünglich aus dem Lateinischen aber für unsere Gesellschaft kommt das Wort aus dem englischen Wort „*corpus*“ (Šulc, 1999, S. 9).

Was versteht man unter dem Begriff *Korpus*?

Unter dem Begriff Korpus versteht man in der Linguistik eine Sammlung von allen schriftlichen Texten aber auch von mündlichen Äußerungen, die in einer elektronischen Form vorkommen. Zu den Produkten der geschriebenen Sprache gehören Zeitungsartikel, E-Mails oder Geschäftsbriefe. Zu den mündlichen Äußerungen gehören Telefongespräche oder ein Gespräch beim Mittagessen. Solche Texte, die in einem Korpus vorkommen, werden „*Primärdaten*“ genannt (Scherer, 2006, S. 3).

Am Anfang der Arbeit mit einem Korpus muss man feststellen, welche Seite der Sprache man analysieren und untersuchen will. Ein Korpus, das sich nur mit einer Varietät beschäftigt, z.B. der Jugendsprache oder der Sprache eines Gebiets, würde ganz anders aussehen als Korpus, das sich ganz allgemein mit der Sprache befasst (Scherer, 2006, S. 4). Die zusätzlichen Auskünfte über den Text wie auch die Auskünfte über den Autor werden als „*Metadaten*“ bezeichnet (Scherer, 2006, S. 9). Die Metadaten dokumentieren den Inhalt eines Korpus und kommentieren die Primärdaten (Lemnitzer/Zinsmeister, 2010, S. 190).

Nach Scherer (2006, S. 10) sind die „*wichtigsten linguistischen Anwendungsgebiete von Korpora*“:

- Erforschung von sprachlichen Strukturen und Varietäten
- Erstellung von Wörterbüchern (Lexikografie)
- Erstellung von Grammatiken (Grammatikografie)

- Fremdsprachunterricht
- Übersetzung
- Computerlinguistik

Bei der Sammlung sollten wichtige sprachwissenschaftliche Merkmale erfüllt werden. Es geht um „*Größe und Inhalt des Korpus, Beständigkeit und Repräsentativität*“ (Scherer, 2006, S. 5).

Scherer (2006, S. 5) behauptet, dass „*die Repräsentativität*“ beim Aufbau eines Korpus die wichtigste Rolle spielt. Damit ein Korpus als repräsentativ bezeichnet werden kann, muss man die Art und Anzahl der Texte in Erwägung ziehen. In diesem Fall ist sehr wichtig zu überlegen, „*welche Arten von Texten in welchen Anteilen in das Korpus aufgenommen werden*“ (Scherer, 2006, S. 6). Die Texte, die in einem 'Korpus enthalten sind, nennt man *Textproben* (Scherer, 2006, S. 7).

Das Kriterium „*Beständigkeit*“ eines Korpus führt an, dass der Inhalt nach der Zusammensetzung unveränderlich ist (Scherer, 2006, S. 6). Dieses Kriterium betrifft nicht die Monitorkorpora, weil sich die Zusammensetzung der Monitorkorpora im Laufe der Zeit verändert (Scherer, 2006, S. 21).

Der nächste wichtige Punkt ist die Größe des Korpus. Hier kommen wir zu einer Frage. Wie viele Texte muss ein Korpus wirklich enthalten? Am Anfang bewegte sich die Standardgröße des Korpus zwischen etwa hundert Tausend und zwei Millionen Textwörtern bis zu zwei Millionen Textwörtern. „*Keines von diesen kleinen Korpora würde heute vor den Linguisten bestehen*“ (Šulc, 1999, S. 14).

Die Korpusgröße hat sich mit dem technischen Fortschritt verändert und die Anzahl von Textwörtern hat sich stark erhöht. Die Korpora des

Institut für Deutsche Sprache (kurz IDS) in Mannheim enthalten über zwei Milliarden Textwörter (Scherer, 2006, S. 7).

## **2.4. Arten von Korpora**

Bei den Korpora werden verschiedene Merkmale und Kriterien klassifiziert. Zu den Merkmalen gehören der Verwendungszweck, die Größe und die Zugänglichkeit. Was die Größe angeht, unterscheidet Scherer (2006, S. 16) vier Korpusgrößen und zwar die sehr großen Korpora, wo sich die Anzahl von Wörtern über eine Milliarde befindet, die großen Korpora, die über hundert Millionen Textwörter enthalten, weiter die mittleren Korpora mit mehr als einer Million Textwörter und die kleineren Korpora, die bis zu einer Million Textwörter enthalten.

Bei der Klassifizierung der Zugänglichkeit unterscheidet man folgende Typen „*online frei zugängliche Korpora*“, also Korpora die man nach vorheriger Registrierung nutzen kann. Dagegen stehen die Korpora, „*wo eine externe Nutzung aus verschiedenen Gründen nicht möglich ist*“ (Scherer, 2006, S. 16).

Scherer (2006, S. 17) stellt eine Übersicht verschiedener Arten von Korpora in einer Tabelle dar:

<b>Speichermedium</b>	<b>computerlesbares Korpus</b>	<b>nicht computerlesbares Korpus</b>
Hierarchie	Gesamtkorpus	Teilkorpus
Vollständigkeit der Texte	Volltextkorpus	Probenkorpus
Abgeschlossenheit	statisches Korpus	Monitorkorpus
Aufbereitung des Korpus	annotiertes Korpus	nicht annotiertes Korpus
Sprachmedium	Korpus der geschriebenen Sprache	Korpus der gesprochenen Sprache
zeitlicher Bezug	Korpus der Gegenwartssprache	historisches Korpus
Geltungsbereich des Korpus	Referenzkorpus	Spezialkorpus
Anzahl Sprachen	einsprachiges Korpus	mehrsprachiges Korpus

Abbildung 1: Arten von Korpora

(übernommen von Scherer, 2006, S. 17)

Einerseits bietet die Tabelle eine gute Übersicht über die verschiedenen Typen der Korpora. Andererseits hilft die Tabelle bei der richtigen Auswahl eines Korpus. Der Benutzer muss in Erwägung ziehen, welche Daten er untersuchen will. Erst nach dieser Entscheidung beginnt er zu recherchieren.

Nach dem Kriterium Speichermedium unterscheidet Scherer (2006, S. 17-18) „*computerlesbares Korpus*“ und „*nicht computerlesbares Korpus*.“ „*Computerlesbare Korpora werden auch elektronische Korpora oder Computerkorpora genannt*“ (Scherer, 2006, S. 17).

Unter den nicht computerlesbaren Korpora versteht Scherer (2006, S. 17) „*Belegsammlungen*“ und „*Textarchive*“ aber auch Papierkorpora (z. B. das Mainzer Zeitungskorpus).

„*Ein Gesamtkorpus*“ besteht aus mehreren Teilkorpora. Die Teilkorpora beziehen sich auf verschiedene historische Epochen, bestimmte Sprechereigenschaften (Texte von Frauen oder Männern), auf verschiedene Textsorten (Zeitungstexte oder Wissenschaftstexte) oder auf Medium, hier handelt sich um gesprochene oder geschriebene Teilkorpora. Alle diese Teilkorpora bilden zusammen ein Gesamtkorpus (Scherer, 2006, S. 18-19).

Scherer (2006, S. 19) legt als Beispiel „*das Mannheimer Wendekorpus*“ vor, „*das sich aus den Teilkorpora ‚Wendekorpus Ost‘, ‚Wendekorpus West‘ und dem Teilkorpus zur Wiedervereinigung zusammensetzt.*“

Wie schon die Bezeichnung verrät, werden die Volltextkorpora nur solche Text beinhalten, die in ihrer gesamten Länge vorkommen. Als Beispiel führt Scherer (2006, S. 19) das digitale Wörterbuch der deutschen Sprache (DWDS-Korpus) an. Die Probenkorpora bestehen aus Texten die eine genormte Größe beinhalten. (z.B. das LIMAS-Korpus) (Scherer, 2006, S. 19).

Weiterhin kann man „*statisches Korpus*“ und „*Monitorkorpus*“ unterscheiden. Der Hauptunterschied zwischen den beiden Korpora liegt in ihrer Zusammensetzung. Bei einem Monitorkorpus verändert sich die Zusammensetzung im Laufe der Zeit, weil das Korpus immer aktualisiert wird (z.B. das Mannheimer-Morgen-Korpus des IDS) (Scherer, 2006, S. 20-21). Im Falle eines statischen Korpus bleibt die Zusammensetzung konstant. Als Beispiel führt Scherer (2006, S. 20) das Mannheimer Wendekorpus.



Scherer (2006, S. 21) unterscheidet nicht nur Korpora die Primärdaten enthalten, sondern auch Korpora, die sog. „*Annotation*“ enthalten. Solche Korpora werden „*annotierte Korpora*“ genannt. Unter dem Begriff „*Annotation*“ kann man sich verschiedene zusätzliche Informationen zum Text vorstellen, wie z.B. grammatische Informationen oder Informationen zur Flexion. Hier liegt der Hauptunterschied zwischen „*annotierten und nicht annotierten Korpora*“.

