



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
Biskuit, Keks, Marzipan und *Nougat*

The grammatical gender variants of German substantives.
A case study using the examples of *Biskuit, Keks, Marzipan* and *Nougat*
corroborated by corpus

Vypracoval: Markéta Machová
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 25. 6. 2014

Poděkování

Úvodem své práce bych chtěla především poděkovat paní Mgr. Janě Kusové, PhD. za odborné vedení mé práce. Na základě jejích rad a připomínek mohla být tato práce dokončena.

Anotace

Rodová variabilita německých substantiv. Případová korpusová studie provedena na příkladu substantiv *Biskuit*, *Keks*, *Marzipan* a *Nougat*.

Autorka se v práci zabývá nejnovějšími tendencemi ve vývoji morfologického systému německých substantiv. Za pomoci korpusu *DeReKo* provádí analýzu kolísání rodu u slov *Biskuit*, *Keks*, *Marzipan* a *Nougat*, přičemž vztahuje svoji pozornost především na jeho diatopickou a sémantickou distribuci. Ve své práci se soustředí mimo jiné na etymologii daných slov. Výsledky analýzy porovnává s údaji, které nám poskytují vybrané slovníky.

V teoretické části se věnuje historii, metodám a potenciálu korpusové lingvistiky. V praktické části provádí analýzu výše uvedených substantiv.

Annotation

The grammatical gender variants of German substantivs. A case study using the examples of *Biskuit*, *Keks*, *Marzipan* and *Nougat* corroborated by corpus.

This thesis focuses on the newest trends in development of morphological system of German nouns. The author analyses four substantivs *Biskuit*, *Keks*, *Marzipan* and *Nougat* on the basis of *DeReKo*. The attention is drawn to the regional and semantic distribution of this variants. Furthermore, the thesis deals with etymology of the loan words. The results of the corpus analysis are compared with data in certain dictionaries.

The theoretical part relates to the methods of corpus linguistic and stress out the advantages of work with the corpus. The practical part includes the analysis of above mentioned substantivs.

Inhalt

1	Einleitung	7
2	Zwei Forschertypen	8
3	Empirismus wie eine Basis für die Korpuslinguistik	9
3.1	Methodische Grundlagen	9
3.2	Methoden der Sprachwissenschaft	10
4	Was ist ein Korpus?	11
4.1	Ist das World Wide Web ein Korpus?	12
4.2	Wie ist ein Korpus aufgebaut?.....	13
4.3	Kriterien für ein Korpus.....	14
4.4	Die angewandten Korpora	15
5	Geschichte der Korpuslinguistik	17
6	Institut für Deutsche Sprache	19
6.1	Das Deutsche Referenzkorpus	19
6.2	Cosmas II (Corpus Search, Management and Analysis System)	21
7	Substantiv im Deutschen	22
7.1	Genus	23
7.2	Genussystem im Deutschen.....	23
7.3	Funktion des Genus.....	24
7.4	Genuszuweisung.....	25
7.5	Genussschwankung und -wechsel	29
8	Analyse	30
8.1	Einleitung.....	30
8.2	Angewandte Regeln für die Genuszuweisung.....	30
8.3	Etymologie der ausgewählten Wörter	33
8.4	Was steht in Wörterbüchern?	36
8.5	Korpusanalyse in Bezug auf semantische und diatopische Aspekte	40
9	Schlusswort	47
10	Resumé	50
11	Literaturverzeichnis	52

1 Einleitung

Die Autorin beschäftigt sich mit dem Thema „Genusvariation deutscher Substantive. Eine korpusgestützte Fallstudie am Beispiel von *Biskuit*, *Keks*, *Marzipan* und *Nougat*.“ Die ausgewählten Substantive verfügen über zwei akzeptierte Genera. Mithilfe der Korpusanalyse wird der Genusgebrauch aus der diatopischen und semantischen Sicht untersucht. Da die Wörter aus Fremdsprachen übernommen wurden, wird die Aufmerksamkeit auch auf die Etymologie der Wörter gezielt. Aufgrund der Analyse wird überprüft, ob der aktuelle Sprachgebrauch in deutschen Wörterbüchern reflektiert wird.

Als erstes wird die Korpuslinguistik vorgestellt. Im Vordergrund stehen die Ziele und das Potenzial der Korpuslinguistik. Diese relativ junge linguistische Disziplin hat eigene methodische Grundlagen. Um eine Sprache zu erforschen, arbeiten die empirisch orientierten Korpuslinguisten mit mündlichen und schriftlichen Äußerungen. Die Äußerungen bilden ein Korpus. Die Arbeit beschäftigt sich auch damit, unter welchen Kriterien die Textproben für ein Korpus ausgewählt werden, wie das Korpus aufgebaut wird und wozu überhaupt die Korpora beitragen. Für die Analyse benutzt man das Deutsche Referenzkorpus, dem die Aufmerksamkeit in dem theoretischen Abschnitt gewidmet wird.

Daran knüpft das nächste bedeutende Kapitel an, in dem sich die Autorin mit dem deutschen Substantiv beschäftigt, nämlich mit der grammatischen Kategorie *Genus*. In diesem Teil wird erklärt, wie das Genusystem im Deutschen funktioniert und wozu das Genus dient. Es werden auch Prinzipien für die Genuszuweisung erwähnt, die in dem praktischen Teil auch angewandt werden können.

Anschließend kommt der praktische Teil, in dem eine Analyse durchgeführt wird. Die Autorin verfährt aufgrund der theoretischen Kenntnisse, die sie im theoretischen Teil gewann. Sie arbeitet mit dem Deutschen Referenzkorpus, das ihr die Angaben für die Analyse gewährt. Mithilfe der extrahierten Belege wird bestätigt oder abgelehnt, was in dem theoretischen Teil steht. Die Ergebnisse aus dem Korpus werden mit Wörterbüchern verglichen.

2 Zwei Forschertypen

Um Korpuslinguistik zu erklären, ist es angebracht, zwei Forschertypen zu erwähnen.

Wie Lemnitzer/Zinsmeister (2010, S. 6 - 7) in ihrem Buch vermitteln, unterscheiden die Sprachwissenschaftskreise zwischen zwei Typen, und zwar zwischen dem „Denker“ und dem „Beobachter“. Der Denker fasst die Forschung wie ein langandauerndes Nachdenken auf, das zu einer Sprachtheorie führt. Anhand der Beispiele, die in seinem Kopf entstehen und die sich infolgedessen nur auf seine eigene Sprachkompetenz beziehen, nimmt er eine Sprachtheorie an oder lehnt sie ab. Was bei dieser Art des Forschens als negativ festgestellt wird, ist die Tatsache, dass dieser Zugang zur Sprachforschung wenig mit dem alltäglichen Sprachgebrauch zu tun hat. „Der Beobachter“ hingegen sieht sich die Sprache als etwas Lebendes an. Er sammelt möglichst große Anzahl von Sprachdaten, von denen die Theorien aufgrund seines Beobachtens abgeleitet werden. Seine Hypothesen werden durch immer neue Daten bestätigt oder verworfen. Beim „Denker“ spielt die Universalgrammatik eine Rolle, die allen Menschen gemeinsam sein soll.

Lemnitzer und Zinsmeister (2010, S. 6 - 7) bezeichnen den „Denker“ als jemanden, der seinen Sessel nur selten verlassen muss. Er bewegt sich dann auf einer theoretischen Ebene. Auf der anderen Seite befasst sich der „Beobachter“ mit der alltäglichen Sprachverwendung und der Vermittlung dieses Sprachgebrauchs. Er strebt nach einer Beschreibung der Sprache und muss deshalb häufig in die Lexikographie und den Sprachunterricht gehen. Beobachten und Beschreiben sind zugleich auch Basis für die Korpuslinguistik.

3 Empirismus wie eine Basis für die Korpuslinguistik

Es gibt zwei Ansätze, und zwar einen theoretischen und einen empirischen. Sie sind parallel zum „Denker“ und „Beobachter“ zu verstehen. Scherer weist z.B. darauf hin, ob – wie immer wieder von Wissenschaftlern und Sprachbenutzern diskutiert wird – nach der Präposition „wegen“ der Dativ oder der Genitiv richtigerweise angewendet werden soll. Weder Wissenschaftler noch Sprachbenutzer können diese Frage anhand bestimmter sprachlicher Regel beantworten, welcher Fall richtig sein soll. Und darum scheitert der theoretische Ansatz, es muss der empirische in Betracht kommen. Die Empiristen arbeiten mit der Vorkommenshäufigkeit und der Sprachgebrauch ist für sie bestimmend (Scherer, 2006, S. 1).

3.1 Methodische Grundlagen

Es wird zwischen zwei methodischen Grundlagen differenziert. Der korpuslinguistischen Forschung wurde die empirische Grundlage zugewiesen und der theoretischen Forschung die rationalistische Grundlage (Lemnitzer, Zinsmeister, 2010, S. 15).

Die zwei angewandten Methoden sind die „Induktion“, die mit dem Empirismus verbunden wird, und die „Deduktion“, die dem Rationalismus zugeordnet wird (Lemnitzer, Zinsmeister, 2010, S. 18).

Mit Induktion wird *„der Übergang vom Besonderen zum Allgemeinen, Schließen von einzelnen Beobachtungen auf Gesetzesaussagen und Möglichkeit der Widerlegung von Gesetzesaussagen durch Beobachtungen [...]“* gemeint. Dagegen weist die Deduktion die folgende Charakteristik an: *„Übergang vom Allgemeinen zum Besonderen, Schluss von Prinzipien und Axiomen auf Regeln und Möglichkeit der Überprüfung der Gültigkeit dieser Regeln durch Beobachtungen [...]“* (Lemnitzer, Zinsmeister, 2010, S. 18 - 19).

3.2 Methoden der Sprachwissenschaft

Wenn man den Sprachgebrauch reflektieren will, dann lässt sich neben dem Mittel des Experiments und der Befragung auch mit authentischen Sprachdaten arbeiten. Diese Daten werden gesammelt und bilden ein Korpus. Um ein Experiment durchzuführen, kann man Versuchspersonen Sätze vervollständigen lassen und davon ableiten, welche Form sie benutzen. Eine andere Möglichkeit ist die Befragung. Die Muttersprachler können befragt werden, welche Form sie bevorzugen. Die Methode des Experiments ist sehr aufwändig und nicht objektiv im Zusammenhang mit einem Mangel an spontanem Sprachgebrauch. Dazu muss man auch die Zufälligkeit der Ergebnisse zur Kenntnis nehmen, wenn sie aus statistischer Sicht eine Rolle spielen. Darum reicht es in manchen Fällen nicht aus, eine Zahl von z.B. zehn Menschen zu befragen. In diesem Fall ist Arbeit mit einer viel breiteren Sammlung nötig. Das ermöglicht ein Textkorpus (Scherer, 2006, S. 2 - 3).

Im Rahmen der Korpuslinguistik unterscheidet man „korpusgestützte“ und „korpusbasierte“ Linguistik. Die Theoretiker können von der empiristischen Einstellung profitieren (Lemnitzer, Zinsmeister, 2010 S. 7). Es ist zu erwähnen, dass korpusgestützte Linguistik als ein Kompromiss zwischen Empiristen und Rationalisten bezeichnet werden kann. Mithilfe eines Korpus werden theoretische Aussagen überprüft. Für die korpusbasierte Linguistik stellt das Korpus die primäre Quelle der Erkenntnis. Aufgrund der Beobachtungen werden Regeln angenommen, verändert oder abgelehnt (Lemnitzer, Zinsmeister, 2010, S. 19).

4 Was ist ein Korpus?

Man kann ein Korpus wie eine Sammlung schriftlicher oder gesprochener Äußerungen vorstellen (Lemnitzer, Zinsmeister, 2010, S. 8). Weil die Korpuslinguistik die technische Entwicklung nutzen sollte, wie Lenz (1999, S. 1) in ihrem Buch behauptet, wurde es dadurch ermöglicht, die Daten des Korpus im Computer zu speichern und dann die Daten einfach zugänglich zu machen (Lemnitzer, Zinsmeister, 2010, S. 8). Das bedeutet nicht, dass früher keine Korpuslinguistik existiert hätte. Im Gegenteil. Ein Korpus war jedoch nur in Form von Papier oder Lochkarten vorhanden. Grundsätzlich kommt bei beiden Arten der Datensammlung dasselbe heraus. Nur können die computergespeicherten Informationen schneller abgerufen werden (Scherer, 2006, S. 5).

