

**Universität Stuttgart
Institut für Maschinelle Sprachverarbeitung
Azenbergstraße 12
D-70174 Stuttgart**

Studienarbeit

Die Verwendung des Rezipientenpassivs im Deutschen

**Eine korpusbasierte Untersuchung für die Auswahl von
Generierungsalternativen**

Vorgelegt von:

Patrick Ziering

Beginn der Arbeit: 20. Juni 2010
Ende der Arbeit: 20. September 2010

Studienarbeit Nr. 108
Betreuer: Prof. Dr. Jonas Kuhn, Sina Zarriß
Prüfer: Prof. Dr. Jonas Kuhn

Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbständig verfasst habe und dabei keine andere als die angegebene Literatur verwendet habe.

Alle Zitate und sinngemäßen Entlehnungen sind als solche unter genauer Angabe der Quelle gekennzeichnet.

Stuttgart, den 20.09.2010

Patrick Ziering

Inhaltsverzeichnis

1	Einleitung	1
2	Das Rezipientenpassiv in der Literatur	2
2.1	Genus Verbi	2
2.2	Vorgangspassiv	4
2.2.1	Unpersönliches Vorgangspassiv	4
2.2.2	Restriktionen des Vorgangspassivs	5
2.2.3	Probleme des Vorgangspassivs	7
2.3	Zustandspassiv	7
2.4	Rezipientenpassiv	10
2.4.1	Unpersönliches Rezipientenpassiv	11
2.4.2	Verteilung der Auxiliare	11
2.4.3	Restriktionen an das partizipiale Vollverb	12
2.4.4	Mögliche Dativformen beim freien Dativ	13
2.4.5	Die Kontroverse um das Rezipientenpassiv	14
2.4.6	Das Rezipientenpassiv in der Grammatikalisierung	15
2.5	Funktionen der Passivierung	16
3	Problemstellung	18
3.1	Verwandte Arbeiten	18
3.1.1	Soft-Constraints spiegeln Hard-Constraints wider	18
3.1.2	Vorhersage der Dativ-Alternation	20
3.2	Fragestellung	21
3.3	Mögliche Anwendung: Generierung aus semantischen Repräsentationen	22
4	Das Experiment	24
4.1	Die Korpora	26
4.1.1	Extraktion von Sätzen im Rezipientenpassiv	27
4.1.2	Extraktion möglicher Rezipientenpassiv-Verben	29
4.1.3	Extraktion von Sätzen mit den häufigsten Rezipientenpassiv-Verben	30
4.2	Datenaufbereitung	30
4.2.1	Parsing extrahierter Sätze	31
4.2.2	Transfer der Features aus den Parses	33
4.3	Daten-Analyse	36
4.3.1	Erstellung von Feature-Sets aus den feature-Dateien	36
4.3.2	Auswertung der Feature-Sets mittels eines Klassifikators	37
4.3.3	Manuelle Auswertung der Feature-Sets	38

5 Ergebnisse und deren Evaluation	40
5.1 Ergebnisse der Auswertung	40
5.2 Schlussfolgerungen	48
6 Zusammenfassung	49
Literaturverzeichnis	51

Abbildungsverzeichnis

2.1	Genus Verbi als verbales Merkmal	2
2.2	Argument-Funktions-Zuweisung in Aktiv und VP; nach (Pittner u. Berman, 2004)	4
2.3	Argument-Funktions-Zuweisung in Aktiv und ZP; nach (Pittner u. Berman, 2004)	8
2.4	Argument-Funktions-Zuweisung in Aktiv und RP; nach (Pittner u. Berman, 2004)	10
4.1	Aufbau des Experiments	25
4.2	Beispiel: C-Struktur	32
4.3	Beispiel: F-Struktur	33
4.4	Beispiel: Transferierte F-Struktur	34
5.1	Diathesen-Verteilung bei <i>Agens-Rezipient-Thema</i> -Beschränkung	41
5.2	Confusion-Matrix der Drei-Diathesen-Klassifikation	42
5.3	Confusion-Matrix der Passiv-Klassifikation	43

Tabellenverzeichnis

2.1	Semantische Rollen zur Aktiv-Passiv-Unterscheidung, (Bussmann; Eisenberg)	3
2.2	Formen-Vergleich VP vs. <i>werden</i> -Kopula-Konstruktion, (Eisenberg, 1999)	7
2.3	Formen-Vergleich ZP vs. <i>sein</i> -Kopula-Konstruktion, (Pittner u. Berman, 2004)	9
3.1	Passivierungsraten im Englischen bzgl. Person/Rolle; nach (Bresnan et al., 2001)	19
4.1	Verwendete Korpora aus CQP; Beschreibung nach IMS-CORPORA (2010)	27
4.2	Anzahl der gefundenen Rezipientenpassive je Korpora	29
4.3	Anzahl der gefundenen Rezipientenpassive je Vollverb	30
4.4	Die relativ häufigsten RP-Vollverben	31
5.1	Anzahl an Feature-Sets für die Klassifikation	40
5.2	Precision, Recall und F-Maß der Drei-Diathesen-Klassifikation	41
5.3	Precision, Recall und F-Maß der Passiv-Klassifikation	43
5.4	Precision, Recall des RPs relativ zum Subkat-Filter	44
5.5	Wichtigste Features bei manueller Auswertung der Daten (1)	45
5.6	Wichtigste Features bei manueller Auswertung der Daten (2)	47

1 Einleitung

Neid muss man sich erarbeiten. Mitleid bekommt man geschenkt.

Dieses alte Sprichwort enthält eine Konstruktion, die für das Deutsche einzigartig und zugleich sehr umstritten ist. Es handelt sich hierbei um das sogenannte **Rezipientenpassiv**, welches sich in diesem Falle aus den Wörtern *bekommen* und *geschenkt* zusammensetzt.

Nun hätte der weise Urheber dieses Zitats durchaus weitere Möglichkeiten gehabt, den zweiten Satz zu formulieren. So wäre der Satz

Mitleid wird einem geschenkt

mit einem **Vorgangspassiv** eine durchaus denkbare Alternative gewesen. Gibt es einen Grund, weshalb er sich für dieses kontroverse Rezipientenpassiv entschieden hat? Beide Varianten würden dasselbe ausdrücken.

Diese Arbeit versucht, auf diese und ähnliche Fragen eine Antwort zu finden. Motiviert durch zwei Arbeiten von Joan Bresnan wird eine Lösung zur folgenden Fragestellung gesucht: Gibt es Kontexte oder Faktoren, die eine Bevorzugung des Rezipientenpassivs erklären?