Korpora, die auf Satzebene annotiert werden, nennt Scherer (2006, S. 22) „*Baumbanken*“.

Nach dem Kriterium „*das Sprachmedium*“ unterscheidet Scherer (2006, S. 23) Texte „*der gesprochenen oder der geschriebenen Sprache*“. Diese Korpora werden als „*Korpora der geschriebenen Sprache*“ und „*Korpora der gesprochenen Sprache*“ bezeichnet. Bei den „*Korpora der gesprochenen Sprache*“ arbeitet man mit den authentischen Materialien dazu gehören Gespräche, Telefonanrufe oder Reden. Die Arbeit mit diesem Korpus ist sehr zeitaufwendig, weil alle Texte die schriftliche Form gewinnen müssen. Deswegen ist die Erstellung des Korpus der geschriebenen Sprache leichter (Scherer, 2006, S. 23).

„*Das britische Nationalkorpus*“ (BNC) enthält die Texte der gesprochenen Sprache, es handelt sich um 10 % des Gesamtumfangs aber auch die Texte der geschriebenen Sprache sind enthalten, hier ist die Anzahl höher es handelt sich um 90 % des Gesamtumfangs. Als weiteres Beispiel nennt Scherer (2006, S. 24) „*das Archiv für Gesprochenes Deutsch des IDS*“ und „*das Bayerische Archiv für Sprachsignale*.“

Zur Verfügung stehen auch Korpora, die sich mit der Kindersprache beschäftigen. Scherer (2006, S. 24) nennt als Beispiel „*das Child Language Data Exchange System*.“

Die Einteilung der Korpora in historische und gegenwartssprachige ist sehr subjektiv. Am Anfang muss man feststellen, ob die Texte eher zu der historischen Sprache oder zu der Gegenwartssprache gehören. Die Grenze zwischen den beiden Korpora ist fließend und nirgendwo festgelegt (Scherer, 2006, S. 25).

Als historisch kann man ganz klar ein Korpus bezeichnen, mit dem das IDS arbeitet. Es wird versucht, den Zeitraum von 1700 bis 1945 abzudecken. Texte des Mittelalters und die althochdeutschen Texte befinden sich in „*der Bibliotheca Augusta, beim Internetportal Mediaevum und in der Datenbank Thesaurus Indogermanischer Text- und Sprachmaterialien (TITUS)*“ (Scherer, 2006, S. 27).

„*Ein Referenzkorpus*“ repräsentiert „*eine Sprache in ihrer Gesamtheit*“, weil er verschiedene Varietäten umfasst. Als gutes Beispiel erwähnt Scherer (2006, S. 27) „*das deutsche Referenzkorpus am IDS in Mannheim*“, weiter nur DEREKO. Die Spezialkorpora orientieren sich nach einer bestimmten Varietät der Sprache (z.B. die Jugendsprache oder die Zeitungssprache) und sind für diesen konkreten Teilbereich repräsentativ. (Scherer, 2006, S. 28)

Je nach Anzahl der enthaltenen Sprachen unterscheidet Scherer (2006, S. 29) „*einsprachige Korpora*“ und „*mehrsprachige Korpora*.“ Die mehrsprachigen Korpora kann man in „*Parallelkorpora*“ und in „*vergleichbare Korpora*“ unterteilen.

Die Parallelkorpora umfassen die Originaltexte in einer Sprache, aber auch die Übersetzung in eine oder mehrere Sprachen. Die vergleichbaren Korpora können in mehrere Teilkorpora gegliedert werden und sind nach denselben Prinzipien aufgebaut. „*Vergleichbare Korpora*“ können sowohl einsprachig als auch mehrsprachig sein.

Als Beispiel für ein mehrsprachiges vergleichbares Korpus erwähnt Scherer (2006, S. 30) „*das PAROLE-Korpus*“ und als Beispiel für ein

einsprachiges vergleichbares Korpus „*das Brown-Korpus, das Lancaster-Oslo/Bergen-Korpus (LOB) und das Kolhapur-Korpus.*“ Die Struktur ist bei allen drei Korpora gleich. Das Brown-Korpus bietet jedoch verschiedene Varianten der englischen Sprache an, die sich auf verschiedene Gebiete beziehen. Die regionalen Varianten sind amerikanisches, britisches und indisches Englisch (Scherer, 2006, S. 30).

## **2.5. Institut für Deutsche Sprache (IDS) in Mannheim und Cosmas II**

*Das Institut für Deutsche Sprache in Mannheim (kurz IDS) wurde im Jahr 1964 gegründet und gerade im Jahr 2014 feiert es das 50-jährige Jubiläum seiner Entstehung. Das IDS stellt „die zentrale außeruniversitäre Einrichtung zur Erforschung und Dokumentation der deutschen Sprache in ihrem gegenwärtigen Gebrauch und in ihrer neueren Geschichte“ dar.*<sup>1</sup>

Institut für Deutsche Sprache bietet uns eine große Menge von Korpora, die vor allem für wissenschaftliche Zwecke benutzt werden. Diese Korpora werden nach der geschriebenen wie auch gesprochenen Sprache klassifiziert und die Texte, die in den Korpora zur Verfügung stehen, sind verschiedener Art. Die Korpora bilden belletristische oder wissenschaftliche Texte, aber den großen Anteil haben die Zeitungstexte (Scherer, 2006, S. 80-81).

Um die Daten zu recherchieren und analysieren, bietet IDS ein spezielles Programm an, COSMAS II. Die Arbeit mit diesem Programm ist kostenlos, aber man fordert eine vorherige Anmeldung, die man ohne Schwierigkeiten online über die Seiten des IDS machen kann (Scherer, 2006, S. 81).

---

<sup>1</sup> <http://www1.ids-mannheim.de/start/>; [zit. am 10.7.2014].

„*COSMAS II* wurde bereits 1993 von Cyril Belica (Leiter des Programmbereichs *Korpuslinguistik*) als Nachfolger von *COSMAS I* konzipiert.“<sup>2</sup> In *COSMAS II* (Corpus Search, Management und Analysis System)<sup>3</sup> sind Menschen aus 108 Ländern registriert und können in 114 Korpora recherchieren und analysieren.<sup>4</sup>

### 2.5.1. DeReKo

Das deutsche Referenzkorpus trägt den Namen *DeReKo* und es stellt die auf der Welt größte Sammlung deutschsprachiger Texte dar. Das Referenzkorpus ist über das Programm *COSMAS II* zugänglich.<sup>5</sup>

*DeReKo* enthält Texte der geschriebenen Sprache verschiedener Art wie z.B. belletristische oder wissenschaftliche Texte wie auch die Zeitungstexte. Die Texte die im *DeReKo* zur Verfügung stehen, beziehen sich nicht nur auf die Gegenwart, sondern auch auf die neuere Vergangenheit.<sup>6</sup>

Dank diesem Korpus kann man mit den Terminen „*Kollokation*“ und „*Kookkurrenz*“ arbeiten. Lemnitzer/Zinsmeister (2010, S. 188) bezeichnen Kollokation als „*das wiederholte gemeinsame Vorkommen zweier Wörter in einer strukturell interessanten Einheit.*“ „*In einer Kollokation beeinflusst ein Wort die Auswahl eines anderen Wortes zugunsten von Wörtern mit gleicher oder ähnlicher Bedeutung.*“

---

<sup>2</sup> <http://www.ids-mannheim.de/cosmas2/projekt/entwicklung/geschichte.html> ;[ zit. am 10.7.2014].

<sup>3</sup> <http://www.ids-mannheim.de/cosmas2/uebersicht.html> ; [zit. am 10.7.2014].

<sup>4</sup> <http://www.ids-mannheim.de/cosmas2/uebersicht.html> ; [zit. am 10.7.2014].

<sup>5</sup> <http://www1.ids-mannheim.de/kl/projekte/korpora.html> ;[zit. am 10.7.2014].

<sup>6</sup> <http://www1.ids-mannheim.de/kl/projekte/korpora.html> ; [zit. am 10.7.2014].

<b>Sprache 1</b>	<b>Sprache 2</b>	<b>Wörtliche Übersetzung</b>
Schlange stehen	sp: hacer cola	Schlange machen
sich die Zähne putzen	fr: se laver les dents	sich die Zähne waschen
den Tisch decken	en: lay the table	den Tisch legen
dichtes Haar	en: thick hair	dickes Haar
harte Währung	fr: devise forte	starke Währung

Tabelle 10:Kollokation in verschiedenen Sprachen

(übernommen von Lemnitzer/Zinsmeister, 2010, S. 142).