Ein Korpus ist keine chaotische Sammlung zufälliger Texte. Alle Texte werden nach bestimmten Kriterien ausgewählt. Wenn man Texte sagt, dann sind damit nicht nur schriftliche, sondern auch mündliche Äußerungen gemeint. Diese Äußerungen können Briefe, Zeitungsartikel sowie Radiosendungen, Vorträge und - für jemand überraschend - auch ein Gespräch am Mittagstisch sein. Man spricht von Primärdaten (Scherer, 2006, S. 3). Diese Äußerungen werden als ganze Texte oder mindestens große Textausschnitte im Korpus gesammelt. Sehr häufig, aber nicht in der Regel, sind diese für einen Gegenstand auch repräsentativ, linguistisch erläutert und durch Metadaten ergänzt (Lemnitzer, Zinsmeister, 2010, S. 40). Metadaten sagen etwas mehr von den gewonnenen Daten aus, z.B. von dem Trägermedium, der Art der Kodierung, den Autoren oder Zeit der Entstehung. Sie bieten einfach mehr Informationen über die Daten an (Lemnitzer, Zinsmeister, 2010, S. 46).

Die Annotation ist auf der korpuslinguistischen Ebene auch wichtig. Sie bereichert die Primärdaten eines Korpus. Das führt zu effizienteren Korpusuntersuchungen. Wortarten, syntaktische Phrasen oder grammatische Funktionen werden interpretiert. Die einzelnen Interpretationen bauen oft

aufeinander auf. Die Wiederverwendbarkeit ist ein Aspekt bei den annotierten Korpora. Ein anderer ist die Multifunktionalität. Diese Quelle kann auch in anderen Bereichen eingesetzt werden. Dazu gehört z.B. eine Erstellung eines Lexikons (Lemnitzer, Zinsmeister, 2010, S. 61 – 62). Mithilfe der annotierten Korpora werden die impliziten Informationen explizit gemacht. Die Annotation bringt alle möglichen Merkmale über verschiedene Phänomene, die markiert sind. Man unterscheidet zwischen der prosodischen, morphologischen, syntaktischen und semantischen Annotation (Scherer, 2006, S. 22).

Es ist auch zu betonen, dass es zahlreiche sprachliche Varietäten gibt, die von Zeit, Raum, Sprechergruppe oder Kommunikationssituation abhängig sind, und deshalb muss es entschlossen werden, ob man eine Sprache allgemein untersuchen will oder nur eine einzelne Varietät (Scherer, 2010, S. 4).

4.1 Ist das World Wide Web ein Korpus?

Wie ist es mit dem World Wide Web? Das Korpus ist doch eine Sammlung von authentischen Texten und es scheint, dass diese Funktion das Internet auch erfüllt. Könnten wir den Begriff *Korpus* auch dem Internet zuweisen? Manche Korpuslinguisten stimmen dem zu. Das Internet ist voll von Texten. Die riesige Textmenge von allen möglichen Sprachen ist aber unübersichtlich. Ein Problem dabei ist, die vielen Sprachen auseinander zu halten und die Texte mit Angaben über die Entstehung zu versehen. Metadaten, die einen Bestandteil von größeren Korpora darstellen, sind im Internet leider nicht zu finden. Ein Risiko im Internet ist, wenn Autoren von Texten nicht Muttersprachler sind und wenn es um seltene Konstruktionen oder Beispiele geht (Lemnitzer, Zinsmeister, 2010, S. 43 - 44). Scherer (2006, S. 76) belegt das Problem mit einem Beispiel: Englischer Terminus *superproportional* wurde als *überproportional* übersetzt. In Wirklichkeit heißt *überproportional* im Englischen *disproportionate*.

Scherer (2006, S. 74) findet ein Defizit im Bereich der Repräsentativität oder der Beständigkeit. Diese Kriterien kann das Internet überhaupt nicht erfüllen, infolgedessen würde die Autorin das Internet nicht als Korpus bezeichnen.

4.2 Wie ist ein Korpus aufgebaut?

Wie es schon erwähnt wurde, kann nicht jede Sammlung von Texten als Korpus bezeichnet werden. Seine Größe, sein Inhalt sowie Beständigkeit und Repräsentativität müssen in Frage gestellt werden. Man muss entscheiden, ob man eine Sprache in Teilen oder in ihrem gesamten Umfang untersuchen will. Korpora haben einen Zweck, und zwar als Ausschnitt einer sprachlichen Gesamtheit zu dienen. Einen Teil kann z.B. die Jugendsprache darstellen. Ein Korpus der Standardsprache sollte alle Bereiche der Standardsprache enthalten. Wenn dieses Kriterium erfüllt wird, darf man ein solches Korpus repräsentativ nennen (Scherer, 2006, S. 5).

Wo gewinnt man die Daten, die das Korpus bilden und die Kriterien dafür erfüllen? Die schriftlichen Texte werden digitalisiert oder manuell eingegeben. Eine gute Quelle ist, wie schon oben erwähnt, das Internet. Diese Daten müssen aber noch weiter bearbeitet werden, um in einem Korpus eingeordnet werden zu können. Eine Rolle spielt auch das Urheberrecht. Dann kommt es zur Kodierung von Daten, häufig via ein Betriebssystem UNICODE – dadurch werden die Angaben in den Computer eingetragen. Schließlich, wenn die Primärdaten gesammelt sind, werden die Metadaten und die Erläuterungen (Annotationen) bearbeitet (Lemnitzer, Zinsmeister, 2010, S. 57 - 58).

Um ein ausgewogenes Korpus zu erstellen, ist ein Zusammenspiel von externen und internen Kriterien sehr wichtig. Zu den externen Kriterien gehört die Anzahl der beteiligten Personen, der Grad der Mündlichkeit und Schriftlichkeit und natürlich die Beschreibung der Situation (Lemnitzer, Zinsmeister, 2010, S. 52). Die internen Kriterien beziehen sich auf zwei in von Sprachwissenschaftlern Interesse stehenden Terminen: die Okkurrenz und die Rekurrenz (Teubert, 1998, S. 115).

4.3 Kriterien für ein Korpus

Teubert (1998) liegt Grundprinzipien vor. Das Korpus sollte möglichst groß im Hinblick auf technologische Möglichkeiten sein, vielfältig angesichts der Textproben sein, in Genres klassifiziert sein und Quellenangaben beinhalten. Ein ursprüngliches Kriterium war auch die gleiche Größe von den Textproben. Das wurde umstritten. Sinclair macht darauf aufmerksam, dass die Proben von gleicher Größe nur Textfragmente sein können, die aus dem Kontext genommen wurden.

Die Repräsentativität eines Korpus spielt eine signifikante Rolle beim Aufbau. Es ist unvermeidbar die Grundgesamtheit klarzumachen. Man muss definiert, wer zu der bestimmte Sprachgruppe, die untersucht werden soll, gehört. Ein Beispiel ist zum Beispiel die Jugendsprache. Um möglichst zuverlässige Informationen zu gewähren, sollte erklärt werden, wer überhaupt ein Jugendlicher ist. (Scherer, 2006, S. 5). Lemnitzer und Zinsmeister (2010, S.51) erwähnt, dass die Grundgesamtheit nicht einwandfrei definiert werden kann. Wenn wir Sprache von alten Menschen erforschen wollen, müssen wir wissen, wer genau zu dieser Gruppe gehört. Noch dazu ist die Kommunikation von der Sprachsituation abhängig, die auch berücksichtigt werden muss. Nur eine intuitive Vorstellung von Sprache der alten Menschen ist leider nicht ausreichend im Fall des Korpusaufbaus. Trotzdem kann intuitives Wissen bei der Kriterienstellung teilweise angewandt werden (Scherer, 2006, S. 6).

Was die Größe betrifft, war eine Million Textwörter lange Zeit die Standardgröße für Korpora. Heutzutage, wenn der technische Fortschritt das erlaubt, wuchs die Textmenge. Das Institut für Deutsche Sprache in Mannheim archiviert zahlreiche Korpora im Umfang von insgesamt über zwei Milliarden Textwörtern. Es ist keine Regel, dass je größer, desto besser das Korpus ist. Auch Korpora mit einer geringeren Anzahl von Wörtern können zuverlässige Aussage über den Sprachgebrauch bringen. Besonders wenn es z. B. um die sog. Funktionswörter geht, die in fast jedem Text auftauchen, sind die kleinen Korpora ausreichend (Scherer, 2006, S. 7).

Ein anderes Kriterium sind Textproben. Die sollten vielfältige Inhalte gewähren, um genügend repräsentativ zu sein. Es muss auch die Größe von angenommen Texten in Frage kommen. Die Textproben sind nur Teile einer Sprachgesamtheit. Es kann entweder der Volltext in den Korpus eingeordnet werden oder nur eine bestimmte Zahl von Textwörtern. Bei langen Werken wie z.B. bei Romanen sind die Proben nicht so unterschiedlich wie bei Telefongesprächen oder Witzen. Das heißt, dass im Fall der langen Romane nicht Volltexte aufgenommen werden. Was auch zur Kenntnis genommen werden soll, ist der Inhalt. Es gibt viele Fragen, welche Textsorten und Sprechergruppen und zu welchen Anteilen usw. umfasst werden sollen, damit das Korpus möglichst repräsentativ ist. Die Metadaten bringen die Informationen, die immer nachvollziehbar sein sollten (Scherer, 2006, S. 8 – 9).

4.4 Die angewandten Korpora

Korpora werden auch in anderen linguistischen Disziplinen angewandt. Der Grund dafür ist die Tatsache, dass ein Korpus auch Informationen über die Vorkommenshäufigkeit und Verwendung von Wörtern, grammatische Kategorien und andere sprachliche Einheiten enthält. Wo die sprachliche Kompetenz eines Sprechers nicht ausreichend ist, lässt sich eine Sprachtheorie anhand eines Korpus überprüfen. Korpora dienen zu verschiedenen Zwecken (Scherer, 2006, S. 10).

Arten von Korpora nach Scherer (2006):

Speichermedium	→	computerisbares Korpus	vs.	nicht computerisbares Korpus
Hierarchie	→	Gesamtkorpus	vs.	Teilkorpus
Vollständigkeit der Texte	→	Volltextkorpus	vs.	Probenkorpus
Abgeschlossenheit	→	statisches Korpus	vs.	Monitorkorpus
Aufbereitung des Korpus	→	annotiertes Korpus	vs.	nicht annotiertes Korpus
Sprachmedium	→	Korpus der geschriebenen Sprache	vs.	Korpus der gesprochenen Sprache
zeitlicher Bezug	→	Korpus der Gegenwartssprache	vs.	historisches Korpus
Geltungsbereich des Korpus	→	Referenzkorpus	vs.	Spezialkorpus
Anzahl Sprachen	→	einsprachiges Korpus	vs.	mehrsprachiges Korpus

Abbildung 1: Arten von Korpora

Wo kann man die Korpora einsetzen?

- Die Erforschung von sprachlichen Strukturen und Varietäten;
- Die Erstellung von Wörterbüchern (Lexikografie);
- Die Erstellung von Grammatiken (Grammatikografie);
- Der Fremdsprachunterricht;
- Die Übersetzung;
- Die Computerlinguistik.

Besonders bei der Herstellung von Wörterbüchern und Grammatiken wächst die Popularität der Korpora. Ein Vorteil ist die Authentizität von Sprachdaten, die in einem Korpus vorkommen. Diese Sprachdaten stehen Lexikografen und anderen Linguisten für die nächste Bearbeitung zur Verfügung. Anhand der Korpora kann man auch Kontextinformationen gewinnen. Daneben kann man auch feststellen, mit welchen Wörtern ein Wort auftritt. Grammatiken sind auf einer ähnlichen korpusbasierten Grundlage aufgebaut. Bestimmte Programme bemühen sich darum, grammatische Strukturen im Korpus zu identifizieren. Diese Ergebnisse bieten wieder eine Überprüfung von Grammatikbeschreibungen und in manchen Fällen auch neue Spracherkenntnisse an (Scherer, 2006, S. 11).

Die einzelnen Varietäten, die nicht jedem bekannt sind, werden auch aufgrund eines Korpus erforscht. Bemerkenswert ist die Verwendung der Korpora beim Fremdsprachenlernen. Eine große Bedeutung haben Korpora bei Textverfassungen, Übersetzungen und insgesamt beim Erwerb einer Fremdsprache. Korpora sind besonders behilflich bei Entwicklungen von Lernmaterialien. Zu erwähnen sind auch der Grundwortschatz und die Grundgrammatik. Sie identifizieren die sprachlichen Einheiten, mit denen die Schüler beim Lesen und Gespräch am häufigsten konfrontiert werden können. Mithilfe eines speziellen Lernerkorpus kann man auch auf verschiedene Schwerpunkte von Fremdsprachenlernen hinweisen. Diese Merkmale, zusammen noch mit der Frequenzrate, machen die Korpusverwendung sehr wertvoll (Scherer, 2006, S. 12 – 14).