Sollte das Rezipientenpassiv mittels bestimmter Gegebenheiten erklärbar sein, so lässt sich im Umkehrschluss auch aus diesen Faktoren die Passiv-Art vorhersagen. Eine mögliche Anwendung dieser Vorhersage stellt die Generierung aus semantischen Repräsentationen dar, die auf einer abstrakteren Ebene befindlich keine Informationen über die Aktiv- oder Passiv-Form enthalten.

Zunächst gibt das 2. Kapitel einen theoretischen Einblick in die Form und Natur des Rezipientenpassivs. Wie wird es in der Literatur bewertet und welche Alternativen, insbesondere das Vorgangspassiv, stehen diesem gegenüber? Im 3. Kapitel werden zwei Arbeiten Joan Bresnans beschrieben, in denen sie sich mit einer ähnlichen Problemstellung beschäftigt hat. Mittels statistischer Methoden kann Bresnan diese Problematik erfolgreich lösen. Weiter werden im 3. Kapitel die Fragestellung auf Basis von Bresnans Arbeiten präzisiert und weitere Fragen aufgeworfen, die diese theoretische Studie leiten sollen. Abschließend wird im 3. Kapitel als mögliche Anwendung einer erfolgreichen Vorhersage ein Generierungsranking aus semantischen Repräsentationen vorgestellt.

Motiviert durch Bresnans Erfolg mittels statistischer Klassifikation, wird im 4. Kapitel ein Experiment beschrieben, durch welches eine Lösung zur gegebenen Fragestellung gefunden werden kann: Über verschiedene Korpora werden Daten gesammelt, diese werden durch diverse Tools aufbereitet und abschließend mittels eines Klassifikators sowie durch eine manuelle Betrachtung der Datenlage ausgewertet.

Im 5. Kapitel werden die Ergebnisse der Auswertung diskutiert und deren Anteil an der Lösung der Problemstellung abgewogen. Wurde die Fragestellung erfolgreich beantwortet? Wie ließen sich eventuelle Defizite im Ergebnis noch beheben?

Abschließend fasst das 6. Kapitel die gesamte Arbeit nochmals zusammen.

2 Das Rezipientenpassiv in der Literatur

Dieses Kapitel befasst sich mit der Form und Natur des Rezipientenpassivs. Was macht diese Konstruktion so einzigartig und lädt zugleich zu einigen kontroversen Diskussionen über ihren Status ein? Was unterscheidet diese Passivart von anderen und welche Funktionen werden dadurch realisiert? Zunächst soll jedoch ein kurzer Überblick über die verschiedenen Genera Verbi, speziell über die Passivformen, im Deutschen darüber Auskunft geben, welche Alternativen es zum Rezipientenpassiv gibt.

2.1 Genus Verbi

Das Verb im Deutschen besitzt 5 Merkmale, nach denen es gekennzeichnet ist (Pittner u. Berman, 2004). Neben *Numerus*, *Person*, *Tempus* und *Modus* zeichnet sich das Verb noch im *Genus Verbi*, der „Art des Verbs“ (Eisenberg, 1999, S. 123), aus, welches 2 mögliche Werte besitzt: Aktiv und Passiv. Diewald (1997) beschreibt das Genus Verbi als eine grammatische Kategorie, die die semantischen Rollen der Verbergänzungen sowie die Gerichtetheit des vom Verb ausgedrückten Ereignisses betrifft.

Das Deutsche bietet drei Passivformen, welche sich hinsichtlich des Aktivs in der Zuordnung der thematischen Rollen zu syntaktischen Funktionen oder, im Falle des Zustandspassivs, in der Semantik unterscheiden: *Vorgangspassiv* (2.2), *Zustandspassiv* (2.3) und *Rezipientenpassiv* (2.4). Pittner u. Berman (2004) schlagen hierzu Grafik 2.1 vor, wobei die Passiv-

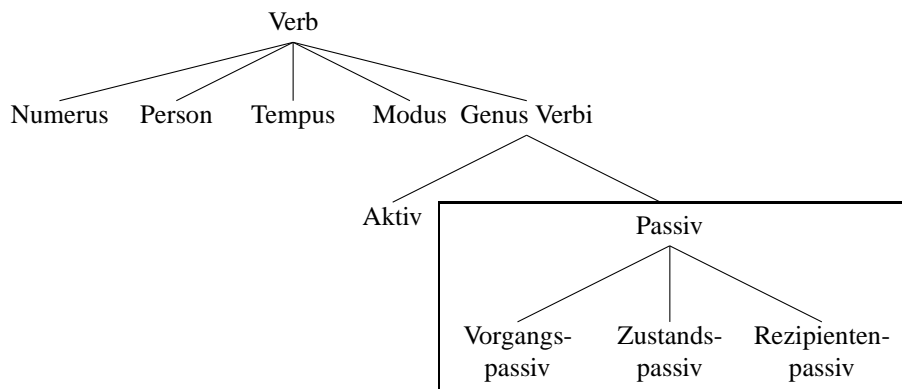


Abbildung 2.1: Genus Verbi als verbales Merkmal

Differenzierung nochmals hervorgehoben wurde.

Die Genera Verbi gehören als valenzmodifizierende Mittel zu den sogenannten *Diathesen*. Bussmann (2002) beschreibt diese als eine „aus dem Griech. übernommene Bezeichnung für Genus Verbi [...] und für andere reguläre Valenzrahmenwechsel wie Applikativ, Akkusativierung, Dativierung u.a.“

Die Begriffe *Aktiv* und *Passiv* zeigen an, was die semantische Funktion des grammatischen Subjekts in den Diathesen ist (Eisenberg, 1999, S. 134).

Die prototypische Form einer Passiv-Konstruktion besteht aus einem Hilfsverb (*werden, sein, bekommen, kriegen, erhalten*) und einer Partizip-II-Form eines passivierbaren Vollverbs. Die Verwendung der unterschiedlichen Hilfsverben realisiert die Form verschiedener Passivarten: *werden* wird im Vorgangspassiv (2.2) verwendet, *sein* ist für das Zustandspassiv (2.3) definiert und *bekommen, erhalten* sowie *kriegen* erhalten den Rezipientenpassivstatus (2.4) zugeteilt.

Das Verhältnis des Aktivsatzes zu seinem Passivsatz besteht darin, dass die beiden Sätze dieselben semantischen Rollen (Mitspielerrollen in der durch das Verb ausgedrückten Szene (Diewald, 1997)) modellieren, allerdings über unterschiedliche grammatische Funktionen. Das Vollverb im Aktiv entspricht der partizipialen Verbform des Passivs. Dem aktivischen Subjekt entspricht eine fakultative Präpositionalphrase, dem (in)direkten aktivischen Objekt entspricht das passivische Subjekt. Die restlichen Ergänzungen des Aktivsatzes bleiben im Passivsatz erhalten und bis auf das Subjekt ändert sich durch die Passivierung bei keiner Ergänzung die Obligatorik des Vorhandenseins (Eisenberg, 1999, S. 125).