*„Als Kookkurrenz wird das gemeinsame Vorkommen zweier oder mehrerer Wörter in einem Kontext von fest definierter Größe bezeichnet. Das gemeinsame Vorkommen sollte höher sein, als bei einer Zufallsverteilung aller Wörter erwartbar wäre.“ (Lemnitzer/Zinsmeister 2010, S. 189).*

### **3. Substantive und ihre Charakteristika. Kontrast zwischen dem Deutschen und Tschechischen**

Nach Povejšil (1994, S. 110) sind Substantive eine Klasse der Wörter, die deklinierbar sind und die nach dem Genus, Numerus und Kasus klassifiziert werden können. Substantive bezeichnen Personen, lebendige, leblose oder abstrakte Dinge. Zu den Substantiven sind ihre Eigenschaften und Charakteristika zuzuordnen. Diese werden von Adjektiven bzw. Pronomen benannt, die auch deklinierbar sind. Helbig/Buscha (2001, S. 203) behaupten, Substantiv trage eine syntaktische und semantische Funktion.

Stellt man Kontrast zwischen Substantiven im Deutschen und im Tschechischen, wird zuerst die Rechtschreibung beachtet. Der erste Buchstabe der Substantive wird im Deutschen immer großgeschrieben (Povejšil, 1994, S. 110).

Im Unterschied zu tschechischen Substantiven werden deutsche Substantive mit einem Artikel, definiten, indefiniten oder Nullartikel gebraucht. Der Artikel hat sowohl seine grammatische Funktion – er gibt Information über Genus, Numerus und Kasus an, als auch stilistische oder semantische Funktion. Der indefinite Artikel wird beim Anwenden eines neuen Wortes benutzt, er kann aber über eine Zahl informieren, wobei der definite Artikel einen konkreten, jeweils bekannten oder außerordentlichen Inhalt bezeichnen kann.

Ansonsten verfügt Deutsch über vier Kasus, demgegenüber gibt es im Tschechischen sieben Kasus. Deutsch hat spezifische Mittel für Äußerung des Numerus. Dies passiert aus dem Grund, dass Deutsch analytische und Tschechisch syntaktische Sprache ist.

## 4. Substantive und ihre Genera

Im Deutschen, ähnlich wie im Tschechischen, gibt es drei Genera – Maskulina, Feminina, Neutra. Jedes der Substantivwörter gehört einer dieser Gruppe an. In Regionalvarianten können Genera dupliziert werden, d.h. einem Substantiv können zwei Genera zugeordnet sein (Povejšil, 1994, S. 111).

Es gibt die Kongruenz im Deutschen, sowie im Tschechischen. Im Deutschen kommt das Substantiv mit dementsprechenden Artikel, bzw. Adjektiv und Pronomen. Trägt der Artikel keine grammatische Information über Genus des Substantives, muss diese Information das Adjektiv tragen – *das gelbe Auto, ein gelbes Auto, der rote Mantel, ein roter Mantel*.

### 4.1. Biologisches Genus

Das Genus richtet sich teilweise nach dem biologischen Genus der Substantivwörter – *der Sohn, die Tochter, der Kater, die Katze, der Löwe, die Löwin, der Maler, die Malerin* usw. Natürlich gibt es Ausnahmen und zwar bei Diminutiva – *das Mädchen, das Tischlein, das Brötchen, oder bei regionalen Varianten – der / das Gulasch*.

Die gleiche Regel gilt bei Berufen – *der Arzt, die Ärztin, der Lehrer, die Lehrerin, der Professor, die Professorin* usw. (Povejšil, 1994, S. 113).

## 4.2. Grammatisches Genus

Das grammatische Genus spielt im Satzbau des Deutschen eine wichtige Rolle. Das Genus kann dank der Morphologie zugeordnet werden und zwar aufgrund der Suffixe, die oft genusspezifisch sind.

### 4.2.1. Genus bestimmende Suffixe

Was die Bestimmung von Genus angeht, gibt es konkrete Suffixe, die bei Maskulina, Feminina und Neutra vorkommen und nach denen Genus zu erkennen ist.

Von den maskulinischen Suffixen sind folgende Gruppen zu erwähnen:

- *-er, -ich, -ig, -ling, -s* – Lehrer, Fahrer, Teppich, Honig, Lehrling, Zwilling, Schwips.
- *-and, -ant, -är, -äst, -är, -ör, -eur, -ent, -ier, -ikus, -ismus, -ist, -oge, -or* – Praktikant, Doktorand, Milionär, Dynast, Masseur, Likör, Student, Kavalier, Liberalismus, Humorist, Autor.
- Null-Suffix – Flug, Zug, Gesang, Hieb.

Bei Feminina führt man folgende Gruppen an:

- *-e, -ei, -heit, -keit, -schaft, -ung, -t* - Erde, Rose, Konditorei, Freiheit, Freundschaft, Bezeichnung, Fahrt, Schlacht, Sucht.
- *-a, -ade, -age, -ance, -äne, -anz, -äre, -elle, -ette, -euse, -ie, -enz, -ere, -ik, -ille, -ine, -isse, -tät, -itis, -ive, -ose, -sis, -se, -ur, -üre* – Kamera, Blockade, Etage, Chance, Fontäne, Eleganz, Frikadelle, Etikette, Familie, Tendenz, Misere, Politik, Kamille.



Bei Neutra sind folgende Gruppen einzuführen:

- Substantive mit Präfix *Ge-* Gebirge, Gerede, Gewölbe, Gebüsch, Gebrüll.
- Deadjektivische und deverbale Substantive – Springen, Lesen, Atmen, Fahren.
- Substantive mit Suffixen *-ett, -eau, -o, -il, -in, -ma, -ment, -um* - Balett, Buffett, Plateau, Benzin, Nikotin, Asthma, Paradigma, Komma, Argument, Dokument, Abonnement, Konto.

#### 4.2.2. Genus nach semantischen Gruppen

Hinsichtlich der Semantik gibt es unter Substantiven Sachgruppen, die zu einem grammatischen Genus gehören.

Bei Maskulina gibt es folgende Gruppen:

- Jahreszeiten, Monate und Tage – der Frühling, Sommer, Herbst, Winter, Sonntag, Morgen, Abend.
- Himmelsrichtungen, Witterungseinflüsse, Winde – der Norden, Süden, Osten, Westen, Schnee, Reif, Hagel, Nebel, Monsun, Föhn, Tornado
- Alkoholgetränke – der Wodka, Kognak, Wein.
- Automobilmarken – der Felicia, Skoda, Audi, Renault, Mercedes.
- Berge und Gebirge – der Montblanc, Harz.
- Steine und Minerale – der Quarz, Granit, Kalk, Schiefer.

Bei Feminina kommen folgende Sachgruppen vor:

- Schiffe, Flugzeuge
- Bäume, Blumen
- Substantivische Numeralia
- Flüsse mit Endung *e, -a*

Bei Neutra werden folgende Sachgruppen eingeführt:

- Metall und Chemiestoffe
- Länder, Insel, Kontinente, Behörde (die ohne Artikel benutzt werden)
- Substantivische nicht flektierbare Wörter und Adjektive – das Ja und Nein, Hin und Her, das Schöne, Gewesene, Gewünschte.

### 4.2.3. Substantive mit zwei Genera

Im Deutschen gibt es Substantive, die über zwei Genera verfügen, deren Bedeutung aber gleich bleibt. Zu diesen Substantivwörtern gehören zum Beispiel – der, die Abscheu, der, das Barock, der, das Bonbon, der, das Filter, der, das Gelee, der, das Gulasch, der, das Keks, der, die Krem, der, das Liter, der, das Marzipan, der, das Sakko, der, das Teil, der, die Wulst, der, das Zubehör (Povejšil, 1994, S. 118).

Es kommen aber auch solche Substantive vor, bei denen zwei Genera zu bemerken sind, deren Bedeutung sich aber ändert (Povejšil, 1994, S. 118 - 119).

der Alp, die Alp

das Band, der Band

der Bulle, die Bulle

der Erbe, das Erbe

der Gefallen, das Gefallen

der Heide, die Heide

der Kaffee, das Café

der Leiter, die Leiter

der Mangel, die Mangel

der Messer, das Messer

der Positiv, das Positiv

der See, die See

der Tau, das Tau

das Wehr, die Wehr

der Weise, die Weise

## 5. Korpusanalyse

Der praktische Teil der Arbeit beschäftigt sich mit der Genuszuweisung bei den ausgewählten Wörtern. Es handelt sich um *Python*, *Samba*, *Rodel* und *Paprika*. Mit Hilfe vom Korpus *DeReKo* und *COSMAS II* wurden alle Substantive analysiert. Das Wort wurde in die Suchanfrage eingegeben, wobei nach der entsprechenden Artikel- und Substantivform mit dem Abstand von maximal 2 Wörtern gesucht wurde.

Das Referenzkorpus sucht jedoch mechanisch, darum müssen in der Ergebnisübersicht viele Wörter gestrichen werden. Manchmal kommen auch solche Belege vor, die einen ganz anderen Kontext haben und für die Analyse unwesentlich und unkorrekt sind. Aus diesem Grund werden bei der Analyse alle diese Belege manuell überprüft und aussortiert. Gestrichen wurden auch die Daten aus Wikipedia, und zwar wegen der unklaren territorialen Herkunft, aus der Analyse ausgelassen, obwohl dort manchmal diskutiert wurde, welches von Genera wo und wer benutzt. Sehr oft stößt man auf Probleme, wo die Wörter den gleichen Artikel haben, obwohl es um einen anderen Kasus und ein anderes Genus geht. Darauf muss man wirklich achten.