5 Geschichte der Korpuslinguistik

Schon in früheren Zeiten beschäftigten sich die Menschen mit Sammeln von verschiedenen Texten. Im Interessenspunkt stand auch die Vorkommenshäufigkeit von den einzelnen Wörtern. Die damaligen Menschen besaßen zu ihrem Nachteil keinen Computer, der zweifellos einen riesigen Anteil an der heutigen Popularität von der Korpuslinguistik hat.¹

Manche teilen die korpuslinguistische Geschichte in zwei Epochen, und zwar in die vor Chomsky und die nach Chomsky, die als moderne Korpuslinguistik betrachtet wird. Es war schon in 1879, als der Kading-Korpus mit ungefähr 11 Millionen deutschen Wörtern entstand. Das Hauptziel von diesem Korpus waren die Abhandlungen über Rechtschreibregel. In der Welt begannen z.B. Franz Boas, Charles Fries oder Leonard Bloomfield mit dem Korpus zu arbeiten.²

Die sechziger Jahre waren für die Korpuslinguistik ganz bedeutend. Die ersten Korpora, die elektronisch zur Verfügung standen, bestanden aus einheitlich kodierten Textsammlungen. Man erwähnt Brown Corpus, das etwa eine Million Textwörter enthält. Es gehört zu den kleineren Korpora aber gleichzeitig zu den sehr signifikanten. In der gleichen Zeit von der Entstehung des Brown Corpus tauchte die Kritik auf die Korpusverwendung auf. Wahrscheinlich der bekannteste Kritiker ist Noam Chomsky (Lenz, 2006, S.6). Er ist der Gründer der generativen Grammatiker. Man kann sagen, dass er der Theoretiker ist, der „seinen Sessel nur selten verlässt“, wie schon am Anfang über die Denker/ Theoretiker gesagt wurde. Seiner Meinung nach lässt sich die Gesamtheit nur annähernd durch eine Anzahl der Äußerungen repräsentieren. Es gibt eine begrenzte Menge von Regeln, aber eine theoretisch unbegrenzte Menge von Äußerungen (Lemnitzer, Zinsmeister, 2010, S. 20). Chomsky

¹ Übersetzt von http://www.essex.ac.uk/linguistics/external/clmt/w3c/corpus_ling/content/history.html, zit. am 2.12.20013.

² Übersetzt von <http://www.cl2011.org.uk/history-of-corpus-linguistics.html>, zit. am 2.12.2013.

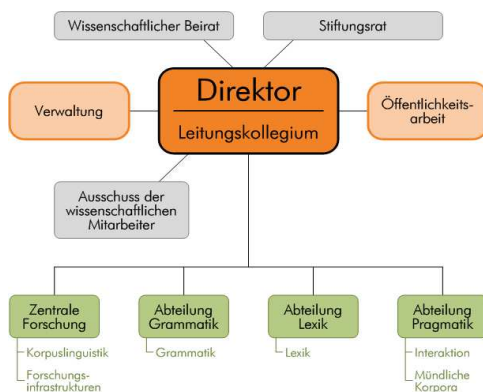
unterscheidet zwischen der „E-Sprache“ (*externalisierte Sprache*), die sich in der Menge der Äußerungereignisse widerspiegelt und der „I-Sprache (*internalisierte Sprache*)“, die die mentalen Zustände der Sprecher umfassen. „Eine Grammatik ist eine Theorie über diese I-Sprache und damit über die mentalen Zustände der Sprecher“ (Lemnitzer, Zinsmeister, 2010, S. 23). Mit diesen Aussagen von Chomsky geriet die empirische Arbeit mit Korpora für eine Zeit ins Abseits. Abgesehen davon ist die Korpuslinguistik aus mehreren Gründen nicht vollkommen verschwunden. Einige davon sind die bedeutungsbasierten Ansätze (verbunden mit „Kontext“, „Kotext“, Kollokation“ und „Kolligation“), die gegen den Formalismus stehen. Dazu noch weisen die Unterstützer von der Korpuslinguistik auf den Mangel an Empirie hin. Sie bezweifeln die Rechtfertigung im Fall der intuitionsbasierten Regeln (Lenz, 1999, S.7).

Nach dieser Ära voller Streitigkeiten kommt dank dem sich verbreitenden Computereinsatz der Aufschwung von den elektronischen Textsammlungen. Zu den in Deutschland bedeutendsten zählt man das LIMAS-Korpus und die Mannheimer Korpora des Instituts für Deutsche Sprache (Lenz, 1999, S. 6).

6 Institut für Deutsche Sprache

Das Institut siedelt seit 1964 in Mannheim. Es ist zwar unabhängig von Universitäten, aber in manchen Forschungsvorhaben arbeitet es neben Projektgruppen auch mit einzelnen Forschern aus diversen Hochschulen und Universitäten. Sein Ziel ist den gegenwärtigen Sprachgebrauch zu erforschen und zu dokumentieren. Alle Daten sind externen Wissenschaftlern auch vorhanden. Institut für Deutsche Sprache steht selbstverständlich Germanisten und anderen Interessierten zur Verfügung. Aufgrund der einflussreichen Position im linguistischen Bereich ist das Institut auch ein Ort, wo zahlreiche Tagungen veranstaltet werden.³

Organisatorische Gliederung des IDS entspricht dem folgenden Diagramm⁴:



6.1 Das Deutsche Referenzkorpus

Die Korpora geschriebener Sprache des IDS entwickelten sich kontinuierlich. Die Sammlung zählt schon 24 Milliarden Wörter. Die Texte kommen aus der Gegenwart oder der neueren Vergangenheit und sind zum größten Teil für die Benutzer zugänglich. Bei der Akquisition achtet man auf Umfang, Variabilität, Aktualität und natürlich auf Qualität.⁵

Das DeReKo wird in verschiedenen Bereichen eingesetzt. Es dient nicht nur den linguistischen Forschern oder Germanisten, sondern auch den Forschern aus anderen Fachbereichen wie z.B. Psychologie, Neurologie oder Medienwissenschaft.⁶

³ <http://www.ids-mannheim.de/org/aufgaben.html>; zit. am 15.4.2014.

⁴ <http://www.ids-mannheim.de/org/orga.html>; zit. am 15.4.2014.

⁵ <http://www1.ids-mannheim.de/kl/projekte/korpora/>; zit. am 24.4.2014.

⁶ <http://www1.ids-mannheim.de/kl/projekte/korpora/einsatz.html>; zit. am 24.4.2014.

Korpusakquisition ist ein ganz komplexer Prozess. Die Textauswahl stellt den ersten Schritt dar. Man muss die Kosten für die Beschaffung der Nutzungsrechte und auch Aufbereitung der Rohdaten zur Kenntnis nehmen. Dann kommt es zu der Textbeschaffung, in den meisten Fällen auch zur Konvertierung zu einem einheitlichen Format und zum Schluss die sehr wichtige Korpuspflege - die Texte müssen nämlich ständig überprüft werden. Wenn es nötig ist, werden durch zusätzliche Angaben erfüllt.⁷ Morphosyntaktische Annotationen sind im Korpus auch eingegliedert. Das hilft den Benutzern sich in den Texten einfach zu orientieren.⁸

⁷ <http://www1.ids-mannheim.de/kl/projekte/korpora/akquisition.html>; zit. am 24.4.2014.

⁸ <http://www1.ids-mannheim.de/kl/projekte/korpora/annotationen.html>; zit. am 24.4.2014.

6.2 Cosmas II (Corpus Search, Management and Analysis System)

Diese Applikation dient zum Recherchieren und zu der Analyse. Zur Zeit stehen ca. 8,9 Mrd. Wortformen zur Verfügung. Um ein Nutzer zu werden, muss man sich kostenlos registrieren.⁹

Scherer (2006, S. 82) behauptet, dass Cosmas II eine Vielzahl von Suchmöglichkeiten anbietet. Man muss nicht das ganze Wort eintragen, sondern nur z. B. den Wortanfang und das Wortende, und die Mitte wird dann von mehr beliebigen Buchstaben ersetzt, denn Cosmas II arbeitet auf der morphologischen Basis.

„Suchanfragen werden u.a. gebildet aus

- Wörtern, Teilwörtern, Wortgrundformen;
- Wortklassen (z.B. Verb, Artikel) und grammatikalischen Mustern;
- Angaben zu Wort- und Satzabstand;
- Angaben zu Textbereichen (z.B. Überschriften) und Position (z.B. erstes Wort eines Satzes).

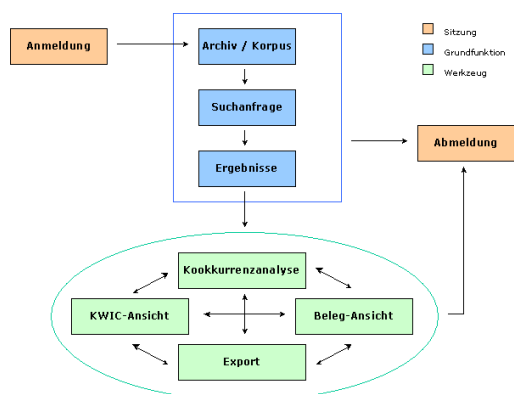
Ergebnisse können u.a.

- nach Entstehungszeit, Erscheinungsland und Thematik sortiert werden;
- durch Statistiken auf häufig verwendete Gebrauchsmuster hin analysiert werden.

Belege werden

- hinsichtlich ihrer Herkunft (Verlag, evtl. Autor, Entstehungszeit, Seitenangabe) dokumentiert;
- mit unterschiedlichen Kontextgrößen exportiert.“¹⁰

Die Korpusangaben werden nach folgendem Diagramm abgerufen¹¹:



⁹ <http://www.ids-mannheim.de/cosmas2/uebersicht.html>; zit. am 24.4.2014.

¹⁰ <http://www.ids-mannheim.de/cosmas2/projekt/nutzung/referenz.html>; zit. am 5.6. 2014.

¹¹ <http://www.ids-mannheim.de/cosmas2/projekt/hilfe/quick.html>; zit. am 5.6. 2014.

7 Substantiv im Deutschen

Das Substantiv besitzt eine starke Position, was die Vorkommenshäufigkeit betrifft. Mehr als die Hälfte des deutschen Wortschatzes wird durch Substantive/Hauptwörter repräsentiert. Die Großschreibung des Substantivs drückt die gesonderte Position in Grammatik und Wortschatz aus. Die Substantive sind den flektierenden Wortarten zugeordnet. In Bezug auf Substantive benutzt man den Terminus „Deklination“ (Hentschel, Vogel, 2009, S. 390).

Aus der morphologischen Sicht bestimmt man bei Substantiven drei grammatische Kategorien: das Genus (Maskulinum, Neutrum, Femininum), den Numerus (Singular, Plural, früher auch Dual) und den Kasus (Nominativ, Genitiv, Dativ, Akkusativ). Eine andere Gliederung wird unter semantischem Aspekt durchgeführt, dann können wir über die Konkreta und Abstrakta sprechen. Die Konkreta umfassen Eigennamen, Gattungsnamen, Sammelbezeichnungen und Stoffbezeichnungen. Dagegen sind die Abstrakta Wörter, die Eigenschaften, Vorgänge, Beziehungen und Maße bezeichnen (Bussman, 2008, S.697 – 698).

Semantisch darf man ein Substantiv auch als Wortbildungsprodukt in Betracht ziehen. Man unterscheidet zwischen „Ergebnis“ (*Verlust*), „Handlung“ (*Schreiben*), „Handelnder“ (*Unternehmer*) und „Mittel“ (*Nadel*). Häufig handelt es sich um Deverbative. Substantive sind artikelfähig. Der Artikel entspricht immer dem Substantiv im Genus, Numerus und Kasus. Der Artikel ist entweder bestimmt oder unbestimmt. Welchen wir benutzen, hängt vom Kontext ab. Etwas neu Angesprochenes, anderen Menschen oder anders Unbekanntes wird mithilfe des unbestimmten Artikels geäußert. Dagegen wird etwas, was für bekannt gehalten wird, durch den bestimmten Artikel markiert (Kessel, Reimann, 2010, S. 65-66).

7.1 Genus

In dieser Arbeit beschäftigt sich die Autorin mit der Genusvariation deutscher Substantive. Aus diesem Grund hat dieses Kapitel eine große Bedeutung für die folgenden Untersuchungen.

Das Genus ist dem Substantiv fest eingeordnet. Man unterscheidet zwischen dem grammatischen Genus und dem natürlichen Geschlecht. Zwischen den beiden lässt sich ganz häufig ein Zusammenhang beobachten. Man muss darauf achten, dass diese Beziehung nicht immer funktioniert. Es gibt Sprachen, die das Genus entweder nach semantischen Kriterien oder nach der externen Form bestimmen. Es ist auch möglich, dass im Rahmen einer Sprache beide Bestimmungsvorgehen vorkommen. Semantisch motivierte Systeme gehen von zahlreichen Kriterien aus, zu denen der Fakt gehört, ob das Substantiv „belebt“ oder „unbelebt“, „vernunftbegabt“ oder „nicht vernunftbegabt“ ist usw. (Hentschel, Weydt, 2003, S.158). Beim natürlichen Geschlecht ist es auch nicht eindeutig, dass wir drei Geschlechter (Maskulinum, Femininum und Neutrum) unbedingt benutzen müssen. Manche Sprachen haben gar kein Genus. Zu entscheiden, ob eine Sprache Genus hat oder nicht, sollten wir besonders auf zwei Aspekte achten: Pronomina (die Wahl des Pronomens ist mit dem Substantivgenus verbunden) und Kongruenz (Zusammenspiel von Adjektiven, Artikel mit Substantiven) (Hentschel, Weydt, 2003, S.158).