Für die hier relevante Aktiv-Passiv-Unterscheidung sind die semantischen Rollen *Agens*, *Rezipient* und *Thema* wichtig, da sie die Aktanten einer durch ein Rezipientenpassiv beschreibbaren Situation darstellen. Tabelle 2.1 erläutert diese Rollen genauer.

Argument	Beschreibung	Beispiel
Agens	Urheber/Verursacher einer Handlung	<i>Er</i> verbraucht das Geld
Rezipient	Der von einer Verbalhandlung indirekt betroffene, typischerweise belebte Mitspieler: Empfänger bei Verben des Besitzwechsels oder Adressat bei Mitteilungsverben	Sie spenden <i>den Flüchtlingen</i> Geld Sie erzählen <i>den Flüchtlingen</i> vieles über Deutschland
Thema	Alternativ: <i>Patiens/Objektiv</i> : Das, worauf sich die vom Verb bezeichnete Handlung richtet	Rainer kauft <i>ein Auto</i>

Tabelle 2.1: Semantische Rollen zur Aktiv-Passiv-Unterscheidung, (Bussmann; Eisenberg)

Eisenberg (1999) nennt neben den hier relevanten Diathesen noch die *Mediale*, die weder passivisch noch aktivisch einzuordnen sind, und *modale Infinitive*.

Diese *Mediale*, auch *Mittelkonstruktionen* genannt, enthalten obligatorisch ein Reflexivpronomen (1).

- (1) a. Das Buch liest sich gut
- b. Morgens lernt es sich am besten

Modale Infinitive bestehen aus dem Hilfsverb *sein* oder *bleiben* und einem infiniten Verb im 2. Status, dem *zu*-Infinitiv (2).

- (2) a. Die Niederlage ist vom Trainer zu verantworten
- b. Der Erfolg deiner Mühe bleibt abzuwarten

Die folgenden Abschnitte befassen sich neben dem *Zustandspassiv* (2.3) speziell mit dem *Vorgangspassiv* (2.2) und *Rezipientenpassiv* (2.4), deren Abgrenzung voneinander im Laufe der Arbeit weiter forciert wird. Abschließend wird ein Blick auf mögliche Funktionen der Passivierung (2.5) geworfen.

2.2 Vorgangspassiv

Die gebräuchlichste Form der Passiv-Diathese bildet das *Vorgangspassiv*.

- (3) a. Ein Bekannter kaufte das Auto
- b. Das Auto wurde (von einem Bekannten) gekauft

Im Aktiv-Satz (3a) steht das Agens im Nominativ und bildet das Subjekt. Das Thema steht im Akkusativ und bildet das direkte Objekt. Das Vollverb steht finit an zweiter Position nach dem Subjekt.

Der Passivsatz (3b) wird gebildet aus dem Auxiliar *werden* und dem infiniten Vollverb im 3. Status, dem Partizip Perfekt. Das Thema steht im Nominativ und bildet das Subjekt. Das Agens wird fakultativ als *von*-Phrase genannt. Die Argumentstruktur des Verbs bleibt erhalten (Pittner u. Berman, 2004).

Abbildung 2.2 stellt diese Argument-Funktion-Relationen im Aktiv und Vorgangspassiv (kurz: VP) dar.

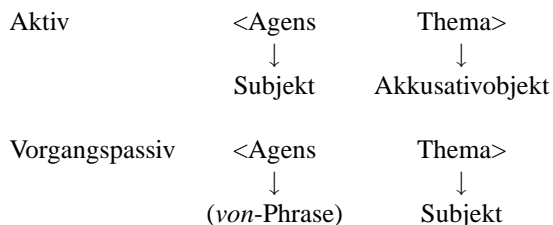


Abbildung 2.2: Argument-Funktions-Zuweisung in Aktiv und VP; nach (Pittner u. Berman, 2004)

2.2.1 Unpersönliches Vorgangspassiv

Das Vorgangspassiv lässt sich auch auf bestimmte Verben anwenden, die kein direktes Objekt fordern.

- (4) a. Man tanzt heute

- b. Heute wird (von allen) getanzt
- c. Man tanzt *transitiv* heute den Tango
- d. Heute wird *der Tango* (von allen) getanzt *transitiv*

In (4a) fehlt ein direktes Objekt, das bei Passivierung zum Subjekt werden kann. Dadurch fehlt bei (4b) das Subjekt, wobei die Agens-Nennung durch eine *von*-Phrase jederzeit möglich ist. Falls das Verb *tanzen* transitiv verstanden wird (4c), so wird das Thema oligatorisch zum Subjekt des Satzes (4d).

Alle anderen vom Verb regierten Komplementfunktionen bleiben bei der Passivierung erhalten (5c, 5e). Das Fehlen eines Subjekts rechtfertigt folglich keine Anhebung einer anderen aktivischen Funktion, etwa eines Genitiv- oder Dativobjektes, in den Subjektstatus. (5d, 5f)

- (5) a. Sie gedachten der Toten
- b. Sie halfen dem Mann
- c. Der Toten wurde (von ihnen) gedacht
- d. * Die Toten wurde (von ihnen) gedacht
- e. Dem Mann wurde (von ihnen) geholfen.
- f. * Der Mann wurde (von ihnen) geholfen.

Die Ergänzung des Passivs in (4b) durch ein impersonales *es* führt bei intransitiver Lesart von *tanzen* zu einem ungrammatischen Satz (6a). Ein Vorkommen von *es* ist nur im Vorfeld (die Position vor einem finiten Verb, das aus seiner Grundposition am Satzende herausbewegt wurde) möglich (6b). Hierbei handelt es sich allerdings um kein formales Subjekt, sondern um ein Vorfeld-*es* (Pittner u. Berman, 2004), welches die Erhaltung des Verbzweitschemas sicherstellt (Eisenberg, 1999)

- (6) a. * Heute wird es (von allen) getanzt *intransitiv*
- b. Es wird heute (von allen) getanzt *intransitiv*

Das unpersönliche Passiv ist im Rezipientenpassiv nicht möglich (2.4.1). Dieser Unterschied wird im Laufe dieser Arbeit nochmals aufgegriffen. Welche Verben für das (unpersönliche) Vorgangspassiv in Frage kommen, wird im nachfolgenden Abschnitt (2.2.2) erläutert.

2.2.2 Restriktionen des Vorgangspassivs

Nicht alle Verben lassen sich ins Vorgangspassiv transformieren. Einige - auch transitive - Verben sind für die Passivierung ausgeschlossen.