Nach dieser Datenreinigung ist die Anzahl der Treffer wesentlich gesunken, weil bei der Analyse nur die richtigen Daten analysiert wurden. Die Arbeit konzentriert sich auf die semantische und diatopische Ebene.

### 5.1. Paprika (die, der)

Nach Langenscheidt – Großwörterbuch Deutsch als Fremdsprache (2010) verfügt das Wort *Paprika* über die folgenden grammatischen und semantischen Charakteristika:

**Paprika** der; -s, -(s) oder die; -, -(s)

1. (der P.) e-e Pflanze mit großen, hohlen Früchten von grüner, gelber od. roter Farbe, die als Gemüse gegessen werden.
2. (der/die P.) die Frucht des Paprikas (1) ≈ Paprikaschote <gedünstete(r), gefüllte(r) P.> : P. in e-n gemischten Salat schneiden.
3. (der P.) nur sg; ein rotes Pulver, das man als (scharfes) Gewürz verwendet. z.B. Gulasch mit Paprika würzen.
4. In der Schweiz Paprika ≈ Peperoni

(übernommen aus dem Langenscheidt – Großwörterbuch Deutsch als Fremdsprache 2010)

Beim Wort *Paprika* wurden insgesamt 473 Treffer geliefert – 252 Treffer mit dem femininen Genus und 221 Treffer mit dem maskulinen Genus. Nach der Datenreinigung wurden beim maskulinen Genus 184 Belege gestrichen, für die Analyse stehen also 37 Belege zur Verfügung. Bei Daten mit dem femininen Genus wurden 224 Treffer gestrichen.

Bei der Datenreinigung wurden alle Belege ausgeschlossen, deren Quelle *Wikipedia* war. Aufgrund der unklaren Herkunft wurde *Wikipedia* für die Arbeit nicht akzeptiert.

Bei der Bearbeitung der maskulinen Variante von *Paprika* muss man darauf achten, dass viele Belege die Genitiv- und Dativform Singular der femininen Variante darstellen. Nächste Belege, die aus dem Korpus geliefert wurden, standen für Genitiv der Pluralform. Solche Daten wurden auch gestrichen. Man muss auch in Erwägung ziehen, dass sich in der Pluralform zwei Möglichkeiten anbieten, wie der Plural ausgedrückt werden kann.

Plural: *die Paprika[s] oder die Paprika*

(Duden - Deutsches Universalwörterbuch, 2007. S. 1254)

Auf den Märkten werden **die** letzten **Paprika** gekauft. Im Fünf- und Zehnkilopack tragen die Frauen sie zu den grossen Feuerstellen, die man überall dieser Tage in Belgrad sieht. Auf Backblechen rösten sie im Freien die Paprika und bereiten in Gläsern mit viel Knoblauch "Ajvar", ein köstliches Paprikamus, zu. [...] (L99/OKT.75052 **Berliner Morgenpost**, 25.10.1999, S. 7, Ressort: POLITIK; "Die drucken Scheine wie Monopoly-Geld")

Ferner wurden aus dem Korpus Belege geliefert, die sich in der Akkusativform befinden. Solche Belege wurden gestrichen, aber man kann hier *Paprika* in der Pluralform mit der Endung *-s* sehen:

und die Paprika zu den begehrtesten Gemüsen. Jetzt im Hochsommer kann man vor allem **die grünen Paprikas** 3 S bis 6 S (St.) und 15 S bis 30 S (kg) von unseren heimischen Bauern, die gelben und die Tomatenparika 4 S bis 6 S (St.) und 20 S bis 35 S (kg) und die roten und gelben Spitzparika 3 S bis 5 S (St.) meist aus dem Burgenland, kaufen. So ist nun die Zeit der vielgeliebten, gefüllten Paprika angebrochen. [...] (N98/AUG.32070 **Salzburger Nachrichten**, 21.08.1998, Ressort: LOKALES; **Paprika**)

Weiter erschienen die Fälle, wo die Paprika mit einem Adjektiv, das die Farbe bezeichnet, vorkommt.

Paprika sorgt nicht nur für Farbe auf den Tellern, sondern ist ausgesprochen gesund. Mit nur 30 Kalorien pro Stück ist **der** grüne **Paprika** die kalorienärmste Sorte, gefolgt vom gelben mit 45 und vom roten mit 55 Kalorien pro Stück. Grüner Paprika ist reich an Vitamin A und trägt damit zu Wachstum und Aufbau von Haut und Schleimhäuten bei. Das im gelben Paprika enthaltene Rutin verlangsamt den Alterungsprozess, Vitamin C wirkt durchblutungsfördernd. [...] (BVZ08/JUL.00958 **Burgenländische Volkszeitung**, 09.07.2008, S. 13; **Vitaminbombe Paprika**)

Daß viel in ihm steckt, mehr als in den meisten anderen Kohllarten, weiß man schon lange. Und trotzdem fragten wir uns immer wieder, warum sich die Italiener schon seit jeher so um den Brokkoli rissen. Der "wilde Bruder" entgiftet besonders im Magen- und Darmbereich, aber auch die Lungen profitieren von der grünen Feuerwehr. Brokkoli enthält etwa dreimal soviel Vitamin C wie Äpfel und beinahe soviel wie **der** rote **Paprika**, er stärkt die Immunabwehr. (K99/MAI.38083 **Kleine Zeitung**, 28.05.1999, Ressort: **Klagenfurt; Er ist der wilde Bruder**)

Nur 15 Minuten bleiben Julia Jöckel (23) und Christian Hammerl (25), um die Gemüsesorten im Kreis anzuordnen - und zwar in der richtigen Reihenfolge. Ob nun der Lollo-Rosso neben der Salatgurke liegen darf, oder **die** gelbe **Paprika** das Pendant zum Radieschen bildet, das mag für den gärtnerischen Laien unerheblich sein - beim Wettstreit der Junggärtner bringt der Griff in die falsche Gemüsebox gleich dicke Minuspunkte. (R99/JUN.45231 **Frankfurter Rundschau**, 09.06.1999, S. 36, Ressort:

**WIRTSCHAFTSSPIEGEL; Der gärtnerische und floristische Nachwuchs kämpfte beim Berufswettbewerb um den Landessieg)**

Andere, im Korpus gefundenen Belege, weisen auf die Herkunft hin: 1. Herkunft des Wortes *Paprika* 2. Herkunft der Gemüsepflanze *Paprika*.

Der Paprika ist aus der ungarischen Küche nicht wegzudenken. Sowohl als Gemüse als auch als Pulver oder Püree ist er ganz typisch für die Gerichte aus dem Magyaren-Land. Ursprünglich kommt **der Paprika** allerdings aus Mittel- und Südamerika, wo er seit mehr als 3000 Jahren kultiviert wird. (M06/MAR.19575 Mannheimer Morgen, 09.03.2006; Paprika entfacht Feuer)

Das etymologische Wörterbuch des Deutschen (2005, S. 968) führt an, dass die Gemüsepflanze aus Südamerika stammt und später in Südosteuropa angebaut wird. Die Herkunft des Namens ist von serbokr. „*pàpar*“ abgeleitet.

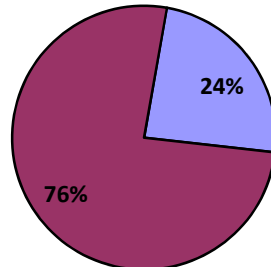
Jeder Mensch kennt die Paprika als Zusatz beim Kochen. Viele Artikel geben Informationen über die Zubereitung der Paprika oder auch, dass die Paprika als ein Ingredienz in ein Rezept gehört.

In einer Pfanne etwas Öl erhitzen, mit der Suppe aufgießen und kochen, bis **der Paprika** weich ist. Mit dem Mixer fein pürieren und dann mit Salz und Pfeffer, Knoblauch und Thymian würzen. (BVZ09/MAR.03252 Burgenländische Volkszeitung, 25.03.2009, S. 39; Scholle natur)

Bei vielen anderen Texten wurde als Quelle Wikipedia angeführt, solche wurden also für ihre unklare territoriale Herkunft gestrichen. Trotzdem kann man als interessante Tatsache erwähnen, dass in der Wikipedia umfangreiche Diskussionen über das Genus von *Paprika* (die, der) und die Anzahl der Vitamine der Paprika geführt werden.

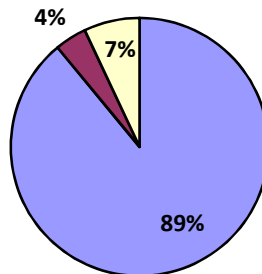
In der Ergebnisübersicht von *der Paprika*, also vom Maskulinum, wurde das häufigste Vorkommen in Österreich (28 Erscheinungen) erkennbar und in Deutschland erschien *der Paprika* nur 9mal.