7.2 Genussystem im Deutschen

Deutsche Sprache verfügt über ein Dreiteilungssystem. Bei den deutschen Substantiven unterscheidet man zwischen Maskulinum, Femininum sowie Neutrum. Zu welchem von diesen drei Genera das bestimmte Substantiv gehört, wird aufgrund der semantischen, morphologischen wie auch phonologischen Kriterien entschieden. Wo der Zusammenhang zwischen Geschlecht und Genus scheitert, dürfen die anderen Kriterien angewandt werden. Bei den meisten Wörtern lässt sich das Genus ableiten (Hentschel, Weydt, 2003, S.159 – 160). Duden (2006, S. 153 – 154) macht darauf aufmerksam, dass am Substantiv selbst das Genus aber nicht direkt erkennbar

ist. Eine besondere Gruppe bilden Schwankungsfälle. In diesem Fall kommen die damit verbundenen regionalen und fachsprachlichen Aspekte in Frage. Duden beweist das auf zwei Beispielen: „*der Dispens*“ (österreichisch: die Dispens), „*der Raster*“ (fachsprachlich: das Raster). Keine besonderen Genusformen erscheinen im Plural. Im Plural benutzt man immer die gleiche Genusform, die mit *die* markiert wird. Es gibt auch Substantive, die nur im Plural vorkommen und bei denen kein Genus festgestellt werden kann.

7.3 Funktion des Genus

Römer (2006, S. 120 - 121) erklärt die Funktion der Genera ganz einfach. Die Genera, *maskulin*, *feminin* und *neutral*, werden durch die Artikel, Pronomen in ihren Flexionsformen markiert. Diese Genusmarkierung dient zur Übersicht im Rahmen der Bezüge zwischen Wörtern.

Es gibt viele Hypothesen, die ganz vielfältig sind, aber nicht ausreichend überprüft werden. Die häufig besprochenen Funktionen sind zum Beispiel die Identifizierung nominaler Konstituenten im Diskurs oder die Disambiguierung von Homonymen (Hentschel, Vogel, 2009, S. 150).

1. Dreiteilungssystem reduziert beim Hören die Anzahl der potentiellen Mitspieler.
2. Mithilfe der Pronomina kann ein Sprecher anaphorisch und kataphorisch auf Einheiten verweisen. Die Genuseigenschaften der Bezugsnomina ermöglichen Disambiguierungsprozess auf Seiten des Hörers.
3. Genusinformation zu Beginn der Nominalphrasen begrenzt die potenzielle Anzahl der Nomina
4. Bei der Klammerung hat das Genus auch eine große Bedeutung. Zu Beginn weisen der Artikel, der die Genusinformation trägt, und zum Ende das Nomen auf eine Informationseinheit hin (Hentschel, Vogel, 2009, S. 150 – 151).

7.4 Genuszuweisung

Die Genuszuweisung geht aus drei Bereichen aus. Diese Bereiche sind Semantik, Morphologie und Phonetik.

7.4.1 Semantische Prinzipien

Semantik bezieht sich auf die Bedeutung des Substantivs. In Betracht kommen Personen- und Sachbezeichnungen und Tierbezeichnungen.

7.4.1.1 Personenbezeichnungen

Das natürliche Geschlecht ist bei der Personenbezeichnung meistens entscheidend. Aus der Sicht der Beziehungen zwischen dem Genus und dem natürlichen Geschlecht darf man drei Substantivklassen ableiten. In der ersten Klasse ist ein Genus dem Substantiv zugewiesen, ohne offensichtliche Differenzierung zwischen den natürlichen Geschlechtern. Dazu gehören Wörter wie z.B. „*der Mensch*“, „*das Kind*“, „*die Person*“. Die nächste Klasse beinhaltet Substantive mit Genus, das von dem natürlichen Genus mitbestimmt ist wie z.B. „*die Frau*“, „*der Mann*“. Eine Ausnahme bildet die dritte Klasse mit abgeleiteten Nomen auf *-chen* oder *-lein*. Dabei sind die morphologischen Regeln stärker als die semantischen. Die andere Gruppe sind Substantive, die mit dem Suffix *-in* gebildet werden. Das ist eine feminine Ableitung. Bei substantivierten Adjektiven und Partizipien benutzt man entweder Maskulinum, Femininum oder Neutrum (Duden, 2006, S. 155 – 160).

7.4.1.2 Tierbezeichnungen

Bei den Nutztieren wird bei der generischen Bezeichnung neutrales Genus verwendet. Wenn es geschlechtsspezifisch ist, spielt das natürliche Geschlecht eine Schlüsselrolle. Jungtiere sind auch Neutra. Andere Tiere haben überwiegend den geschlechtsneutralen Gebrauch. Allgemein betrachtete Bezeichnungen (generisch) haben keine festen Regeln für die Genuszuweisung. Die Neigungen beziehen sich auf die Größe der Tierarten. Bei größeren, attraktiven und starken Tieren benutzt man

eher maskulin, bei kleineren und eher hässlichen feminin (Duden, 2006, S. 160). Bei Tiergattungen, mit denen weibliche als auch männliche Individuen verbunden sind, ist das Genus arbiträr. Gute Beispiele dafür sind „*der Wurm*“ oder „*das Krokodil*“. Keine offensichtliche semantische Motivation ist zu beobachten (Hentschel, Weydt, 2003, S. 160).

7.4.1.3 Sachbezeichnungen

Die einzelnen Wortfelder haben eine breite Skala von Regeln. Man kann bei manchen offensichtliche Tendenzen beobachten.

Wortfelder mit einer deutlichen Begriffshierarchie haben ein ganz stabiles System. Oberbegriffe kommen meistens im neutralen Genus vor. Im Vergleich mit den Oberbegriffen sind die Basisbegriffe sehr oft vom bestimmten Wortfeld abhängig. Obstorten sind meist feminin. Alkoholische und nicht alkoholische Getränke gehören zum maskulinen Genus, aber es gibt viele Ausnahmen wie z.B. „*das Wasser*“, „*das Bier*“, „*die Milch*“, „*die Limonade*“. Unterbegriffe richten sich nach dem Genus vom Basisbegriff, demzufolge übernehmen z.B. verschiedene Bierarten neutrales Genus (Duden, 2006, S. 160 – 161). Kurzwörter richten sich gewöhnlich nach den zugrunde liegenden Ausdrücken („*die Kripo*“ – „*die Kriminalpolizei*“). Auch diese Gruppe hat viele Sonderfälle („*das Kino*“ – „*der Kinematograph*“). Anderes Wortfeld sind artikellose Länder- und Ortschaftsnamen. Diese gehören zu Neutra. Länder mit dem Artikel sind oft feminin. Berge sind Maskulina („*der Brocken*“), aber bei Zusammensetzungen ist das Grundwort bestimmend („*das Matterhorn*“). Gebirgsnamen sind entweder feminin oder maskulin sowie pluralisch. Flüsse in Deutschland sind vorzugsweise feminin, aber manche auch maskulin. Bei den Flüssen außerhalb des deutschsprachigen Gebiets wiegt das maskuline Genus vor, ausgenommen diejenigen, die auf *-e* oder *-a* enden. Schiffsnamen sind feminin. Man kann auch ein ursprüngliches Genus benutzen, wenn es sich um einen Eigennamen mit einem Artikel handelt. Flugzeugbezeichnungen sind zum Femininum eingeordnet, nur erkennbare Zusammensetzungen übernehmen das Genus vom Grundwort. Einzelne Typen von Autos sind maskulin, wobei Motorräder feminin sind. Sehr interessante Gruppe bilden die Namen von Hotels, Kinos und

Gasthäusern. Die mit Personennamen verbundenen Bezeichnungen sind meist neutral. Bei anderen ist das Genus der Gattungsnamen entscheidend (Duden, 2006, S. 160 – 164).

7.4.2 Morphologische Prinzipien

Ein Wort kann aus mehreren Teilen bestehen, den sog. Morphemen. Das Genus richtet sich nach dem letzten Bestandteil. Das gilt für die Zusammensetzungen, die Ableitungen mit Präfixen und Ableitungen mit Suffixen. Ableitungen mit Präfixen haben eine Ausnahme und zwar, dass Sammelbezeichnungen mit *Ge-* Neutra („*das Geflügel*“ aber „*der Flügel*“) sind (Duden, 2006, S. 165).

Man stößt auf Probleme mit Suffixen, die nicht selbständig vorkommen und demzufolge nicht in Wörterbüchern eingetragen werden. Man muss darauf achten, welches Suffix angeknüpft ist, und davon ausgehen. Tabellen mit typischen Suffixen wurden in manchen Grammatiken dokumentiert. Suffixe **-nis, -sal und -tum** können zwei Genera repräsentieren. Diese Regeln können ganz einfach in Bezug auf Sachbezeichnungen angewandt werden. Demgegenüber hängen Personenbezeichnungen eher mit den semantischen Regeln eng zusammen (Duden, 2006, S. 164 – 168).

Hentschel und Vogel (2009) erwähnen noch Null-Substantivierung von Wörtern und Phrasen, denen das neutrale Genus zugeordnet ist. Es wird auf dem Beispiel „*das Laufen*“ gezeigt. Ein ähnlicher Fall sind Deverbative mit Nullsuffix (maskulin) und mit **-t** (feminin).

7.4.3 Phonologische Prinzipien

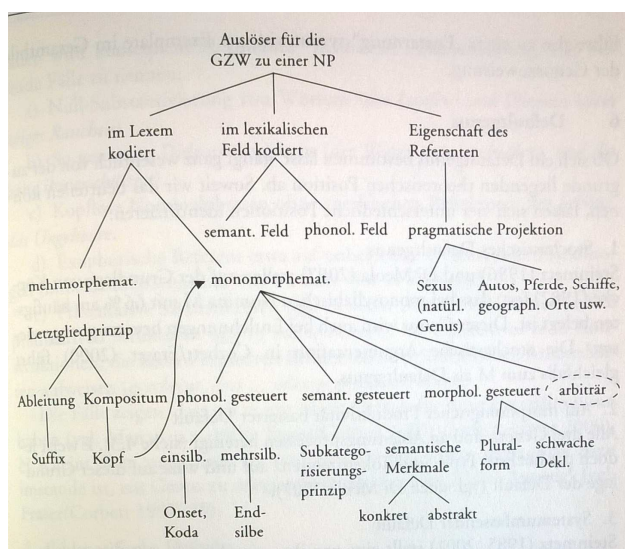
Bei der Genuszuweisung aufgrund der lautlichen Struktur des Wortes sind zahlreiche Tendenzen zu bemerken, aber es gibt keine feste Regel. Einsilbige Wörter, die auf **-ft** oder **-cht** enden, sind eher feminin. Einsilbige Wörter, die am Anfang oder am Ende eine Konsonantengruppe haben, gehören überwiegend zum maskulinen

Genus. Die einzelnen Tendenzen beeinflussen sich gegenseitig und es hängt davon ab, welche Regel in dem bestimmten Fall stärker ist (Duden, 2006, S. 168).

Obwohl die Prinzipien in Duden (2006) aufgelistet wurden, wird noch immer kontrovers diskutiert, ob es überhaupt diese Prinzipien der Genuszuweisung gibt. Die Theoretiker stellen sich die Frage, ob diese Prinzipien dann eher semantischer oder formalgrammatischer Natur sind. Es wurde geschätzt, dass bei 50% von den 1000 hochfrequenten deutschen Substantiven das Genus abgeleitet werden kann (Römer, 2006, S. 120). Zu den Substantiven, bei denen die Genera ganz einfach erklärbar sind, gehören

- Komposita, die das Genus des Grundworts haben;
- Substantive mit Suffixen;
- sexusmarkierte Substantive;
- Substantive auf –b sind nicht feminin (*der Dieb*);
- Entlehnte Fremdwörter, die das Genus eines assoziativ verbundenen Substantivs haben (*das Spiel*, *das Match*) (Römer, 2006, S. 120).

Die Genuszuweisung verläuft nach diesem Schema:



(übernommen von Hentschel, Vogel, 2009, S. 147).

7.5 Genussschwankung und -wechsel

Manche Substantive haben mehr als ein Genus. Das bezieht sich zum großen Teil auf die Sprachgeschichte. Der deutsche Wortschatz bilden auch Wörter, die aus anderen Sprachen übernommen wurden. Zu solchen Sprachen gehören auch die genuslosen. In diesem Fall spielen alle oben erwähnten Prinzipien für die Genuszuweisung eine Rolle, wobei häufig nicht klar ist, welcher der Faktoren sich endgültig durchgesetzt hat. Infolgedessen entsteht die Variabilität im Genusgebrauch. Eine andere Gruppe wird von den Substantiven, die aus den Sprachen mit Genus entlehnt wurden, geprägt. Bei diesen Wörtern werden Genera teilweise bewahrt, aber sie sind teilweise auch den deutschen Regeln untergeordnet. Bei den fachsprachlichen Entlehnungen kommt es gewöhnlich nicht zum Genuswechsel. Sonderfälle erscheinen, wenn die Begriffe in die Allgemeinsprache eingegliedert werden, dann wechselt man manchmal das Genus. Als unstabil wurden schon im Mittelhochdeutschen die maskulinen Sachbezeichnungen wahrgenommen, die eine Nominativform auf **-e** hatten. Heute gehören sie häufig zu Feminina (Duden, 2006, S. 169).