- (7) a. haben, besitzen, bekommen, kriegen, erhalten
- b. kosten, wiegen, gelten, enthalten, fassen
- c. erstaunen, freuen, verwundern, entsetzen, schmerzen

Eisenberg (1999) zählt hierunter Besitzverben (7a), Verben, die Mengenverhältnisse beschreiben, (7b) und Verben der Wahrnehmung (7c). Der Grund hierfür liegt darin, dass keines der Verben in (7) ein agentives Subjekt fordert. Dieses ist jedoch für eine Passivierung obligatorisch.

Eine Ausnahme hierzu bilden Verben mit nichtagentiver Lesart wie *verbinden*, *trennen*, *teilen* oder *bedecken* (8). Bei diesen ist ein Vorgangspassiv möglich (Eisenberg, 1999, S. 131).

- (8) a. Die beiden Zimmer werden durch einen Gang verbunden
- b. Die Stadt wird durch die Eisenbahnlinie geteilt
- c. Die Berge werden von Schnee bedeckt

Allerdings gibt es auch Einschränkungen auf Verben, die ein Agens regieren. Dies folgt aus dem Schluss, dass Konstruktionen mit Reflexivpronomen nur im Aktiv realisierbar sind (9c) und nicht im Passiv mit einem Reflexivpronomen als Subjekt (9d).

- (9) a. Karla beschuldigt Klaus
- b. Klaus wird (von Karla) beschuldigt
- c. Karla beschuldigt sich
- d. * Sich wird (von Karla) beschuldigt

Eisenberg (1999) nennt hierfür folgende Gründe:

1. Das Reflexivpronomen *sich* kommt als solches nur im Akkusativ vor. Eine nominative Subjekt-Beschreibung ist daher nicht möglich.
2. Das Reflexivpronomen verweist anaphorisch auf das aktivische Subjekt. Eine Passivierung würde den Bezug des Pronomens zu seinem Antezedenten vernichten.
3. Da sowohl aktivisches Subjekt als auch Reflexivpronomen auf die Agens-Rolle verweisen, würde eine Passivierung zu keiner Veränderung der Obligatorik des Vorhandenseins des Agens führen.

Zwar können Verben wie *beschuldigen* passiviert werden (9b), jedoch ist dies für obligatorisch reflexive Verben (10) trotz eines agentiven Subjekts nicht möglich (Eisenberg, 1999, S. 129).

- (10) sich beeilen, entschließen, vertragen, kümmern, bessern, bewerben,...

Eine Ausnahme bildet das „Reflexivpassiv“, das bei solchen Verben einem unpersönlichen Passiv (2.2.1) gleicht wie in (11) ausgeführt. Diese Konstruktion wird verwendet, wenn der Sprecher eine energische Aufforderung aussprechen möchte.

- (11) Hier wird sich nicht geschämt/verkrochen/...

Überraschend ist, dass Verben wie *schämen*, die kein Agens fordern, durchaus im Reflexivpassiv auftreten können.

Eine weitere Restriktion stellt der inkorporierte Akkusativ dar. Für einen Satz wie (12a) gibt es keine passive Diathese mit *Karten* als Subjekt (12c), was an der Subjekt-Verb-Kongruenz deutlich wird (vgl. (12b)). Der Grund hierfür ist, dass *Karten* mit dem Verb ein komplexes Prädikat bildet (Pittner u. Berman, 2004).

- (12) a. Heute Abend spielen wir Karten
- b. Heute Abend wird Karten gespielt
- c. * Heute Abend werden Karten gespielt

- d. * Heute Abend werden sie gespielt

Hier hat die NP *Karten* keinen Satzgliedstatus. Sie ist nicht pronominalisierbar (12d) und nicht permutierbar (Pittner u. Berman, 2004, S. 71).

2.2.3 Probleme des Vorgangspassivs

Ein Problem beim Vorgangspassiv stellt die Differenzierung jener Diathese zu aktivischen Kopula-Konstruktionen mit dem formgleichen Kopulaverb *werden* dar. Die beiden Konstruktionen unterscheiden sich lediglich in der partizipialen Form von *werden* (Tabelle 2.2). Diese ist im Vorgangspassiv *worden* und in einer Kopula-Konstruktion *geworden*.

Vorgangspassiv	Kopula-Konstruktion
wird geteilt	wird ärgerlich
wurde geteilt	wurde ärgerlich
wird geteilt werden	wird ärgerlich werden
ist geteilt worden	ist ärgerlich geworden
war geteilt worden	war ärgerlich geworden
wird geteilt worden sein	wird ärgerlich geworden sein

Tabelle 2.2: Formen-Vergleich VP vs. *werden*-Kopula-Konstruktion, (Eisenberg, 1999)

So wird durch die Form *worden* eine Konstruktion mit zwei Partizipien vermieden (13).

- (13) * Das Vaterland ist geteilt geworden.

2.3 Zustandspassiv

Eine weitere Passivform bildet das sogenannte *Zustandspassiv*, welches einen aus einer Handlung resultierenden Zustand bezeichnet (Bussmann, 2002, S. 871).

- (14) a. Sie verschließt die Tür
 b. Die Tür ist (* von ihr) verschlossen

So beschreibt (14b) den Endzustand des in (14a) genannten Vorgangs. Es wird durch das Hilfsverb *sein* zusammen mit der Partizipialform eines Vollverbs gebildet.

Anders als bei Vorgangs- und Rezipientenpassiv ist eine fakultative Nennung des Agens in den meisten Fällen nicht möglich.

Abbildung 2.3 stellt diese Argument-Funktion-Relationen im Aktiv und Zustandspassiv (kurz: *ZP*) dar.

In seltenen Fällen kann jedoch auch beim Zustandspassiv eine *von*-Phrase auftreten (15).

- (15) a. Die Berge waren *von Schnee* bedeckt
 b. Er ist *vom Pech* verfolgt

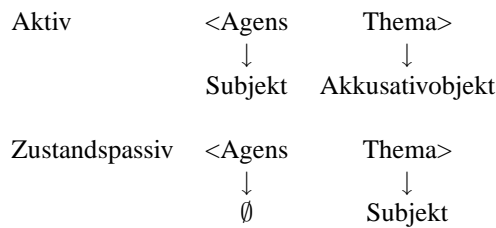


Abbildung 2.3: Argument-Funktions-Zuweisung in Aktiv und ZP; nach (Pittner u. Berman, 2004)

Unpersönliches Zustandspassiv

Die unpersönliche Form des Zustandspassivs ist äußerst selten (16a).