### Paprika als Maskulinum



Das Femininum bei Paprika gibt es insgesamt 28mal, das Vorkommen ist in Österreich und in der Schweiz niedrig. In Deutschland erschien die Paprika 25 Mal.

### Paprika als Femininum



### Zusammenfassung:

Beim Substantiv *Paprika* musste man darauf achten, dass viele Beispiele beim Wort *der Paprika*, die das Korpus geliefert hat, nicht im Nominativ waren, sondern als Genitiv und Dativ Singular vom Femininum.

Andere Beispiele waren wieder im Genitiv der Pluralform. Damit war die Datenreinigung anspruchsvoll.

In Österreich kamen 28 Treffer vor, bei denen *Paprika* als Maskulinum verwendet wurde, in Deutschland erschien *der Paprika* nur 9mal.

Feminine *Paprika* erschien am meisten in Deutschland (25 Erscheinungen), in Österreich und in der Schweiz ist das Vorkommen niedrig (insgesamt 3 Belege).

## 5.2. Samba (die, der)

Das Wort *Samba* weist nach Duden – Universalwörterbuch (2007, S. 1430) folgende grammatische und semantische Zeichen auf:

**Samba**, die; -, -s, auch, österr. nur : der; -s, -s [port. (bras.) samba, aus einer afrik. Spr.]: *beschwingter und spritziger Gesellschaftstanz im 2/4- Takt (nach einem brasilianischen Volkstanz).*

Beim Wort *Samba* wurde der Gebrauch vom femininen und maskulinen Genus verglichen. Wie schon oben gesagt wurde, spielt bei dem maskulinen Genus die identische Form mit dem Genitiv und Dativ Femininum und mit dem Genitiv der Pluralform eine wesentliche Rolle. Solche Belege wurden gestrichen, wie auch die Texte, deren Herkunft Wikipedia war.

Vorerst kommen allerdings nur die Altrhoi-Schnougen ins Schwitzen, die für ihren Auftritt proben. Für den Tanz ist nicht nur ein gekonnter Hüftschwung zu meistern, sondern auch die eine oder andere Schrittkombination. "Und ohne Kostüme kommt man eh nicht so in Schwung", meint die Choreografin Christa Müller, die mit Gerda Packlin die Tanztruppe leitet. Doch bei **der Samba** nehmen die Frauen ihre Rasseln zur Unterstützung und schon kommt der richtige Rhythmus von ganz alleine. (M07/JAN.06061 **Mannheimer Morgen, 26.01.2007; Schnougen schwitzen schon**)

Außer Daten, die in Wikipedia gefunden wurden, wurden vor allem bei der Suchanfrage *der Samba* und *die Samba* auch solche Belege abgelehnt, bei



denen sich das Genus nicht auf das Wort Samba bezog. Es handelt sich hauptsächlich um die Texte, wo *Samba* den Namen einer Tanzgruppe trägt.

In diesem Jahr war auch ein besonders attraktives Programm zusammengestellt worden: Die Kinder-Bauchtanz-Gruppe der Braunschweiger „Kawanserei“ führte auf dem Sportplatz ihre Künste vor. Neben Trommlern aus Gifhorn und der Meiner Kindertanzgruppe konnte **der „Samba Cor do Brasil“** aus Hannover gewonnen werden. **(BRZ09/AUG.07956 Braunschweiger Zeitung, 18.08.2009; Samba-Klänge verzücken)**

Beim Wort Samba wurden insgesamt 816 Treffer exportiert, es handelte sich also um umfangreichstes Material. Nach der Datensortierung, vor allem wegen der Wikipedia-Artikel oder wegen des Zusammenhangs mit anderen Kasus, kann man mit 125 Belegen arbeiten, davon 66 bei maskulinem Samba und 59 bei femininem Samba.

Viele im Korpus gefundenen Belege werden mit einem Adjektiv verbunden. Vor allem handelt es sich dabei um das Adjektiv: *brasilianisch, heiß und kubanisch*.

Nicht nur leichtfüßig, auch leichtfüßig muss **die** brasilianische **Samba** sein – man soll ja dazu tanzen können, und zwar stunden- oder nächtelang, gerade beim Karneval. Die Sängerin Marisa Monte, in ihrer Heimat ein Superstar, hatte nun die schöne Idee, eine der großen Samba-Schulen von Rio de Janeiro aufzusuchen und die alte Garde von Musikern, die dort unterrichtet, ihre geheimen Schatzkästchen öffnen zu lassen. Und das hat die Velha Guarda da Portela dann auch gern getan: Wie beim kubanischen Buena Vista Social Club geben die betagten Sänger und Spieler auf „Tudo Azul“ (Red Circle) ihr Bestes – und bringen Lieder zu Gehör, die man sonst kaum kennen kann. Anregend und anrührend! **(NUZ08/FEB.02494 Nürnberger Zeitung, 25.02.2008, S. 7; CD-Tipps: Weltmusik - Eine Rose ohne Dornen)**

Während im Konzertsaal einige beim Fußball abhängen und aus dem Nachbarraum **die** heiße **Samba** des brasilianischen Kulturtages im KUZ herüberdröhnt, hält Bessler vor 20 Interessierten den ersten Vortrag seines Lebens. Im Galopp berichtet der funk- und fernseherfahrene Aktivist von der Unterdrückung der Yanomami durch Siedler und Goldsucher, erklärt die Folgen der Regenwald-Rodung und nennt dafür verantwortliche Firmen beim Namen. "War wohl noch etwas zu schnell", räumt er hinterher ein. [...] **(RHZ96/JUN.14466 Rhein-Zeitung, 25.06.1996; Wanderer übt Überleben)**

Der Samba ist ein brasilianischer Tanz, der mit dem weltbekannten Karneval in Rio de Janeiro verbunden ist. Wann der Samba entsteht? Wohin greifen die Wurzeln des Samba? Wie bedeutungsvoll ist der

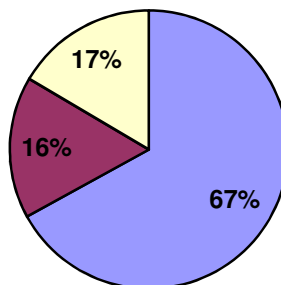
Samba? Das alles sind die Themen, die in den Texten vorkommen. Was für den Brasilianer **die Samba** ist, ist dem Argentinier der Tango. So einfach könnte man es ausdrücken. [...] (RHZ06/SEP.00182 Rhein-Zeitung, 01.09.2006; Tango erklingt in Kirche)

Sehr oft kommen auch Artikel vor, bei denen es sich um ein Tanzspiel handelt, und zwar am meisten in der Rubrik Sport.

Für *den maskulinen Samba* wurden 66 Treffer gefunden. In der Schweiz und in Österreich wurde das Vorkommen wieder das niedrigste. In Deutschland erschien *der Samba* 44mal.

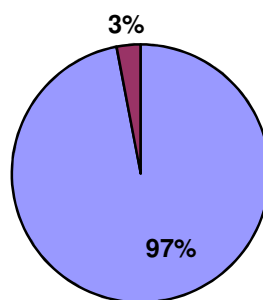
#### Samba als Maskulinum

■ Deutschland ■ Österreich ■ Schweiz



*Die Samba* kommt deutlich häufiger fast im ganzen deutschsprachigen Raum vor. Für *die Samba* wurden im Korpus 59 Fälle identifiziert. In Deutschland kam *die Samba* mit 57 Erscheinungen vor und für Österreich blieben nur 2 Belege zur Verfügung. Hier kann man einen markanten Unterschied sehen.

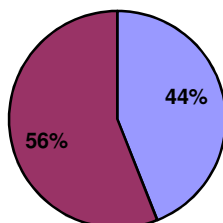
### Samba als Femininum



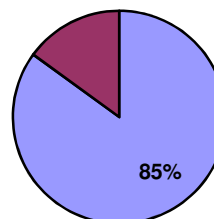
### Zusammenfassung:

Im Allgemein kann man sagen, dass *Samba* mit beiden Genera häufiger in Deutschland auftritt. In Österreich steht an der ersten Stelle *der Samba*, hier stimmt auch die Bemerkung von Duden – Universalwörterbuch (2007). Und für die Schweiz kommen nur solche Belege vor, die sich mit maskulinem *Samba* binden (100%).

### Genus in Deutschland



### Genus in Österreich



### 5.3. Python (die, der)

*Duden – Universalwörterbuch* (2007, S. 1338) bestimmt das Substantiv *Python* grammatisch und semantisch folgend:

**Python**, der; -s, -s, nicht fachspr. auch : die; -, -s,

**Pythonschlange**, die [*nach der von Appolo getöteten Schlange Python, die das Orakel in Delphi behütete*]: in Afrika, Südasien u. Nordaustralien lebende Riesenschlange.