Bis heute benutzen verschiedene Mundarten ein früher übliches Genus. Das verursacht auch die Variabilität beim Genusgebrauch (Duden, 2006, S. 169 – 171).

Zu betonen ist auch die Tatsache, dass die Substantive im Deutschen, die über zwei Genera verfügen, zwei Hauptgruppen bilden. Falls die Form und die Bedeutung gleich sind, spricht man von der Schwankung des Genus. Wenn die Form jedoch gleich bleibt, die Bedeutung aber anders ist, spricht man über die Homonyme (Helbig, Buscha, 2001, S. 249). Engel (2009) behauptet, dass der Gebrauch von regionalen Aspekten abhängt, wobei ein einzelner Sprachbenutzer nur eine bestimmte Genusvariante benutzt.

8 Analyse

8.1 Einleitung

In diesem Teil befasst sich die Autorin mit dem Schwanken des Genus. Das Ziel ist die Theorie aufgrund der Korpusanalyse praktisch zu überprüfen. Vier Substantive, bei denen das Schwanken vorkommt, wurden ausgewählt. Mithilfe des Deutschen Referenzkorpus (DeReKo II), das durch das System Cosmas II elektronisch verfügbar ist, werden die Tendenzen im Genusgebrauch beschrieben.

Die vier ausgewählten Substantive *Biskuit*, *Keks*, *Marzipan* und *Nougat* werden in die Suchmaschine eingegeben. Man bemüht sich um die Analyse anhand der aktuellen Neigungen im Genusgebrauch. Bei allen verwendet man das doppelte Genus. Die Substantive sind entweder Neutra oder Maskulina. Unsere Aufgabe ist diese Behauptung zu bestätigen oder abzulehnen. Das Hauptziel bezieht sich auf die diatopischen und semantischen Aspekte, die zweifellos eine Rolle im Genusgebrauch spielen. Die Etymologie ist auch nicht wegzudenken. Die oben genannten Wörter wurden aus Fremdsprachen übernommen und das muss auch reflektiert werden.

8.2 Angewandte Regeln für die Genuszuweisung

Im theoretischen Teil wurde schon erwähnt, dass die Genuszuweisung verschiedene Umstände beeinflussen. Zu nennen sind die semantischen, die morphologischen und phonologischen Prinzipien. Im Fall der Substantive *Biskuit*, *Keks*, *Marzipan* und *Nougat*, kann man die morphologischen Prinzipien ausschließen. Die Wörter haben keine Präfixe und Suffixe und darum lässt sich das Genus nicht ableiten. Einfach gesagt, diese Wörter bestehen nur aus einem Teil, wobei keine morphologischen Regeln angewandt werden können. Man kann erwarten, dass die semantischen Prinzipien Einfluss haben können. Duden (2006) bietet eine breite Skala von Regeln. Keine Regel passt jedoch zu den ausgewählten Substantiven. Aufgrund der exportierten Angaben ist offensichtlich, dass es auch zu keiner Änderung in Wortbedeutung kommt. Es scheint, als ob der Genusgebrauch nur ein Zufallsprodukt wäre. Da muss man noch die phonologischen Prinzipien und auch die

Tatsache betrachten, dass diese Wörter aus dem Italienischen, Englischen und Französischen übernommen wurden und das Genussystem bei diesen Sprachen unterschiedlich ist.

Aus der phonologischen Sicht ist die lautliche Struktur bestimmend. Einsilbige Wörter, die auf –ft und –cht enden, sind häufig feminin. Diese Regel ist in unserem Fall nutzlos. Sehr interessant ist, dass die einsilbigen Wörter, die am Anfang oder am Ende eine Konsonantengruppe haben, gehören eher zum maskulinen Genus. „Keks“ schwankt ebenso zwischen zwei Genera. Das Maskulinum erscheint häufiger als das Neutrum. Hier gilt wahrscheinlich die oben genannte Regel. Es handelt sich um ein einsilbiges Wort mit der Konsonantengruppe an seinem Ende. Dieses Prinzip kann bei den anderen Substantiven leider nicht angewandt werden. Sie weisen keine Elemente aus, die von den Regeln unterstützt werden können. Die Regeln aus dem theoretischen Teil haben sich, außer den phonologischen Prinzipien im Fall des Substantivs „Keks“, hier nicht durchgesetzt. In Betracht kommen dann noch die einzelnen Genussysteme der Sprachen, von denen diese Wörter übernommen wurden. Um möglichst präzise Analyse durchzuführen, findet die Autorin unvermeidbar die Genusprinzipien der einzelnen Sprachen zu erklären.

Genussystem im Französischen

Französisch gehört zur romanischen Sprachenfamilie und verfügt über zwei Genera, maskulin und feminin. Die Genuszuweisung richtet sich nach dem grammatischen wie auch dem natürlichen Geschlecht. Das Genus wird mit Artikeln markiert. Bei Personenbezeichnungen steht Sexus bei der Kategorisierung im Vordergrund. Trotz der ganz großen Zahl der sexusneutralen Substantive hat das Französische kein Neutrum.¹²

¹² https://www.uni-due.de/imperia/md/content/prodaz/das_genus_in_verschiedenen_sprachen_alphabetisch_geordnet.pdf; zit. am 29.5. 2014.

Genussystem im Italienischen

Italienisch gehört zur romanischen Sprachfamilie und unterscheidet zwischen dem Maskulinum und dem Femininum. Ähnlich wie Französisch spielt grammatisches und natürliches Geschlecht eine Rolle. Genus und Numerus können dank den Endungen erkannt werden. Die *o*-Endung benutzt man für maskulin, die *a*-Endung für feminin und die *e*-Endung kann für beide stehen. Italienisch hat auch kein Neutrum.¹³

Genussystem im Englischen

Englisch gehört zu den westgermanischen Sprachen. Im Gegenteil zum Französischen und Italienischen unterscheidet Englisch zwischen drei Genera: maskulin, feminin und neutral. Es verfügt nur über das natürliche Geschlecht. Zu Besonderheiten gehört, dass Englisch kein Substantivgenus hat, sondern das Genus ist durch den Gebrauch von Pronomen markiert.¹⁴

¹³ https://www.uni-due.de/imperia/md/content/prodaz/das_genus_in_verschiedenen_sprachen_alphabetisch_geordnet.pdf; zit. am 29.5. 2014.

¹⁴ https://www.uni-due.de/imperia/md/content/prodaz/das_genus_in_verschiedenen_sprachen_alphabetisch_geordnet.pdf; zit. am 29.5. 2014.

8.3 Etymologie der ausgewählten Wörter

Die Etymologie bezieht sich auf die Herkunft der Wörter. Jedes Wort hat seine Geschichte und davon berichten die etymologischen Wörterbücher. Die Angaben geben uns Auskunft über die Entwicklung. Bei den Fremd- und Lehnwörtern wird die Entwicklung bis in die Ursprungssprache verfolgt.¹⁵ Es wurde schon darauf aufmerksam gemacht, dass sich die Sprache entwickelt. Manche Wörter wurden entlehnt. Es hängt von verschiedenen Umständen ab, ob sich die Prinzipien für die Genuszuweisung durchsetzen oder ob die Sprache auch das Genus aus der Ursprungssprache übernimmt. Zu beobachten sind auch regionale Varietäten, die eine einzige Möglichkeit bevorzugen.

Substantiv	Biskuit
Herkunft	Französisch
Genus in der Ursprungssprache	Italienisch maskulin → Französisch maskulin
Entwicklung	n. 'leichtes Gebäck'. Das im 17. Jh. ins Dt. Übernommene frz. <i>biscuit</i> verdrängt älteres, aus ital. <i>biscotto</i> 'Zwieback, Biskuit' entlehnte <i>spiscot</i> (13. Jh.), <i>biscott</i> (16. Jh.). Sowohl die frz. als auch die ital. Form gehen auf lat. <i>bis</i> 'zweimal' und lat. <i>coctum</i> , Part. Perf. von lat. <i>coquere</i> 'kochen, siedend, backen', zurück, bedeuten also 'zweimal Gebackenes, Zwieback' (vgl. mlat. <i>biscoctus</i> , <i>biscottus</i>) ¹⁶

Substantiv	Keks
Herkunft	Englisch
Genus in der Ursprungssprache	Neutral
Entwicklung	Keks m. n. 'trockenes Kleingebäck'. Der Plural engl. <i>cakes</i> 'Kuchen-, (süße) Gebäckstücke' wird im 19. Jh. ins Dt. übernommen (<i>Leibniz-Cakes</i> 1889). Wahrscheinlich ist engl. <i>cake</i> 'Kuchen, (süßes) Gebäck' eine Entlehnung aus gleichbed. anord. <i>kaka</i> ; vgl. isl. schwed. <i>kaka</i> , dän. <i>kage</i> , die im Ablaut zu dt. Kuchen (s. d.) stehen. Die Entlehnung <i>Cakes</i> wird zu <i>Keeks</i> (1905), dann zu <i>Keks</i> eingedeutscht. Vorschläge für eine Singularform <i>Keek</i> oder <i>Kek</i> werden nicht angenommen, vielmehr wird der Plural <i>Keks</i> singularisiert und eine neue Pluralform <i>Kekse</i> hinzugebildet (1911). ¹⁷

¹⁵ <http://www.duden.de/hilfe/herkunft>; zit. am 29.5. 2014.

¹⁶ <http://www.dwds.de/?qu=Biskuit>; zit. am 29.5. 2014.

¹⁷ <http://www.dwds.de/?view=1&qu=Keks>; zit. am 29.5. 2014.

Substantiv	Marzipan
Herkunft	Italienisch
Genus in der Ursprungssprache	Italienisch maskulin
Entwicklung	<p>Marzipan n. weiche Masse aus Mandeln und Zucker, seit dem Anfang des 16. Jhs. im Dt. bezeugt. Wort und Sache werden mit dem Italienhandel (ital. <i>marzapane</i>) übernommen. Die weitere Herkunft ist ungewiß. Eine (sehr hypothetische) Herleitung versucht KLUYVER in: ZfdWf. 6 (1904/05) 59 ff.: Arab. <i>mauṭabān</i> 'wer seinen Sessel nicht verläßt', eine Ableitung von arab. <i>waṭaba</i> 'still sitzen', wird vor 1200 bei südarabischen Händlern Bezeichnung für eine byzantinische Münze mit dem Bild eines auf dem Thron sitzenden Christus. Eine von den Venezianern geprägte Nachahmung (seit 1193) wird venezian. <i>mat(t)apān</i> genannt. In Syrien (1202) bezeichnet <i>marzapanus</i> eine auf importierte Waren erhobene Abgabe, in Venedig (13./14. Jh.) <i>marzapane</i> eine kleine Schachtel. Schließlich wird (im 14. Jh.) die Bezeichnung auf den Inhalt der Versandschachtel 'Teig aus Zucker und Mandeln, Süßigkeiten daraus' übertragen. Bei dieser Annahme bleibt jedoch zumindest die Lautentwicklung (zumal im Konsonantismus) unerklärt. Neuerdings wird auf die burmesische Stadt <i>Martaban</i> und von dort stammende Porzellangefäße verwiesen, vgl. CORTELAZZO/ZOLLI3,725. Der älteste dt. Beleg <i>marczapan</i> folgt der ital. Lautform (s. oben). Die zuerst Mitte des 16. Jhs. belegte, heute geläufige Form <i>Marzipan</i> beruht wohl auf einer volksetymologischen Deutung, die von mlat. *<i>Marci panis</i> 'Markusbrot' ausgeht.¹⁸</p>

Substantiv	Nougat
Herkunft	Französisch
Genus in der Ursprungssprache	Französisch maskulin
Entwicklung	<p>Nugat Nougat m. und n. 'Nuß- oder Mandelkonfekt', Übernahme (19. Jh.) von gleichbed. frz. <i>nougat</i>, aus prov. <i>nougat</i> 'aus Mandeln und Honig bereiteter Kuchen', aprov. <i>nogat</i> 'Ölkuchen, Rückstände ausgepreßter Nüsse', das auf vlat. *<i>nucātum</i> 'aus Nüssen Bereitetes' zurückgeht; zu lat. <i>nux</i> (Gen. <i>nucis</i>) 'Nuß'.¹⁹</p>

¹⁸ <http://www.dwds.de/?view=1&qu=Marzipan>; zit. am 29.5. 2014.