- (16) a. Damit ist ihm nicht geholfen
 b. * Heute ist (von allen) getanzt

Restriktionen des Zustandspassivs

Nicht alle Verben können im Zustandspassiv vorkommen. Dies folgt unmittelbar aus der Tatsache, dass das Zustandspassiv ein Ergebnis beschreibt, welches aus einer Handlung resultiert. Folglich können nur solche Verben eine Zustandspassiv-Konstruktion bilden, die auch ein Vorgangspassiv bilden können.

Dies lässt sich in Formel 2.1 als Teilmengenbeziehung aller Verben beschreiben. Die Menge aller zustandspassivierbaren Verben ist Teilmenge aller vorgangspassivierbaren Verben.

$$VERBEN_{ZP} \subseteq VERBEN_{VP} \quad (2.1)$$

Eisenberg (1999) nennt ein weiteres Kriterium für zustandspassivierbare Verben. So muss ein Vorgang, aus dem ein im Zustandspassiv beschriebenes Resultat hervorgeht, **inkrementell** sein, d.h. „er vollzieht sich kontinuierlich in Schritten derselben Art“ (Eisenberg, 1999, S. 131). Wird ein Zustandspassiv mit agentiven, aber nicht-inkrementellen Verben gebildet, so ist dessen Grammatikalität zweifelhaft (17).

- (17) ? Die Studentin ist begrüßt; Das Bild ist angeblickt; Das Haus ist gezeigt; Die Arbeit ist gelobt

Daher handelt es sich bei den im Zustandspassiv realisierbaren Verben um eine echte Teilklasse der agentiven Verben, die ein Vorgangspassiv bilden können (Formel 2.2).

$$VERBEN_{ZP} \subset VERBEN_{VP} \quad (2.2)$$

Probleme des Zustandspassivs

Analog zur Problematik im Vorgangspassiv (2.2.3) stellt sich auch beim Zustandspassiv die Frage, ob es sich hierbei um eine Passivdiathese oder eine aktivische Kopula-Konstruktion mit

dem Kopulaverb *sein* handelt (18a). In (18b) tritt dieses Problem nicht auf, da das Prädikativ keine Partizipialform hat, es ist ein Adjektiv (Pittner u. Berman, 2004).

- (18) a. Die Tür ist geöffnet
 b. Die Tür ist offen

Eine Möglichkeit zur Unterscheidung bietet bedingt die Komparation („eine morphologische Kategorie von Adjektiven und Adverbien zum Ausdruck von Gradangaben und Vergleichen“ (Bussmann, 2002)). Demnach sind Verben nicht komparierbar und so von Adjektiven abgrenzbar (19).

- (19) a. Die Tür ist offener als die andere
 b. * Die Tür ist geöffneter als die andere

Allerdings lässt sich *geöffnet* mit *un-* präfigieren und mittels eines Kopulaverbs in anderen Kopula-Konstruktion verwenden, was gegen eine Analyse von *geöffnet* als verbale Form spricht (20).

- (20) Der Brief bleibt/ist noch ungeöffnet

Anders als beim Vorgangspassiv lässt sich beim Zustandspassiv die Unterscheidung zur Kopula-Konstruktion auch durch keine Tempusform festlegen (Tabelle 2.3).

Zustandspassiv	Kopula-Konstruktion
Die Tür ist geöffnet	Die Tür ist offen
Die Tür war geöffnet	Die Tür war offen
Die Tür ist geöffnet gewesen	Die Tür ist offen gewesen
Die Tür war geöffnet gewesen	Die Tür war offen gewesen
Die Tür wird geöffnet sein	Die Tür wird offen sein
Die Tür wird geöffnet gewesen sein	Die Tür wird offen gewesen sein

Tabelle 2.3: Formen-Vergleich ZP vs. *sein*-Kopula-Konstruktion, (Pittner u. Berman, 2004)

Eine klare Abgrenzung von Zustandspassiv zu *sein*-Kopula-Konstruktionen bleibt daher unmöglich.

2.4 Rezipientenpassiv

Die letzte hier vorgestellte Diathese stellt den Kern dieser Arbeit dar, das *Rezipientenpassiv*. Alternative Benennungen für diese Passivform sind nach Diewald (1997); DUDEN (2009):

- bekommen-Passiv
- Dativ-Passiv
- Adressatenpassiv
- Benefizientenpassiv

Für die Bildung des Rezipientenpassivs stehen die Hilfsverben *bekommen*, *erhalten* sowie *kriegen* zur Verfügung. Auch hier steht das passivierte Vollverb im Partizip II (21).

- (21) a. Er bekam ein Buch geschenkt
 b. Sie erhält den Nobelpreis verliehen
 c. Die Frau kriegt die Bretter zugeschnitten

Durch das Rezipientenpassiv (kurz: *RP*) wird es ermöglicht, den Rezipienten, welcher dem Dativobjekt im Aktivsatz entspricht, als Subjekt zu realisieren (Pittner u. Berman, 2004). Der Rezipient (nach Diewald (1997) die „Zielrolle“) entspricht dem Handlungspartner des Agens, ist belebt, meist menschlich und trägt damit ausreichende Qualifikation, um zur Nominativergänzung promoviert zu werden. Das meist unbelebte Thema, welches im Vorgangspassiv das Subjekt konstituiert, bleibt nun als direktes Objekt der Aktivform entsprechend gleich. Das aktivische Subjekt wird fakultativ als *von*-Phrase realisiert.

Abbildung 2.4 stellt diese Argument-Funktion-Relationen im Aktiv und Rezipientenpassiv dar.

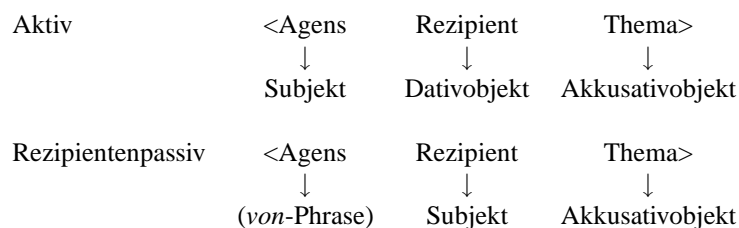


Abbildung 2.4: Argument-Funktions-Zuweisung in Aktiv und RP; nach (Pittner u. Berman, 2004)

Die Topikalisierung (Bewegung ins Vorfeld) der Dativergänzung bei Beibehaltung aller aktiver Rollenordnungen ist keine zufriedenstellende Alternative zum Rezipientenpassiv, da dadurch eine markierte Satzstellung erzeugt würde, die nicht immer möglich ist (Diewald, 1997).

Eisenberg (1999) deutet an, dass das Rezipientenpassiv eine hilfreiche Ergänzung zum Aktiv und Vorgangspassiv darstelle, insofern es nun möglich sei, jede der drei hier diskutierten semantischen Rollen (*Agens*, *Rezipient*, *Thema*) als Subjekt zu kodieren (22).