Vom Deutschen Referenzkorpus wurden 267 maskuline Treffer exportiert, davon konnte man mit 54 Treffern arbeiten. Für die Variante *die Python* wurden aus dem Korpus insgesamt 132 Treffer gefunden, von denen nur 13 Vorkommen relevant sind.

Außer Daten, die in Wikipedia gefunden wurden, wurden hauptsächlich bei der Suchanfrage *der Python* solche Belege ausgelassen, bei denen sich das Genus nicht auf das Substantiv *Python* bezog und Belege, wo das Wort als Genitiv und Akkusativ Femininum steht. Bei dieser Analyse muss man nicht auf die Pluralform achten, wie z.B. beim Wort *Rodel*, weil bei beiden Genera (die, der) die Pluralform die Endung *-s* enthält.

#### Die Vergleich DER und DIE Python

	Singular DIE	Singular DER	Plural
<b>Nominativ</b>	die Python	der Python	die Pythons
<b>Genitiv</b>	der Python	des Pythons	der Pythons
<b>Dativ</b>	der Python	dem Python	den Pythons
<b>Akkusativ</b>	die Python	den Python	die Pythons

Belege, bei denen Python als eine Programmiersprache, die sich als Software beim Computer befindet und sich nicht auf *Python* als Tier (Schlange) bezieht, wurden auch ausgelassen.

Wie schon oben gesagt wurde, Python ist eine extrem lange Schlange, die in Afrika, Südasien und Australien lebt. Viele Belege geben die Informationen über die Herkunft der Python an, sowohl auch Informationen über verschiedene Gattungen der Pythons. Andere Zeitungsartikel informieren die Gesellschaft darüber, was die Python gegessen hat.

Im australischen Kuanda hat ein Hundebesitzer durch einen beherzten Einsatz seinen Chihuahua vor einem Python gerettet. Die drei Meter lange Schlange hatte sich in einem Garten an den winzigen Hund herangemacht. Doch der Besitzer entriss der Schlange das Tier rechtzeitig, wie die Zeitung «Cairns Post» berichtete. Der Chihuahua überlebte, **der Python** wurde in der Wildnis ausgesetzt. (A10/MAR.04221 St. Galler Tagblatt, 13.03.2010, S. 10; Wie bitte?)

Sydney: Erst fraß **der Python** einen Jack-Russell-Terrier, dann lauerte er einem Chihuahua auf – dessen australischer Besitzer entriss der Riesenschlange jedoch das Tier. Der Chihuahua überlebte, berichtete die Zeitung „Cairns Post“ am Freitag. Der drei Meter lange Python hatte sich in einem Garten in Kuranda 15 Kilometer nordwestlich von Cairns in Nordaustralien an den kleinen Hund herangemacht. Ein Schlangenspezialist fing den Python ein und setzte ihn weit außerhalb der Stadt in der Wildnis aus. (HAZ10/MAR.00706 Hannoversche Allgemeine, 13.03.2010; Chihuahua vor Python gerettet)

In den Nürnberger Nachrichten kommt ein Artikel vor, der sich mit dem Genus beschäftigt.

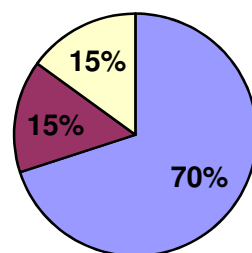
Übrigens: Die Einzahl heißt „**der Python**“ und nicht „die Python“, wie sogar die meisten Erwachsenen glauben (und sagen). Richtig ist allerdings „die Python-Schlange“. (NUN09/JUL.03238 Nürnberger Nachrichten, 01.07.2009, S. 23; So lang wie ein Kleinbus - Pythons erlegen ihre Beute, indem sie diese erdrücken)

Der oben angeführte Artikel stellt uns ein Beispiel vor, auf Grund dessen viele andere Belege aus der Analyse gestrichen wurden, weil sich der feminine Artikel mit dem Substantiv *Schlange* verbindet und nicht mit dem Wort *Python*. Deswegen hat der Python den Artikel *die*.

Beim Wort *der Python* wurden nach der Datenreinigung 213 Treffer gestrichen. Für die Analyse bleiben 54 Erscheinungen zur Verfügung. Das häufigste Vorkommen wurde in Deutschland und zwar mit 38 Erscheinungen. Österreich und die Schweiz haben den gleichen Anteil (8 Erscheinungen) gewonnen.

### Python als Maskulinum

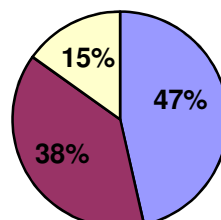
■ Deutschland ■ Österreich □ Schweiz



Zu der femininen Genuszuweisung bei *Python* wurden im Referenzkorpus 132 Treffer gefunden. Nach der Datenreinigung konnte man mit nur 13 Belegen arbeiten. Das niedrigste Vorkommen wurde für die Schweiz mit zwei Belegen nachgewiesen. In Österreich wurde *die Python* mit 5 Erscheinungen bestätigt und das häufigste Vorkommen ist wieder in Deutschland mit 6 Erscheinungen.

### Python als Femininum

■ Deutschland ■ Österreich □ Schweiz

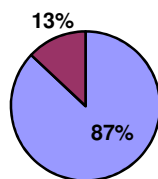
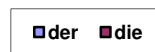


### Zusammenfassung:

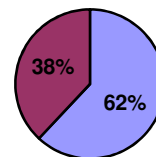
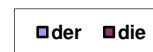
In der Ergebnisübersicht von *dem Python*, also vom Maskulinum, verfügte das Korpus 267 Treffer. Nach der Datenreinigung konnte man mit 54 Erscheinungen arbeiten. Das häufigste Vorkommen wurde in Deutschland (38 Erscheinungen). In Österreich und in der Schweiz kam diese Genusvariante nur 8mal vor.

Das niedrigste Vorkommen beim Wort *die Python* war in der Schweiz (2 Belege). In Österreich befunden sich 5 Erscheinungen. In Deutschland kamen 6 Belege vor.

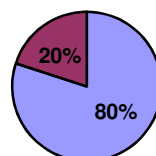
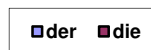
#### Genus in Deutschland



#### Genus in Österreich



#### Genus in der Schweiz



## 5.4. Rodel (die, der)

Das Substantiv *Rodel* wird nach *Duden – Universalwörterbuch* (2007, S. 1403) grammatisch und semantisch folgend charakterisiert:

**Rodel**, der; -s, -

1. [*spätmhhd. rodel* = *Urkunde, Register* < *lat. rotula*, ↑ *Rolle*] (*südwestd., schweiz.*) : *Liste, Verzeichnis*.
2. [*H.u.*] (*bayr.*) : *Rodelschlitten*.

**Rodel**, die; -, -n [*1: zu* ↑*2 Rodel*; *2: zu mundartl. rodeln* = *rütteln, schütteln*] (*österr.*) :

1. *kleiner Schlitten*
2. *Kinderrassel*

Nach *Duden – Universalwörterbuch* (2007) kann man behaupten, dass *die Rodel* häufiger in Österreich vorkommt und die männliche *Rodel* wieder in Deutschland und in der Schweiz. Nach meiner Analyse werden wir sehen, ob wir diese Behauptung bestätigen oder widerlegen.

Bei der Suchanfrage wurden insgesamt 475 Treffer exportiert. 317 Treffer davon werden für das Wort *der Rodel* gefunden und der Rest 158 Treffer verbindet sich mit der Variante *die Rodel*.

Wie bei dem Wort *Paprika* wurden auch hier die femininen Genitiv- und Akkusativformen gestrichen. Eine Menge von Daten verlor man nach der Datensortierung, vor allem wegen der Präpositionen. *Zum Rodel*, also zum Maskulinum, wurden 317 Treffer gefunden, von denen kann man aber nur mit 17 Vorkommen arbeiten kann. Die Mehrheit von der maskulinen Variante *der Rodel* ist mit der Präposition verbunden und zwar 129 Treffer sind mit der Präposition *mit* verbunden, an der zweiten Position mit 81 Treffer ist die Präpositin *auf* und den dritten Platz belegte die Präposition *von* (27 Treffer).



Man kann noch andere Präpositionen finden, wie z.B. *an* (7 Belege), *bei* (2 Belege) oder *unter* (2 Belege). Diese Präpositionen regieren in diesen Belegen den Dativ bei Feminina. Die Aufmerksamkeit muss man noch den anderen Kasus zuwenden und zwar in der Pluralform befindet sich auch der Artikel *der* aber *der* gilt für den Genitiv.

Wenn die Quelle Wikipedia war, wurden diese Beispiele gestrichen.