¹⁹ <http://www.dwds.de/?view=1&qu=Nougat>; zit. am 29.5. 2014.

Die Wörter *Biskuit*, *Keks*, *Marzipan* und *Nougat* wurden aus dem Französischen, Italienischen und Englischen übernommen. *Biskuit* stammt aus dem Italienischen, das später vom Französischen verdrängt wurde. Beide Sprachen unterscheiden zwischen maskulin und feminin und in Bezug auf das Wort *Biskuit* gehört es in beiden Fällen zum Maskulinum. Das Wort wurde im 17. Jahrhundert ins Deutsche entlehnt. *Keks* wurde aus dem Englischen übernommen. Englisch unterscheidet zwischen drei Genera und es verfügt nur über natürliches Geschlecht. *Keks* entstand aus dem Englischen Plural *cakes*, während der Singular *cake* neutral ist. Es wurde im 19. Jahrhundert ins Deutsche übernommen. *Marzipan* bereicherte den Deutschen Wortschatz schon im 16. Jahrhundert. Es wurde durch den Italienhandel in der Sprache integriert. Italienisch benutzt das Wort *marzapane* und es geht um ein Maskulinum. *Nougat* wurde im 19. Jahrhundert aus Französischen *nougat* übernommen. Im Französischen ist das Substantiv maskulin.

8.4 Was steht in Wörterbüchern?

Im theoretischen Teil wurde klargemacht, wo die Korpora angewandt werden können. Sie dienen zu verschiedenen Zwecken und daneben auch zur Erstellung von Wörterbüchern. Da die Lexikographie teilweise eine korpusgestützte Disziplin ist, will die Autorin einige Wörterbücher durchsuchen, ob sie den aktuellen Genusgebrauch reflektieren.

	Biskuit	Artikelwort
Duden	Bis kuit [bɪ'skvi:t, auch:...'kvi:t], das, auch: der; - [e]s, -s, auch: -e [frz. biscuit < afrz. Bescuit, unter Einfluss von ital. biscotto (↑ Biskotte) zu: cuire < lat. coquere = kochen, backen]: 1. Feingebäck aus Mehl, Eiern, Zucker. 2. kurz für ↑ Biskuitporzellan (Duden, 2007, S. 311).	das, der
Langenscheidt	Bis kuit ['bɪskvi:t, bɪs'kvi:t] <i>das/der; -s, -s;</i> ein leichtes Gebäck, das ohne Fett hergestellt wird: <i>ein Tortenboden aus B. K-Biskuitrolle, Biskuitteig</i> (LANGENSCHIEDT 2010, S. 229).	das, der
FIN Publishing	r, s Biskuit ((e)s s n. e) ['bɪskɪt]	der, das

	Keks	Artikelwort
Duden	Keks , der, seltener: das; - u. es, - u. -e, österr.: das; -, - [e] [engl. cakes, Pl. von: cake = Kuchen]: <ol style="list-style-type: none"> a) <o. Pl.> <i>trockenes, haltbares Kleingebäck</i>: K. backen, essen; eine Dose K.; b) Stück Keks (1 a): -e backen; einen Keks essen; eine Dose Kekse. (salopp) <i>Kopf</i>: sich den K. stoßen; * einen weichen K. haben (<i>nicht recht bei Verstand sein, verrückt sein</i>); jmdm. auf den K. gehen (↑ Nerv 3): das geht mir langsam auf den K. (Duden, 2007. S. 943). 	der, das (seltener, in Österreich)
Langenscheidt	Keks <i>der; -es, -e od bes österr. auch das; -, -(e)</i> ; ein kleines, flaches, haltbares Gebäck (das in Dosen od. Packungen verkauft wird): <i>Kekse u. Waffeln K- Keksdosen K- Butterkeks, Schokoladenkeks ID j-m auf den K. gehen</i> umg; j-n nervös machen, j-m auf die Nerven gehen NB: ↑ Plätzchen (LANGENSCHIEDT 2010, S. 627).	der, das (in Österr.)

FIN Publishing	r, s Kek(s, e)	der, das
----------------	----------------	----------

	Marzipan	Artikelwort
Duden	Mar zi pan [auch: ' - - -], das, österreich. , sonst selten: der; -s, -e [ital. Marzapane, H.u.] <i>weiche Masse aus fein gemahlenden Mandeln, Puderzucker u. Rosenwasser, aus der Süßigkeiten hergestellt werden</i> (Duden, 2007. S. 1117).	das, der (seltener)
Langenscheidt	Mar zi pan das der; -s, -e; mst sg; e-e weiche Masse aus Mandeln u. Puderzucker, aus der man Süßigkeiten macht K-Marzipanbrot, Marzipankartoffel, Marzipanschweinchen (LANGENSCHIEDT 2010, S. 735).	das, der
FIN Publishing	r, s Marzipan (s, e)	der, das

	Nougat	Artikelwort
Duden	Nu gat , Nougat, der od. das; -s, (Sorten:) -s [frz. nougat, über das provenz. u. Vlat. Zu lat. Nux (Gen. : nucis) = Nuss]: <i>aus fein zerkleinerten gerösteten Nüssen od. Mandeln, Zucker u. Kakao zubereitete Masse (als Süßware bzw. Als Füllung für Süßwaren)</i> (Duden, 2007, S. 1219).	der, das
Langenscheidt	Nu gat ['nu:gat] <i>der, das; -s; nur sg; e-e weiche, süße, braune Masse aus gemahlenden Nüssen, Zucker u. Kakao, mit der man oft Pralinen o. Ä. Füllt K- Nugatfüllung, Nugatmasse, Nugatschokolade K- Nussnugat</i> (LANGENSCHIEDT 2010, S. 807).	der, das
FIN Publishing	Keine Angaben	-

Drei verschiedene Wörterbücher wurden untersucht, um den Genusgebrauch bei den ausgewählten Wörtern zu erforschen. In allen drei Wörterbüchern sind die vier Substantive mit beiden Artikelwörtern *der* und *das* eingetragen. Nur *Nougat* erscheint im FIN PUBLISHING (2004) überhaupt nicht. Beim Suchen nach dem Wort *Nougat* wurde erfahren, dass in Wörterbüchern nicht nur *Nougat*, sondern auch *Nugat* steht. Das hat in Bezug auf den Genusgebrauch jedoch keinen Unterschied verursacht. *Nugat* als auch *Nougat* haben zwei Genera. Was das Wort *Biskuit* betrifft, geben uns Duden (2007) und Langenscheidt (2010) eine kurze Auskunft über häufigeres Vorkommen eines der Genera. Duden (2007) bestätigt, dass das

Artikelwort *das* in wenigeren Fällen erscheint. Den gleiche Bericht gibt uns auch Langenscheidt (2010), worin steht, dass das Artikelwort *das* eher im österreichischen Sprachgebiet benutzt wird. Duden (2007) betont auch die Präferenzen im Genusgebrauch beim Wort *Marzipan*. Seltener benutzt man maskulin.

Das Schwanken der Genera wurde in diesen Fällen bestätigt. Es wurde nicht erfahren, dass es mit verschiedenem Genus zur Änderungen in Bedeutung kommt. *Biskuit*, *Nougat* und *Marzipan* haben nur einzige Bedeutungen und das Schwanken zwischen zwei Genera haben offensichtlich keinen Einfluss auf die Bedeutung. Die Ausnahme ist *Keks*, der über zwei Bedeutungen verfügt. Eine der Bedeutung ist idiomatisiert und lehnt sich an das Wort *Kopf* an, wobei ausschließlich das Artikelwort *der* benutzt wird. Eines der Idiome ist „**einen weichen Keks haben**“, das uns sagt, dass jemand nicht recht bei Verstand ist. Anderes Idiom, das mit dem Wort *Keks* (nämlich Kopf) verbunden ist, ist „**jemandem auf den Keks gehen**“. Das sagt uns, dass uns etwas nervt.

Bei der Korpusanalyse wurden auch manche Proben exportiert, die die Redewendung „der Keks ist gegessen“ beinhalten. Diese Redewendungen haben zwar mit *Kopf* nichts zu tun, aber ihre Bedeutungen werden auch im Internet oft diskutiert. Bei Google-Searching bekam die Autorin bezüglich der Bedeutungen diese Antworten:

- es ist vorentschieden;
- hat sich erledigt;
- vorzeitig vorbei.²⁰

Die exportierten Proben beweisen den aktuellen Gebrauch und die Bedeutung selbst ist offenbar.

Herr Minister, ob die Mittel, die Sie beschrieben haben, mittelfristig ausreichend sind, das weiß ich nicht. **Der letzte Keks** ist noch nicht gegessen. Das hat gar nichts allein mit dem Urteil zu tun. Die Frage ist ja, welchen Aufgaben wir uns noch stellen müssen. Vor uns liegt ein sehr, sehr langer Weg. Momentan orientieren sich alle an der 35- % -Marke bis 2013. Das ist nur ein Richtwert. De facto gibt es ab August 2013 einen Rechtsanspruch. Darauf müssen wir uns einstellen. **(PSH/W17.00033 Protokoll der Sitzung des Parlaments Landtag Schleswig-Holstein am 18.11.2010. 33. Sitzung der 17. Wahlperiode 2009-2012. Plenarprotokoll, Kiel, 2010)**

²⁰ <http://www.gutefrage.net/frage/der-keks-ist-gegessen>; zit. am 7.6. 2014.

Der letzte Keks ist noch nicht gegessen sagt uns, dass es noch etwas zu erledigen oder zu diskutieren gibt. Das hat die oben erwähnten Bedeutungen bestätigt.

Herr Präsident! Meine Damen und Herren! Ja, wir brauchen das Regionalisierungsgesetz. Ich sage Ihnen aber auch ganz klar: Wenn hier so getan wird, als sei **der Keks** schon längst gegessen worden, dann muss ich Ihnen entgegenhalten, dass wir noch darum kämpfen müssen, das durchzukriegen, Herr Dinkla. Bislang scheitert die Kompromissfindung ja daran, dass sich die CDU-geführten Länder diesem Kompromiss bislang verweigert haben. (PNI/W14.00104 Protokoll der Sitzung des Parlaments Landtag Niedersachsen am 24.04.2002. 104. Sitzung der 14. Wahlperiode 1998-2003. Plenarprotokoll, Hannover, 2002 [S. 10356])

„*als sei der Keks schon längst gegessen worden*“ macht klar, dass es schon zu spät sein könnte.

Die Wörterbücher sagen über die statistische Gebrauchshäufigkeit leider überhaupt nichts aus, darum widmet sich die Autorin diesem Bereich im Kapitel „*Korpusanalyse in Bezug auf semantische und diatopische Aspekte*“. Mit Hilfe der Wörterbücher wurde erfahren, ob beide Genera in Betracht gezogen werden und bzw. ob es aus der semantischen oder diatopischen Sicht einen Unterschied gibt. Außer *Keks* haben alle Wörter nur eine einzige Bedeutung, abgesehen davon, ob das Wort ein Neutrum oder Maskulinum ist. *Keks* steht als Maskulinum oder Neutrum, aber in den Fällen, wo es sich auf die Bedeutung *Kopf* bezieht, wird ausschließlich das maskuline Genus erfordert. Nur *Keks* ist diatopisch analysiert. In Langenscheidt (2010) steht, dass *das Keks* eher in Österreich benutzt wird.

8.5 Korpusanalyse in Bezug auf semantische und diatopische Aspekte

Mithilfe der Wörterbücher wurde erfahren, welche Genera zugewiesen werden und demzufolge welche davon als grammatisch korrekt betrachtet werden. Der aktuelle Genusgebrauch spiegelt sich jedoch in einem Korpus. Dazu ermöglicht ein Korpus auch, mit den exportierten Proben aus der diatopischen und semantischen Sicht zu arbeiten.

Für diese Analyse wurde das Deutsche Referenzkorpus ausgewählt. Die Angaben konnten durch das Cosmas-II-System exportiert werden. Die Suchanfrage wurde eingegeben, in unserem Fall allmählich alle vier Substantive *Biskuit*, *Keks*, *Marzipan* und *Nougat*, und anschließend wurden die Ergebnisse ausgesucht. Es wurden die Ergebnisübersicht, KWIC, Belege, Kookkurrenzen und Suchbegriff-Expansionslisten abgerufen. Das System ist so eingerichtet, dass es nicht nur das Vorkommen extrahiert, sondern auch die Kookkurrenzanalyse ermöglicht. Bei den ausgewählten Wörtern wurde wegen des Mangels an spezifischen Kookkurrenzen diesem Bereich fast keine Aufmerksamkeit gewidmet.