- (22) a. Die Karla (*Agens*) füllt dem Karl das Formular aus
 b. Das Formular (*Thema*) wird dem Karl von der Karla ausgefüllt
 c. Der Karl (*Rezipient*) bekommt von der Karla das Formular ausgefüllt

2.4.1 Unpersönliches Rezipientenpassiv

Anders als bei den bisherigen Passivdiathesen ist beim Rezipientenpassiv eine unpersönliche Form ausgeschlossen (Pittner u. Berman, 2004, S. 74).

- (23) a. * Heute bekam verziehen
 b. * Natürlich bekam die Aufgaben erklärt

Daraus lässt sich schließen, dass jede Rezipientenpassiv-Konstruktion einen Rezipienten beinhaltet, aber nicht jedes Vorgangspassiv ein Thema. Im späteren Verlauf der Arbeit wird sich zeigen, ob dieser Unterschied für die Auswahl der richtigen Passiv-Variante von Bedeutung ist.

2.4.2 Verteilung der Auxiliare

Dieser Abschnitt befasst sich mit der Distribution sowie Einschränkung der Hilfsverben des Rezipientenpassivs auf semantischer sowie textsortaler Basis, wie sie in (Pimanyonok, 2004, S. 11ff) beschrieben ist. Für das Rezipientenpassiv lassen sich die Auxiliare *bekommen*, *kriegen* und *erhalten* verwenden. Diese sind semantisch und in ihrer Verwendungsmöglichkeit jedoch nicht äquivalent.

Erhalten wird in literarischen sowie Sach- und Anweisungstexten verwendet und oft dem „Behördenstil“ zugeordnet. *Kriegen* ist in der Schriftsprache sehr selten und wird größtenteils umgangssprachlich verwendet. *Bekommen* findet den häufigsten Gebrauch, weshalb oft auch vom *bekommen-Passiv* gesprochen wird.

Es zeigt sich eine deutliche Abgrenzung der Auxiliare *bekommen* und *kriegen* von dem Auxiliar *erhalten*. Grammatisch (aber nicht stilistisch) betrachtet lässt sich *bekommen* stets durch *kriegen* substituieren und umgekehrt. Diese Äquivalenz gilt für *erhalten* nicht. Bei den häufig markierten, teils inakzeptablen Konstruktionen mit Verben des „Wegnehmens“ ist das Auxiliar *erhalten* nicht möglich (24a). Das Rezipientenpassiv kann ebenfalls nicht mit *erhalten* gebildet werden, wenn das Vollverb intransitiv ist (24b) oder in einer Konstruktion mit einem Pertinenzdativ (vgl. (2.4.4)) wie in (24c) auftaucht.

- (24) a. * Er erhält den Führerschein entzogen
 b. * Die Kinder erhielten dann vorgelesen
 c. * Er erhielt auf die Schulter geklopft

Unter dem stilistischen Aspekt sind die Auxiliare *erhalten* und *kriegen* als eine Gruppe in Opposition zu *bekommen* zu betrachten. Sie unterliegen anders als das neutrale *bekommen* stärkeren Verwendungsrestriktionen. So spielt die stilistische Charakteristik des Vollverbs eine entscheidende Rolle in der Auswahl der kombinierbaren Auxiliare. Beispielsweise lässt sich *vorjammern* als stilistisch eher umgangssprachlich klassifiziert nicht mit dem gehobeneren

erhalten vereinen (25b), jedoch ist eine Realisierung mit *kriegen* grammatischer anzusehen (25c).

- (25) a. Sie jammert ihm ständig vor, dass ...
b. * Er erhält (von ihr) ständig vorgejammert, dass ...
c. ? Er kriegt (von ihr) ständig vorgejammert, dass ...

2.4.3 Restriktionen an das partizipiale Vollverb

Zunächst lässt sich festhalten, dass nur Verben für die Rezipientenpassiv-Bildung in Frage kommen, die auch im Vorgangspassiv möglich sind (Pimanyonok, 2004). Wie bei allen bisherigen Passivformen ist auch für die Bildung des Rezipientenpassivs ein agentives Vollverb eine notwendige Voraussetzung:

- (26) a. Er bekam die Zeitung vorgelesen
b. * Er bekam die Arbeit gut geglückt

Das Rezipientenpassiv wird am häufigsten von gewöhnlichen ditransitiven Verben realisiert, d.h. von Verben des *Gebens*, *Nehmens*, *Mitteilens*, *Verbergens*, usw... (DUDEN, 2009, S. 550). Das Vollverb subkategorisiert also die Argumente *Agens*, *Thema* und *Rezipient*:

- (27) Peter (*Rezipient*) bekam von seiner Mutter (*Agens*) ein Fahrrad (*Thema*) geschenkt

Ditransitive Verben, die keinen Rezipienten fordern, sind bei der Bildung des Rezipientenpassivs ausgeschlossen (Diedrichsen, 2004):

- (28) a. Du schuldest mir 100 Euro
b. * Ich bekomme von dir 100 Euro geschuldet

Neben den Dativ regierenden Verben sind auch solche möglich, die einen *freien Dativ* (2.4.4) erlauben:

- (29) a. Seine Oma strickt *Peter* einen Pulli
b. Peter bekommt von seiner Oma einen Pulli gestrickt

Eine Rezipientenpassiv-Konstruktion bei zweistelligen Verben ist selten, aber wird durchaus als grammatisch aufgefasst (Pittner u. Berman, 2004, S. 74):

- (30) a. Wer heute hilft, bekommt morgen geholfen
b. Er bekommt verziehen

„Passivsätze, die mit *bekommen* (*erhalten*, *kriegen*) gebildet sind, enthalten (fast) immer ein Akkusativobjekt. Sie unterscheiden sich dadurch grundsätzlich von anderen Passivsätzen.“ (DUDEN, 2009)

Wie auch beim Vorgangspassiv (2.2.2) ist das Rezipientenpassiv nicht durch transitive Verben generierbar, bei welchen das Dativkomplement als Reflexivum realisiert ist (Pimanyonok, 2004):

- (31) a. Er stellt sich die Frage
b. * Er bekommt von sich die Frage gestellt

2.4.4 Mögliche Dativformen beim freien Dativ

Neben den Verben, die eine Dativergänzung fordern, ist das Rezipientenpassiv auch bei Verben möglich, die einen *freien Dativ* erlauben.