### Der Vergleich DER RODEL

	Singular	Plural
<b>Nominativ</b>	<b>der</b> Rodel	die Rodel
<b>Genitiv</b>	des Rodels	<b>der</b> Rodel
<b>Dativ</b>	dem Rodel	den Rodeln
<b>Akkusativ</b>	den Rodel	die Rodel

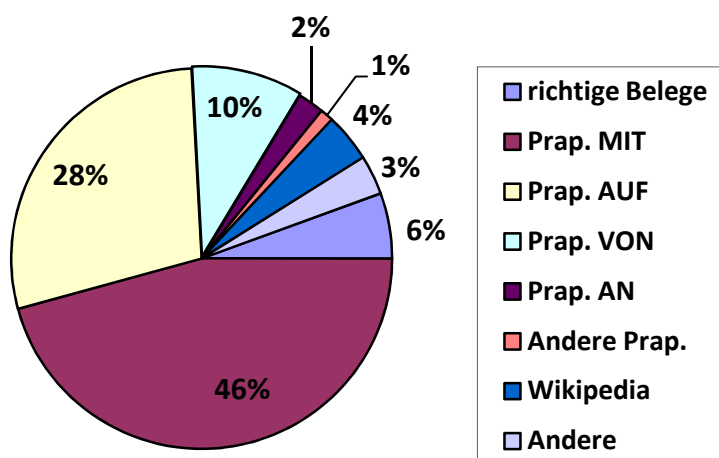
### Der Vergleich DIE RODEL

	Singular	Plural
<b>Nominativ</b>	die Rodel	die Rodeln
<b>Genitiv</b>	<b>der</b> Rodel	<b>der</b> Rodeln
<b>Dativ</b>	<b>der</b> Rodel	den Rodeln
<b>Akkusativ</b>	die Rodel	die Rodeln

Zielpunkt ist der gemütliche Alpengasthof Mahdegg in 1200 Meter Höhe. Bei guter Schneelage ist die Rückfahrt mit **der Rodel** die sicherlich lustigste und schnellste Abstiegsvariante. (N95/DEZ.50455 Salzburger Nachrichten, 28.12.1995; Dem Hochkönig gegenüber)

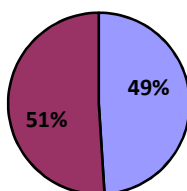
Der Tischberger war der Kohlenhändler, und im Winter transportierten sie die Briketts auf **der Rodel** heim. Das war billiger, weil sie sich dadurch das Geld für die "Zubringung" sparten. Auch das Brennholz holten sie auf die gleiche Art heim. Nur gingen sie dazu mit ihrer Rodel bis in den Passianerwald. Dieser Wald liegt in Ebelsberg und um dorthin zu gelangen, mußten sie über die Traunbrücke. [...] (X96/DEZ.29896 Oberösterreichische Nachrichten, 07.12.1996, Beilage, Ressort: Magazin; FRUAHLIACHD)

## Eine Übersicht über der Datensortierung (der Rodel)



Von den 17 „maskulinen Rodeln“ hat dieses Wort 7mal die deutsche Herkunft, die zehn anderen kommen in der Schweiz vor. Für Österreich wurde kein Beleg gefunden. Nach der Analyse der Daten kann man bestätigen, dass die männliche Variante wirklich in Deutschland und in der Schweiz auftritt.

### Rodel als Maskulinum



Zur Form *die Rodel*, also zum Femininum, wurden insgesamt 158 Treffer gefunden. Bei vielen wurde als Quelle wieder Wikipedia angegeben. Man muss wieder aufpassen, dass viele Belege in der Pluralform vorkommen. Das Wort *der Rodel* hat im Genitiv Plural keine Endung *-n* (siehe in der Tabelle oben) und man kann es leicht mit dem Nominativ Maskulinum

Singular verwechselt. Wenn man aber den Vollbeleg durchliest ist aus dem Kontext klar, dass es sich um den Plural handelt. Diese Belege wurden aus der Analyse weggelassen.

Nach der Aussortierung konnte man mit 37 Erscheinungen rechnen. Sehr oft handelte sich um Texte, die in der Rubrik Sport vorkommen. Diese Texte geben dem Leser Informationen über *Rodel* und zwar Rodel als ein Sport. Schlitten fahren macht Spaß und für die Kinder ist das im Winter das beliebteste Gerät, auch solche Texte sind in der Analyse zu finden.

COMPUTERSPIELE HIN, FERNSEHEN HER sobald der erste Schnee fällt, steht bei allen Kindern **die Rodel** ganz hoch im Kurs: Diesen Kindern in Westendorf scheint das rasante Vergnügen jedenfalls mächtig Spaß zu machen. Und solange der Schnee liegt, wird jeder Hang genützt denn der nächste Wärmeeinbruch kommt bestimmt. **(I98/NOV.47934 Tiroler Tageszeitung, 30.11.1998, Ressort: Regional Unterinntal; Der Rodelspaß kommt nie aus der Mode)**

Trotz Snowboard und Carvingski ist **die Rodel** nach wie vor eines der beliebtesten Wintersportgeräte in Salzburg geblieben. Aus einfachsten Anfängen hat sie sich zu einem perfekten Sportgerät entwickelt. Die ältesten Rodeln hatten Kufen aus Rinderknochen, später wurden dann die Kufen aus Brettern gesägt oder aus krummen Hölzern gehobelt. Namen wie Brettlhupfer oder Rumpler erinnern an deren eher unruhige Fahrweise. 1883 fand in Davos das erste Schlittenrennen statt, bei den Nordischen Spielen 1904 in Mürzzuschlag bildete das Preisrodeln bereits einen Höhepunkt. Dabei waren schwere Stahlschlitten im Einsatz. [...] **(O98/FEB.14772 Neue Kronen-Zeitung, 13.02.1998, S. 22; Vom Schlitten zur Rennrodel)**

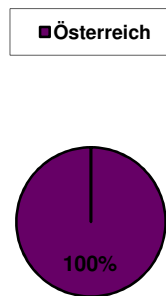
Die Zeitungen führen auch die zweite Seite der Rodel an und zwar, dass mit der Rodel viele Verletzungen und Unfälle geschehen können. Der Unfall ereignete sich bereits am Samstagnachmittag, 29. Juli. Wie die Staatsanwaltschaft Trier unserer Zeitung auf Anfrage mitteilte, war **der Rodel** mit den beiden Insassen bei hoher Geschwindigkeit aus der Führungsschiene der Bahn geraten und hatte sich mehrfach überschlagen. **(RHZ06/AUG.13059 Rhein-Zeitung, 15.08.2006; Schlitten wurde aus der Bahn geworfen)**

In der Frankfurter Rundschau wurde ein Text gefunden, der sich mit der Genuszuweisung (maskulin, feminin) beschäftigt. Bei den anderen analysierten Wörtern (*Samba, Paprika*) wurde diese Diskussion nur auf Wikipedia geführt.

Sind Rodel männlich oder weiblich? Bei den österreichischen Nachbarn heißt es "die Rodel" anstatt "**der Rodel**". Uneingeschränktes Rodel-Vergnügen jeglichen Geschlechts garantieren eine kurze Ausbildung für richtiges Lenken und Bremsen sowie regelmäßig präparierte Pisten. Gelegenheitsrodler wissen um das Gefühl, auf längeren oder steilen Wegen ihr Gerät nur unzureichend im Griff zu haben und nach einer unübersichtlichen Kurve hilflos über Eis und Steine zu rumpeln. (R98/JAN.05956 Frankfurter Rundschau, 24.01.1998, S. 2, Ressort: REISE; Im Tiroler Wipptal wird Schlittenfahren ernst genommen)

Die Belege, die für das Wort *die Rodel* nach der Selektion untersucht wurden, stammen alle aus Österreich. Im Allgemeinen kann man sagen, dass *die Rodel* „nur“ in Österreich vorkommt.

### Rodel als Femininum



### Zusammenfassung:

Nach den Angaben, die in Korpus für das Substantiv *Rodel* geliefert wurden, stehen 457 Treffer zur Verfügung. Eine große Überraschung war, dass viele Belege wegen der Verbindung mit einer Präposition gestrichen wurden und nicht wegen Wikipedia.

In dieser Analyse hat man auch die Behauptung von *Duden – Universalwörterbuch* (2007) bestätigt, und zwar, dass *die Rodel* in Österreich vorkommt und *der Rodel* häufiger in der Schweiz und Deutschland auftritt.

## 6. Schluss

Das Hauptthema, mit dem sich diese Bachelorarbeit beschäftigt, ist die „*Genusvariation deutscher Substantive*“ am Beispiel von *Paprika*, *Rodel*, *Python* und *Samba*. Das Substantiv kann man deklinieren und es gehört zu den wichtigsten und häufig verwendeten Wortarten der deutschen Sprache. Bei einem Substantiv können die folgenden grammatischen Kategorie identifiziert werden: das Genus, der Numerus und der Kasus.

In dem teoretischen Teil dieser Bachelorarbeit wurden das grammatische und biologische Genus vorgestellt. Im Rahmen des grammatischen Genus unterscheidet man nach Povejšil (1999, S. 111) zwischen dem Maskulinum, Femininum und Neutrum. Das natürliche Genus wird durch das Geschlecht der bezeichneten Person oder des Tiers bestimmt.