Das System gewährte uns Angaben über die Herkunft. Anhand dieser Informationen ließ sich erfahren, in welchen Gebieten welche Tendenzen vorkommen. Was die Bedeutung betrifft musste man mit den extrahierten Texten unbedingt arbeiten. Die erste Ansicht bot uns KWIC (key word in context) an. Man stieß ganz häufig auf Probleme, dass sich das Artikelwort nicht direkt auf das Lemma bezog. Am Anfang musste man bestimmen, welcher Abstand zwischen dem Artikelwort und dem Lemma erlaubt wurde. Das verursachte, dass auch Artikelwörter, die nicht mit dem Lemma verbunden wurden, in der Nähe erschienen. Wenn das so war, wurden diese Proben entweder gestrichen, oder das richtige Artikelwort wurde nachgesucht. Unabhängig von dem Artikelwort kam es in meisten Fällen zu keiner Änderung in der Bedeutung. Die Mehrdeutigkeit wurde nur beim Wort *Keks* entdeckt. Hier musste man mit dem entsprechenden Kontext näher

arbeiten. Das DeReKo berücksichtigt Inhalte und Zusammenhänge überhaupt nicht und deswegen musste man sich die extrahierten Belege genauer ansehen.

Die Proben beinhalteten auch Texte aus der Wikipedia. Diese Proben wurden in die Korpusanalyse nicht einbezogen. Es gab Probleme mit Herkunftsangaben und das heißt, dass es damit diatopisch nicht gearbeitet werden konnte. In manchen Fällen wurde in der Wikipedia nur über das Schwanken diskutiert und welches Artikelwort „richtiger“ sein sollte.

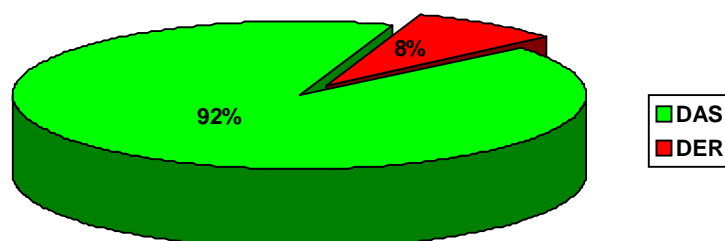
8.5.1 Biskuit

Biskuit stammt aus dem Italienischen, das später von dem Französischen verdrängt wurde. In beiden Sprachen gehört das Wort *Biskuit* zu Maskulina. In allen drei Wörterbüchern der deutschen Sprache, mit denen die Autorin arbeitete, wurde angegeben, dass *der* wie auch *das* ohne Unterschied benutzt werden können. Die Korpusanalyse entdeckte, welches Artikelwort in der Sprache bevorzugt wird. Insgesamt wurden 112 Treffer exportiert.

99 Treffer gehörten zum neutralen Genus. Zwei Proben wurden ausgelassen, weil sie aus der Wikipedia kamen und die Herkunft unbekannt war. Der Rest, 97 Treffer, konnten in die Analyse einbezogen werden.

13 Treffer wurden wie Maskulina exportiert. Während der Untersuchung wurde erfahren, dass nur neun davon gültig waren. Die Beispiele aus der Wikipedia wurden automatisch aussortiert. Außerdem stieß man auf das Problem mit Artikelwörtern, die sich nicht auf das bestimmte Lemma bezogen. Der Treffer "(*der Zigeuner*) und „*Lou Tamborinaire*“ (*der Trommel*) . ***Biskuit***" ist ein gutes Beispiel dafür. Obwohl *Biskuit* artikellos in diesem Fall war, wurde diese Probe markiert. Von dieser Probe ließ sich nicht ableiten, zum welchen Genus *Biskuit* gehörte und man konnte mit dieser Probe nicht weiterarbeiten

Biskuit als Neutrum oder Maskulinum



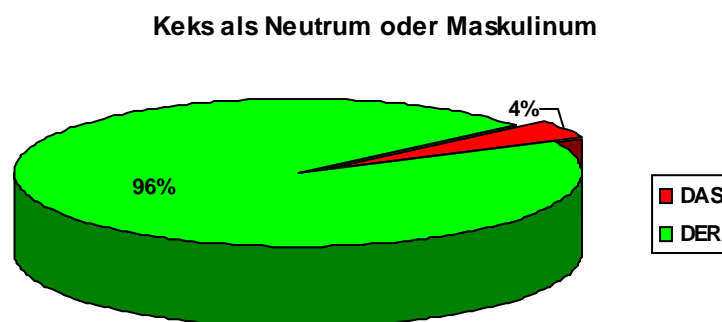
Nach der Datenreinigung blieben 106 Treffer übrig. 97 Treffer waren neutral und neun maskulin. Die Bedeutung blieb gleich in allen Fällen. Trotz der geringeren Zahl der Treffer mit *der*, wurde sehr interessante Tatsache entdeckt u. zw., dass *der* ausschließlich in deutschen Quellen benutzt wurde. Dagegen wurde die neutrale Genuszugehörigkeit in 88 Fällen in österreichischen Zeitungen und in 9 Fällen in Schweizer Quellen gefunden. Für die Verallgemeinerung stehen mit Hinblick auf das maskuline Genus keine ausreichenden Daten zur Verfügung. Trotzdem lässt sich beobachten, dass *der* eher in Deutschland und *das* eher in Österreich und in der Schweiz verwendet wird. In untersuchten Wörterbüchern steht das Artikelwort *das* in Duden und Langenscheidt an der ersten Stelle. Ob das die Präferenzen im Genusgebrauch zeigt, ist nur reine Spekulation. Die Ursprungssprache (Französisch) benutzt das maskuline Genus. Das entspricht dem Genusgebrauch in Deutschland. Ob es einen Zusammenhang dazwischen gibt, ist nicht erfahren worden.

8.5.2 Kekes

Keks wurde aus dem Englischen übernommen. Englisch verfügt nur über natürliches Geschlecht. *Keks* entstand aus dem Englischen Plural *cakes*, während der Singular *cake* neutral ist. In den Wörterbüchern steht, dass beide Genera möglich sind. Duden und Langenscheidt gewähren Zusatzinformation über das benutzte Genus. Das Artikelwort *der* soll häufiger als *das* erscheinen und in Langenscheidt wird dazu hinzugefügt, dass das neutrale Genus überwiegend in Österreich vorkommt. Um dieses zu bestätigen oder abzulehnen, wurde die Korpusanalyse ähnlich wie am Beispiel von *Biskuit* durchgeführt. Insgesamt wurden 195 Treffer exportiert. Es kam wieder zur Datenreinigung und demzufolge senkte die Zahl zu 121 Proben, mit denen man weiter arbeiten konnte.

19 Treffer wurden für das neutrale Genus exportiert. Beiträge aus der Wikipedia wurden ausgelassen. In drei Fällen wurde *Keks* ohne Artikelwort benutzt und einmal bezog sich das Artikelwort auf ein anderes Lemma. Die fünf übriggebliebenen Treffer kamen 3-mal aus österreichischen und 2-mal aus deutschen Quellen.

176 Treffer wurden maskulinem Genus zugewiesen. Nach der Datenreinigung blieben 116 Belege übrig. Zwischen den ausgeschlossenen Fällen waren Vereinsnamen, die häufig erschienen. *Keks* wurde auch wie ein Hundename benutzt. Es wurde schon erwähnt, dass *Keks* auch mit Kopf verbunden wird und das spiegelt sich in zahlreichen Redewendungen wider. Diese Redewendungen wurden in die Analyse einbezogen. Die exportierten Belege berichten in den meisten Fällen über den goldenen Keks, der gestohlen wurde. Er symbolisierte einen wirklichen Keks und demzufolge konnten diese Belege auch analysiert werden. Das maskuline Genus wurde 5-mal in den Schweizer, 7-mal in den österreichischen und 104 -mal in den deutschen Quellen benutzt.



Im Einklang mit Duden und Langenscheidt wurde herausgefunden, dass das Artikelwort *das* seltener als *der* und eher im österreichischen Sprachgebiet vorkommt. Die phonologischen Prinzipien für die Genuszuweisung setzten sich in diesem Fall durch. Substantive mit Konsonantengruppen am Ende gehören eher zu Maskulina. Anteil der Maskulina war wesentlich größer.

8.5.3 Marzipan

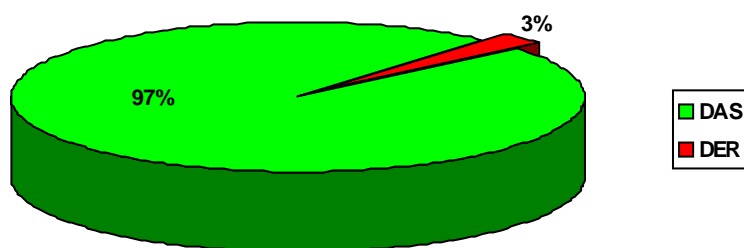
Marzipan wurde aus dem Italienischen übernommen. In der Ursprungssprache geht es um ein Maskulinum. Alle drei untersuchten Wörterbücher akzeptieren das maskuline sowie neutrale Genus. Duden erkennt, dass *der* seltener erscheint. Das wurde mithilfe der Korpusanalyse auch bestätigt.

Insgesamt wurden 150 Treffer exportiert. Es kam genauso zur Datenreinigung und die Zahl wurde um 46 Treffer reduziert. Die Belege aus Wikipedia wurden ausgelassen.

Das maskuline Genus wurde nach der Datenreinigung nur bei drei Belegen anerkannt. Aus zahlreichen Proben konnte wegen des artikellosen Gebrauchs nicht erfahren werden, zu welchem Genus das Substantiv gehörte. Zwei Treffer hatten den Schweizer und ein Treffer den deutschen Ursprung. Laut Duden ist der maskuline Genusgebrauch seltener und das bestätigte auch diese Analyse.

Die Zahl beim neutralen Genus sank um 31 Belegen. Die Quellen aus der Wikipedia wurden in die Analyse nicht einbezogen und in manchen Fällen bezog sich das Artikelwort nicht auf das Lemma. In zahlreichen Fällen wurde *Marzipan* ohne Artikel benutzt. 85 Treffer wurden aus den deutschen, zwölf Treffer aus den österreichischen und vier aus den Schweizer Texten ausgeschnitten. Die Bedeutung blieb abgesehen vom Genusgebrauch in allen Fällen gleich. Das kann auf folgenden Beispielen nachgewiesen werden: „*Nach Bedarf noch etwas Wasser zugeben, **der Marzipan** muss gut zusammenhalten.*“ und „***Das Marzipan** mit dem Amaretto leicht erwärmen*“. Semantik spielt bei diesen ausgewählten Belegen keine Rolle.

Marzipan als Neutrum oder Maskulinum



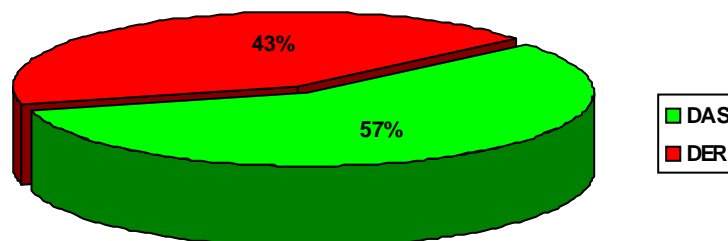
Das benutzt man häufiger als *der*. Der Genusgebrauch von der Ursprungssprache setzte sich offensichtlich bei meisten Fällen nicht.

8.5.4 Nougat

Nougat wurde aus dem Französischen *nougat* übernommen. Im Französischen ist das Substantiv maskulin. In Langenscheidt und Duden sind beide Genera vertreten. Sie gewähren uns keine näheren Informationen hinsichtlich der Diatopik oder Semantik. Bei der Untersuchung von den Wörterbüchern wurde herausgefunden, dass *Nougat* eine alternative Schreibung hat. *Nougat* sowie *Nugat* sind nämlich korrekt.

Insgesamt wurden 22 Treffer exportiert. Nach der Datensortierung hatte man nur 14 Treffer. Sechs Treffer hatten das maskuline Genus. Alle Proben wurden aus den deutschen Quellen extrahiert. Die Treffer mit neutralem Genus hatten 6-mal die deutsche und 2-mal die österreichische Herkunft. Der Genusgebrauch von beiden Genera scheint aufgrund dieser Folgerung ausgeglichen zu sein. Für die objektive Verallgemeinerung mangelt es jedoch an ausreichendem Material.

Nougat als Neutrum oder Maskulinum



9 Schlusswort

Die Autorin beschäftigte sich mit dem Thema „Genusvariation deutscher Substantive. Eine korpusgestützte Fallstudie am Beispiel von *Biskuit*, *Keks*, *Marzipan* und *Nougat*.“ Die Arbeit wurde in drei Hauptteile gegliedert. Der erste Teil bezog sich auf die Korpuslinguistik, der zweite Teil befasste sich mit dem Genus bei deutschen Substantiven und im dritten Teil wurde die Analyse der oben erwähnten Wörter durchgeführt.

Der theoretische Teil war die Basis für den anschließenden praktischen Teil. Um mit einem Korpus zu arbeiten, wurde es am Anfang für wichtig gehalten, Ziele und Potenzial der Korpuslinguistik zu beschreiben und natürlich zu erklären, was überhaupt ein Korpus ist.