Nach Bussmann (2002); Helbig u. Buscha (2005) gibt es folgende Gruppen von freien Dativen:

Dativus ethicus Der ethische Dativ dient dem Ausdruck einer persönlichen Stellungnahme
→ *Fahr mir aber vorsichtig*

Träger-Dativ (z.B.: Träger eines Kleidungsstücks)

→ *Dem Jungen rutscht die Hose* (mit Beziehung auf das Subjekt)

→ *Er zieht ihr den Mantel an* (mit Beziehung auf das Objekt)

→ *Ich trete ihm auf den Schuh* (mit Beziehung auf die Adverbialbestimmung)

Pertinenzdativ Der possessive Dativ

→ *Meinem Vater schmerzt der Kopf* (mit Beziehung auf das Subjekt)

→ *Wir waschen uns die Hände* (mit Beziehung auf das Objekt)

→ *Er sieht seiner Tochter in die Augen* (mit Beziehung auf die Adverbialbestimmung)

Dativus Comodi/Incomodi Der Dativ des Interesses bezeichnet die Person oder Sache, zu deren Gunsten/Nachteil etwas geschieht

→ *Er strickte ihr einen Pulli*

Dativus ludicantis Der Standpunkt-Dativ bezeichnet den Standpunkt, von welchem aus eine Äußerung Gültigkeit hat

→ *Er ist mir zu intelligent*

Dativ des Zustandsträgers

→ *Dem Schüler ist diese Zensur ein Trost*

So kann statt einer Dativergänzung auch ein Pertinenzdativ Subjekt des Rezipientenpassivs werden (32a). Weitere mögliche Dativformen sind der Dativus Comodi (32b) und entsprechend der Dativus Incomodi (32c)

- (32) a. Er bekommt auf den Fuß getreten
b. Er bekommt (von seiner Tochter) die Schuhe geputzt
c. Er bekommt den Antrag abgelehnt

Pimanyonok (2004) schließt den Rezipientenpassiv jedoch für Konstruktionen mit Pertinenzdativ und Subjekt (33b) oder Pertinenzdativ, Subjekt sowie Situativkomplement (33d) aus. Gleiches gilt für intransitive Verben mit Dativus Incomodi (33f)

- (33) a. Der Rücken tat ihm weh
b. * Er bekam vom Rücken wehgetan
c. Der Eintopf lag ihm im Magen
d. * Er bekam vom Eintopf im Magen gelegen
e. Sie lief ihm davon
f. * Er bekam von ihr davongelaufen

2.4.5 Die Kontroverse um das Rezipientenpassiv

Von vielen Grammatiken wurde das Rezipientenpassiv als in die Umgangssprache gehörend oder als zu vernachlässigende Konkurrenzform des Vorgangspassivs degradiert (Diewald, 1997). Jedoch hat sich das Rezipientenpassiv im heutigen Deutsch ausreichend etabliert, um anerkannt zu werden. In den meisten Grammatiken wird heute das Rezipientenpassiv als eine sich in der **Grammatikalisierung** befindliche Diathese betrachtet (2.4.6). Ob es sich beim Rezipientenpassiv um eine reguläre Passivform handelt, ist in der Literatur jedoch stark umstritten. Hierfür kann folgender Grund genannt werden:

Die Hilfsverben *bekommen*, *kriegen* und *erhalten* verfügen in den jeweiligen Konstruktionen meist über noch sehr viel eigenständige Semantik und sind oft schwer von den entsprechenden Vollverben zu trennen (Pittner u. Berman, 2004, S. 74). Dies hat mehrere Folgen. Einmal wirken Konstruktionen mit Vollverben, die semantisch das Gegenteil beschreiben, oft ungrammatisch oder stark markiert (34).

- (34) a. ? Er bekam das Fahrrad gestohlen
b. ? Er bekam den Führerschein entzogen

Desweiteren führt die Polyfunktionalität von *bekommen* zu mehreren Lesarten.

- (35) Peter bekommt einen Apfel gewaschen

So hat Satz (35) vier mehr oder weniger markierte Lesarten:

1. Rezipientenpassiv: *Jemand wäscht Peter einen Apfel*
2. Gelingen-Lesart: *Peter gelingt es, einen Apfel zu waschen*
3. freies Objekt-Prädikativ: *Peter bekommt einen gewaschenen Apfel*
4. freies Subjekt-Prädikativ: *Peter bekommt gewaschen einen Apfel*

Mehrheitlich wird (35) als ein Satz mit mehreren syntaktischen Strukturen betrachtet: das Partizip des Rezipientenpassivs steht topologisch im Verbalkomplex (rechte Satzklammer) wohingegen das freie Prädikativ in der Vollverb-Lesart von *bekommen* im Mittelfeld positioniert ist (Pittner u. Berman, 2004, S. 75):

- (36) a. Wir bekamen das Auto vom Händler [*Verbalkomplex* gewaschen]
b. Wir bekamen [*Mittelfeld* das Auto gewaschen vom Händler]

Neben der topologischen Unterscheidung lässt sich noch die *un*-Präfigierung als Abgrenzungskriterium verwenden. So ist es möglich, freie Prädikative als Adjektive mit *un*- zu präfigieren (37a), was mit verbalen Partizip-Formen nicht möglich ist (37b).

- (37) a. Peter bekam einen Apfel ungewaschen
b. * Peter bekam das Fahrrad ungestohlen

Der Versuch einer Interpretation von (37b) führt unmittelbar zur Vollverb-Lesart mit freiem Prädikativ (*Peter bekam das Fahrrad, welches nicht gestohlen war*).

Eine mögliche Lösung dieser Kontroverse bietet die Theorie der Grammatikalisierung im nachfolgenden Abschnitt (2.4.6).

2.4.6 Das Rezipientenpassiv in der Grammatikalisierung

Dieser Abschnitt orientiert sich größtenteils an dem Grammatikalisierungsprozess des Rezipientenpassivs, wie ihn Diewald (1997, S. 32ff) skizziert. Bussmann (2002, S. 289) beschreibt die Grammatikalisierung als eine „Bezeichnung eines Sprachwandelprozesses, in dessen Verlauf eine autonome lexikalische Einheit allmählich die Funktion einer abhängigen grammatischen Kategorie erwirbt“. Im aktuellen Fall erwerben *bekommen*, *erhalten* und *kriegen* die Funktionen eines Auxiliars zur Bildung einer Passivdiathese.

Wie bereits in (2.4.5) beschrieben, haben Sätze in Form eines Rezipientenpassivs oft mehrere Lesarten. Einerseits kann *bekommen* als Vollverb betrachtet werden und andererseits als Auxiliar (38b).