*der Vater – der Mutter*

*der Stier – die Kuh*

Das Genus kann dank der Morphologie zugewiesen werden und zwar aufgrund der Suffixe.

Ferner wurde sowohl die gegenwärtige Korpuslinguistik vorgestellt, wie auch der historische Überblick. Die Korpuslinguistik ist eine Disziplin, die sich mit der Untersuchung der menschlichen Sprache befasst und einen engen Zusammenhang mit anderen Nachbardisziplinen hat. Am Ende dieses Kapitels werden sowohl die Arten von Korpora vorgestellt, wie auch ihre Verwendung.

Im praktischen Teil wurde eine Korpusanalyse, mit Hilfe von COSMAS II und *DeReKo*, bei den ausgewählten Substantiven durchgeführt.

Die Analyse konzentriert sich vor allem auf die Genuszuweisung der deutschen Substantive.

Bei den analysierten Substantiven stieß man ziemlich oft auf interessante Bemerkungen zur Genusschwankung wie z.B. in der Frankfurter Rundschau oder auf Wikipedia. Diese Texte befassten sich gerade mit der Genuszuweisung, oder wo und welche Genera benutzt werden oder benutzt werden sollen.

Die Ergebnisse der Korpusanalyse zeigten, dass im Korpus viele Belege für alle Substantive gefunden wurden. Das umfangreichste Material wurde bei *Samba* exportiert (insgesamt 816 Belege). Nach der Analyse wurde festgestellt dass *der Samba* am häufigsten in Deutschland vorkommt. In der Schweiz, sowohl auch in Österreich wurde *der Samba* nur 11mal verwendet *Die Samba* kommt deutlich häufiger fast im ganzen deutschsprachigen Raum vor. Für *die Samba* wurden im Korpus 59 Fälle identifiziert. In Deutschland kam *die Samba* mit 57 Erscheinungen vor und für Österreich blieben nur 2 Belege zur Verfügung.

Beim Wort *Rodel* wurde eine zahlreiche Menge von Belegen wegen ihrer Verbindung mit einer Präposition gestrichen. Obwohl das Material für die Analyse nach der Datensortierung kleiner wurde, kann man die Behauptung von Duden-Universalwörterbuch (2007) bestätigen, und zwar dass *die Rodel* am häufigsten in Österreich vorkommt und *der Rodel* am meisten in Deutschland verwendet wird.

*Paprika* als Maskulinum erschien mit 28 Belegen in Österreich und nur 9mal in Deutschland. *Die Paprika* wurde wieder am häufigsten in Deutschland verwendet, in Österreich und in der Schweiz ist das Vorkommen sehr niedrig.

Im Allgemeinen kann man sagen, dass *Python* mit beiden Genera am meisten in Deutschland vorkommt (*der Python* 38mal und *die Python* 6mal). In Österreich steht an der ersten Stelle *der Python* und an der zweiten Stelle *die Python*. In der Schweiz erschien *der Python* 8mal und *die Python* nur 2mal.

Nach der Arbeit mit einem Korpus kann man bestätigen, dass für die Untersuchung der Gegenwartssprache das Korpus wirklich nutzbar ist.

## 7. Resumé

Stěžejní téma této práce bakalářské bylo zkoumání kolísání rodu u německých substantiv *Paprika*, *Rodel*, *Python* a *Samba*. Substantivum je jeden ze základních slovních druhů, který můžeme skloňovat. Určujeme u něj tedy rod, číslo a pád. V německém jazyce je rod s podstatným jménem pevně spojen, tudíž není možné si člen zvolit.

Povejšil (1999, S. 111) v rámci gramatického rodu rozlišuje rod na: ženský (femininum), mužský (maskulinum) a střední (neutrum). Avšak rod přirozený (biologický), který se nejčastěji vyskytuje u označení osob a zvířat, rozlišujeme pouze mezi maskulinem a femininem. V německém jazyce může být rod určen i na základě morfologie, a sice pomocí sufixů. Sufix může tedy podstatnému jménu přiřadit jeden ze tří rodů. (Povejšil, 1999, S. 114)

V teoretické části byla dále charakterizována korpusová lingvistika a to jak ze současného, tak i z historického hlediska. V neposlední řadě byly představeny druhy korpusů a jejich využití. Na konci teoretické části byl představen Deutsches Referenzkorpus (*DeReKo*) a jeho program COSMAS II, na jehož základě byla provedena korpusová analýza.

V následující praktické části byla provedena korpusová analýza vybraných substantiv. Tato analýza se soustředila především na kolísání rodu německých podstatných jmen. Za pomoci programu COSMAS II a korpusu *DeReKo* byly zjištěny výsledky, které byly zpracovány a vyhodnoceny. Analýza bude soustředěna především v oblasti sémantické a diatopické.



Ke slovu *Samba* bylo exportováno velké množství materiálu. Po vyhodnocení výsledků se ukázalo, že se slovo *Samba* jako maskulinum nejvíce vyskytuje v Německu, méně pak v Rakousku a Švýcarsku. Stejně tak *Samba* jako femininum. Její největší výskyt je opět v Německu.

U slova *Rodel* bylo značné množství materiálu vyškrtáno hlavně z toho důvodu, že bylo ve spojení s předložkou. Tato předložka nám dala najevo fakt, že se nejedná o slovo v nominativu, proto muselo být z analýzy vyškrtáno. Dle analýzy můžeme potvrdit, jak uvádí Duden-Universalwörterbuch (2007), že *Rodel* jako femininum se nejvíce vyskytuje v Rakousku. *Rodel* jako maskulinum naopak v Německu.

*Paprika* v mužském rodě se nejvíce vyskytuje v Rakousku a to 28krát. V Německu bylo nalezeno pouze 9 výskytů. *Paprika* rodu ženského se nejvíce užívá v Německu, méně pak v Rakousku a Švýcarsku.

Substantivum *Python* jako maskulinum se nejvíce vyskytuje v Německu. Na druhé pozici se nachází Rakousko a třetí pozici obsadilo Švýcarsko. *Python* v rodě ženském se nejčastěji vyskytuje opět v Německu a sice 6krát. V Rakousku se vyskytuje 5krát a ve Švýcarsku byly nalezeny pouze 2 příspěvky.

Ačkoliv se metoda práce s korpusem řadí mezi novější, lze potvrdit, že práce s ním, je pro výzkum současného jazyka velice přínosná.

## 8. Literaturverzeichnis

- 1) BELICA, Cyril, 1995. Statistische Kollokationsanalyse und Clustering. Korpuslinguistische Analysemethoden. Mannheim: Institut für Deutsche Sprache. Mannheim.
- 2) DUDEN. Deutsches Universalwörterbuch, 2007. 6., überarb. und erw. Auflage. Mannheim: Dudenverlag. ISBN 978-3-411-05506-7.
- 3) DUDEN. Die Grammatik, 2005. Band 4. 7., völlig neu erarbeitete und erweiterte Aufl. Mannheim: Dudenverlag. ISBN 34-110-4047-5.
- 4) Etymologisches Wörterbuch des Deutschen, 2005. München. ISBN 3-423-32511-9.
- 5) HELBIG, Gerhard a Joachim BUSCHA, 2001. Deutsche Grammatik: Ein Handbuch für den Ausländerunterricht. Berlin: Langenscheidt. ISBN 3-468-49493-9.
- 6) LANGENSCHIEDT. Großwörterbuch Deutsch als Fremdsprache, 2010. Langenscheidt KG, Berlin und München. ISBN 978-3-468-49038-5 .
- 7) LEMNITZER, Lothar a Heike ZINSMEISTER, 2010. Korpuslinguistik: eine Einführung. 2., durchgesehene und aktualisierte Aufl. Tübingen: Narr. ISBN 978-382-3365-556.
- 8) LENZ, Susanne, 2000. Korpuslinguistik. Tübingen: Groos. ISBN 38-727-6848-4.
- 9) POVEJŠIL, Jaromír, 1999. Mluvnice současné němčiny. Praha: Academia. ISBN 80-200-0076-3.
- 10) SCHERER, Carmen, 2006. Korpuslinguistik. Heidelberg: Winter. ISBN 38-253-5164-5.
- 11) ŠULC, Michal, 1999. Korpusová lingvistika. Praha: Karolinum. ISBN 80-718-4847-6.

### Internetquellen

<http://www1.ids-mannheim.de/start/>; [zit. am 10.7.2014 ]

<http://www.ids-mannheim.de/cosmas2/projekt/entwicklung/geschichte.html> ;  
[zit. am 10.7.2014]

<http://www.ids-mannheim.de/cosmas2/uebersicht.html> ; [zit. am 10.7.2014]

<http://www.ids-mannheim.de/cosmas2/uebersicht.html> ; [zit. am 10.7.2014]

<http://www1.ids-mannheim.de/kl/projekte/korpora.html> ;[ zit. am 10.7.2014]

<http://www1.ids-mannheim.de/kl/projekte/korpora.html> ;[ zit. am 10.7.2014]