Die Korpuslinguistik bemüht sich darum, den Sprachgebrauch zu beobachten und zu beschreiben. Sie ist eng mit Empirie verbunden. Die Sprache entwickelt sich allmählich und das spiegelt sich im Sprachgebrauch wider. Es gibt auch verschiedene sprachliche Varietäten und darum lässt sich die Sprache als etwas Lebendes behandeln. Die Texte, die unter bestimmten Kriterien gewählt werden, sind in Korpora einbezogen. Korpora sind Textsammlungen von schriftlichen und mündlichen Äußerungen. Die Daten in einem Korpus sind häufig noch linguistisch bearbeitet und durch Metadaten ergänzt, die noch etwas über das Trägermedium, die Art der Kodierung, die Autoren oder die Zeit der Entstehung aussagen. Mit diesen Angaben lässt sich dann weiterarbeiten. In dieser Arbeit wurden diese Angaben für die diatopische und semantische Analyse verwendet.

Die Autorin beschäftigte sich primär mit Genusschwankung bei den Wörtern *Biskuit*, *Keks*, *Marzipan* und *Nougat*. Es war nötig die grammatische Kategorie Genus zu erklären. In Bezug auf die Analyse waren die Kriterien für die Genuszuweisung sehr wichtig. Anhand dieser Kriterien wurde überprüft, welche eine Rolle bei den ausgewählten Substantiven spielten. In manchen Fällen setzte sich überhaupt kein Kriterium durch. Bei der Genuszuweisung hat ebenso die Ursprungssprache einen Einfluss und das wurde in der Analyse in Betracht gezogen.

Ein Kapitel widmete sich der Etymologie der Wörter. Die vier ausgewählten Substantive wurden aus dem Italienischen, Französischen und Englischen übernommen. Demzufolge wurden die Genussysteme der einzelnen Gebersprachen und auch die ursprünglichen Genera der Wörter erwähnt. In den Tabellen wurden Angaben aus einem etymologischen Wörterbuch präsentiert. Die Angaben wurden bei der Analyse auch berücksichtigt.

Die Analyse lehnte sich teilweise an die Arbeit mit Wörterbüchern an. Drei verschiedene Wörterbücher wurden untersucht, um den akzeptierten Genusgebrauch bei den Wörtern zu erfahren. Es wurde herausgefunden, dass Duden, Langenscheidt sowie FIN Publishing bei den Wörtern *Biskuit*, *Keks*, *Marzipan* beide Genera, also maskulin und neutral, akzeptieren. Langenscheidt und Duden akzeptieren auch beide Genera bei *Nougat*, aber in FIN Publishing wurde das Substantiv überhaupt nicht angegeben. Langenscheidt und Duden gewähren bei *Keks* auch diatopische Information über häufigeres Vorkommen. Die Korpusanalyse wurde mithilfe des Deutschen Referenzkorpus durchgeführt. Die Belege wurden aus dem Korpus extrahiert. Aufgrund der Daten wurde erfahren, welche Tendenzen im Genusgebrauch zu beobachten sind. Die Korpusanalyse war im Einklang mit den untersuchten Wörterbüchern. Die Genusschwankung wurde bei allen Fällen bestätigt. Beim Wort *Keks* wurde seltener *das* und eher in Österreich benutzt. Das steht auch in Duden (2007) und Langenscheidt(2010). Dem Wort *Marzipan* wurde seltener das maskuline Genus zugewiesen. Das gibt auch Duden (2007) an. *Biskuit* wurde aus der diatopischen Sicht in den Wörterbüchern nicht näher definiert. Die Korpusanalyse entdeckte, dass das Wort häufiger als Neutrum benutzt wird. Das Artikelwort *der* wurde ausschließlich in deutschen Quellen benutzt. Aus der semantischen Sicht blieb die Bedeutung bei *Biskuit*, *Marzipan* und *Nougat* im maskulinen wie auch im neutralen Genus gleich. Beim Wort *Keks* wurde die Mehrdeutigkeit entdeckt. *Keks* kann ebenso mit *Kopf* verbunden werden. Wenn sich *Keks* auf das Wort *Kopf* bezieht, benutzt man es ausschließlich als Maskulinum. In anderen Fällen lässt sich *das* wie auch *der Keks* verwenden. Dem Wort *Keks* wurde in den meisten Fällen maskulines Genus zugewiesen. Im Einklang mit den Regeln für die Genuszuweisung setzten sich hier wahrscheinlich die phonologischen Prinzipien durch. Der Genusgebrauch bei *Biskuit*, *Marzipan* und *Nougat* scheint in den meisten

Fällen ein Zufallsprodukt zu sein, weil keine Prinzipien für die Genuszuweisung angewandt werden konnten und noch dazu wurde das Genus in den meisten Fällen offensichtlich nicht aus den Ursprungssprachen übernommen.

Die Kookkurenzanalyse bot uns überhaupt keine spezifischen Ergebnisse und darum wurde sie in die Arbeit nicht einbezogen.

10 Resumé

Autorka se v této práci zabývá variabilitou rodu německých substantiv. Případová studie se opírá o příklady čtyř substantiv, a sice *Biskuit*, *Keks*, *Marzipan* a *Nougat*. Práce je rozdělena do třech hlavních částí. První část obrací svou pozornost ke korpusové lingvistice samotné. Následující část rozebírá německé substantivum a pojednává také o gramatické kategorii *rod*. Na základě nabytých znalostí z příslušné odborné literatury byla provedena korpusová analýza, která se vztahovala především na diatopickou a sémantickou distribuci kolísání rodu u výše zmíněných substantiv.

Část, která se vztahuje k tématu korpusové lingvistiky, vysvětluje postupy, se kterými zástupci této relativně mladé jazykové disciplíny pracují. V opozici stojí empiristé a racionalisté. Empiristé kladou důraz na praktickou zkušenost, kterou jim v rámci korpusové lingvistiky poskytuje pozorování a popisování užívaného jazyka. Racionalisté naopak zkoumají jazyk z teoretického hlediska. Tyto dva přístupy se v podstatě nevyklučují. Na základě empirického zkoumání se dají ověřit etablované teorie, které jsou v důsledku zkušenosti buďto potvrzeny, nebo zamítnuty. V praktické části byl přínos obou metodických postupů ověřen. V následujících kapitolách autorka zmiňuje korpus. Korpus je sbírka textů, které jsou důkladně vybírány za předem určených podmínek. Konkrétně „das Deutsche Referenzkorpus“ nám pro korpusovou analýzu poskytl anotace a metadata, se kterými se dalo následně pracovat pro zjištění diatopické distribuce. Práce nastiňuje mimo jiné, jakým způsobem a v jakých jiných jazykových disciplínách může být korpus ještě využit.

Neméně důležitou kapitolu představuje *Substantiv im Deutschen*, ve které je tento slovní druh blíže definován. Dále se pak autorka zaměřuje na gramatickou kategorii *rod*. Ta stojí v popředí našeho zájmu, jelikož se práce vztahuje právě na rodovou variaci. Kromě funkce rodu ze syntaktického hlediska, jsou uvedeny principy, které se dají pozorovat v přiřazování určitého rodu substantivu. Odhalit případné principy se snaží autorka v praktické části.

U některých substantiv dochází ke kolísání rodu, které může být mimo jiné způsobeno jejich původem. Proto je v práci obsažena i kapitola „Etymologie der ausgewählten Wörter“. Ta odhaluje, z jakých jazyků ona slova pocházejí a jaký rod v původní řeči mají. V analýze je buď potvrzeno nebo vyvráceno, zda němčina přejala slovo i s rodovou příslušností. Na tuto kapitolu navazuje práce se slovníky. Slovníky v některých případech blíže určují, je-li některý z rodů případně upřednostňován. Pokud má slovo více významů, jsou tam ve většině případů také uvedeny. Nicméně někdy korpusová analýza odhaluje, na co obor lexikografie ještě nestihl zareagovat, a proto je tato disciplína autorkou považována za velmi přínosnou.

Posledním oddílem je korpusová analýza vztahující se na diatopickou a sémantickou distribuci vybraných substantiv. Exportované výsledky z „*das Deutsche Referenzkorpus*“ daly základ pro zkoumání současných tendencí. Jelikož vzorky textů poskytují mimo jiné informace o původu, lze pozorovat, zda ke kolísání dochází na základě regionální variace. Jednotlivé texty, které nejsou exportovány v plném rozsahu, ale přesto dávají dostatečně informací pro odhalení významu slova, umožňují navíc rovněž pracovat v sémantické rovině. V tomto oddíle jsou zohledněny jak principy pro určování rodu, tak i původ slov a údaje obsažené ve vybraných slovnících.

11 Literaturverzeichnis

BUSSMANN, Hadumod, 2008. *Lexikon der Sprachwissenschaft*. Stuttgart : Alfred Kröner Verlag. ISBN 978-3-520-45204-7.

Deutsches Universalwörterbuch, 2007. Mannheim : Dudenverlag. ISBN 34-110-5506-5.

Die Grammatik, 2009. Mannheim : Dudenverlag. ISBN 978-3-411-04047-6.

ENGEL, Ulrich, 2009. *Deutsche Grammatik – Neubearbeitung*. München : Iudicium. ISBN 978-3-89129-914-2.

Německo-český česko německý slovník, 2004. Olomouc : FIN PUBLISHING. ISBN 80-86002-74-8

GÖTZ, Dieter, Günther HAENSCH und Hans WELLMANN, 2010. *Großwörterbuch Deutsch als Fremdsprache: das einsprachige Wörterbuch für alle, die Deutsch lernen*. Neubearbeitung. Berlin : Langenscheidt. ISBN 978-346-8490-422.

HELBIG, Gerhard a Joachim BUSCHA, 2001. *Deutsche Grammatik: Ein Handbuch für den Ausländerunterricht*. Berlin : Langenscheidt. ISBN 3-468-49493-9. 49

HENTSCHEL, Elke a Harald Weydt, 2003. *Handbuch der deutschen Grammatik*. 3., völlig neu bearb. Auflage. Berlin : Walter de Gruyter. ISBN 3-11-017501-0.

HENTSCHEL, Elke a Petra M. Vogel, 2009. *De Gruyter lexikon: Deutsche Morphologie*. Berlin : Walter de Gruyter. ISBN 978-3-11-018562-1.

KESSEL, Katja a Sandra REIMANN, 2010. *Basiswissen deutsche Gegenwartssprache*. 3. Auflage. Tübingen : A. Francke Verlag. ISBN 978-3-8252-2704-3.

LEMNITZER, Lothar a Heike ZINSMEISTER, 2010. *Korpuslinguistik: eine Einführung*. Tübingen : Narr. ISBN 978-382-3365-556.

LENZ, Susanne, 2000. *Korpuslinguistik*. Tübingen : Groos. ISBN 3-87276-848-4.

RÖMER, Christine, 2006. *Morphologie der deutschen Sprache*. Tübingen : A. Francke Verlag. ISBN 978-3-7720-8179-8.

SCHERER, Carmen, 2006. *Korpuslinguistik*. Heidelberg : Winter. ISBN 38-253-5164-5.

TEUBERT, Wolfgang, 1998. *Neologie und Korpus*. Tübingen : Gunter narr Verlag. ISBN 3-8233-5141-9.

INTERNETQUELLEN

http://www.essex.ac.uk/linguistics/external/clmt/w3c/corpus_ling/content/history.html, zit. am 2.12.2013.

<http://www.cl2011.org.uk/history-of-corpus-linguistics.html>, zit. am 2.12.2013.

<http://www.ids-mannheim.de/org/aufgaben.html>; zit. am 15.4.2014.

<http://www.ids-mannheim.de/org/orga.html>; zit. am 15.4.2014.

<http://www1.ids-mannheim.de/kl/projekte/korpora/>; zit. am 24.4.2014.

<http://www1.ids-mannheim.de/kl/projekte/korpora/einsatz.html>; zit. am 24.4.2014.

<http://www.ids-mannheim.de/cosmas2/uebersicht.html>; zit. am 24.4.2014.

<http://www.ids-mannheim.de/cosmas2/projekt/nutzung/referenz.html>; zit. am 5.6. 2014.

<http://www.ids-mannheim.de/cosmas2/projekt/hilfe/quick.html>; zit. am 5.6. 2014.

https://www.uni-due.de/imperia/md/content/prodaz/das_genus_in_verschiedenen_sprachen_alphabetisch_geordnet.pdf; zit. am 29.5. 2014.

<http://www.duden.de/hilfe/herkunft>; zit. am 29.5. 2014.

<http://www.dwds.de/?qu=Biskuit>; zit. am 29.5. 2014.

<http://www.dwds.de/?view=1&qu=Keks>; zit. am 29.5. 2014.

<http://www.dwds.de/?view=1&qu=Marzipan>; zit. am 29.5. 2014.

<http://www.dwds.de/?view=1&qu=Nougat>; zit. am 29.5. 2014.

<http://www.gutefrage.net/frage/der-keks-ist-gegessen>; zit. am 7.6. 2014.