- (38) a. Sie bekommt die Bretter (vom Schreiner) schon passend zugeschnitten
b. Sie bekommt den Katalog zugeschickt
- Vollverb: *Sie bekommt den Katalog, und zwar zugeschickt und nicht etwa am Kiosk*
 - Auxiliar: *Der Katalog wird ihr zugeschickt*
- c. August bekam den Führerschein weggenommen

Satz (38a) wird als Ausgangspunkt der Grammatikalisierung angenommen. Hierbei ist *bekommen* das Vollverb und *zugeschnitten* eine Prädikativangabe: *Sie bekommt die Bretter, welche bereits passend zugeschnitten sind*.

Nun sind die Vollverben *bekommen*, *erhalten* und *kriegen* keine prototypischen transitiven Verben. Sie fordern als externes Argument kein Agens, sondern einen Rezipienten. Ein Agens kann hier (wie beim Passiv) nur durch eine fakultative Präpositionalphrase genannt werden. Hinsichtlich dieser Einzigartigkeit beschreiben diese Vollverben bereits eine Sondergruppe als „lexikalische Konversen“ zu *geben*. Sie kehren die in *geben* beschriebene Handlungsrichtung (*Agens* → *Verb* → *Thema* → *Ziel*) um in die *Rezipientenperspektive* (*Agens* ← *Verb* ← *Thema* ← *Ziel*). Bei dieser Sondergruppe handelt es sich um eine äußerst kleine Klasse, wodurch die „paradigmatische Variabilität“, ein Kennzeichen für Nichtgrammatikalisierung, bereits im Vollverbgebrauch stark eingeschränkt ist.

Den für (38b) beschriebenen Verwendungen des Auxiliars liegen zwei Grammatikalisierungsstufen zugrunde. Während die Vollverb-Verwendung von *bekommen* den Ausgangspunkt der Grammatikalisierung ausmacht, hat in der Verwendung als Auxiliar bereits eine *Reanalyse* (eine „Restrukturierung der Konstituentenstruktur“) stattgefunden.

Diewald (1997) nennt zwei Bedingungen für diese Reanalyse. Einmal müssen bestimmte aktuelle Verhältnisse zwischen dem Vollverb *bekommen* und dem Partizip vorliegen. Während der Rezipient in (38a) die Bretter erst nach dem Zuschneiden erhält, finden das Zuschneiden und das Bekommen in (38b) gleichzeitig statt. Der Zwang zu einer resultativen Lesart besteht nun nicht mehr und eine notwendige Voraussetzung zur Reanalyse ist damit erfüllt.

Eine weitere Bedingung zur Reanalyse stellt die Valenzstruktur des im Partizip denotierten Vollverbs dar. Hier sind zunächst nur ditransitive Verben möglich, die neben der Nominativergänzung noch eine Dativ- und Akkusativergänzung fordern. Wird vom Verb kein Akkusativobjekt regiert, so wirkt eine resultierende Rezipientenpassiv-Konstruktion häufig ungrammatisch oder von vielen Sprechern nur partiell akzeptiert (39).

(39) ? Er bekommt geschmeichelt/geholfen/auf die Füße getreten

Weshalb im Rezipientenpassiv eine Akkusativergänzung erforderlich ist, scheint auf den ersten Blick nicht klar. So benötigt auch weder das Vorgangspassiv (2.2) noch das Zustandspassiv (2.3) ihrerseits eine Dativergänzung. Betrachtet man jedoch das Rezipientenpassiv als ein noch nicht gänzlich vollzogener Grammatikalisierungsprozess mit einer prädikativen Verwendung des Partizips als Ausgangslage, so zeigt sich, dass die Akkusativergänzung benötigt wird, um den Übergang des Partizips von der Prädikativangabe zum partizipialen Vollverb nachzuvollziehen, indem das Akkusativobjekt den Bezug des Partizips sicherstellt. Erreicht die Grammatikalisierung ihren Endzustand, wird kein Akkusativobjekt mehr notwendig sein (39). Ein weiteres Beispiel für die Endstufe der Grammatikalisierung sind Konstruktionen mit dem Vollverb *bekommen* semantisch gegensätzlichen Partizipien (38c). Hier ist eine Vollverb-Lesart ausgeschlossen, *bekommen* kann nur als Auxiliar auftreten (Pimanyonok, 2004).

Dieser Einblick in die Theorie der Grammatikalisierung soll für diese Arbeit reichen. Die einzelnen Schritte des Grammatikalisierungsprozesses sind in (Diewald, 1997, S. 36ff) detailliert beschrieben.

2.5 Funktionen der Passivierung

Aktivsätze und Passivsätze bedeuten „im Wesentlichen“ dasselbe (Eisenberg, 1999, S. 125). Demnach stellt sich die Frage, welche Funktion die Diathesen eines Verbs haben.

Als mögliche Funktionen einer Passivierung wird u.a. die Realisierung eines Perspektivenwechsels genannt. Im Aktivsatz wird eine Begebenheit in der *Handlungsperspektive* dargestellt, wohingegen im Passivsatz das Ereignis aus der *Geschehensperspektive* wiedergegeben wird (40).

- (40) a. Hausmeister Krause ruft die Feuerwehr an.
b. Die Feuerwehr wird von Hausmeister Krause angerufen.

In (40a) wird das Anruf-Ereignis aus Sicht des Agens betrachtet (*der Hausmeister telefoniert*) und in (40b) aus Sicht des Themas (*der Anruf geht bei der Feuerwehr ein*) (Pittner u. Berman, 2004).

Eine weitere Funktion bietet die Passivierung durch die fakultative Realisierung des Agens als Präpositionalphrase. Dadurch findet eine Valenzreduktion statt (Pittner u. Berman, 2004).

Die Realisierung einer thematischen Rolle als Subjekt macht diese wichtiger und ihre fakultative Erwähnung als *von*-Phrase macht sie *beiläufig*. So lassen sich Satzglieder hervorheben oder *beiläufig* benennen (Eisenberg, 1999, S. 133).

Allerdings kann die Unterscheidung von wichtigen und unwichtigen Ergänzungen nicht allein auf Subjekt und *von*-Phrase zurückgeführt werden. Vielmehr beruht eine solche Unterscheidung auf der *Informationsstruktur*: Ausdrücke in einem Satz können an bereits Bekanntes anknüpfen oder mit bereits Genanntem koreferieren. Das Bekannte in einem Satz wird *Thema* genannt und ist hier nicht mit gleichnamiger Theta-Rolle zu verwechseln. Der Rest des Satzes enthält die neue Information und wird *Rhema* genannt. So kann ein Text erst als *kohärent* betrachtet werden, wenn die *Thema-Rhema*-Markierung mit der erwarteten *Thema-Rhema*-Aufteilung übereinstimmt.

Die Thematisierung wird im Deutschen durch das Subjekt sowie durch Topikalisierung vollzogen. Ob beide Kriterien für ein Thema unabhängig voneinander zu betrachten sind, bleibt umstritten, da das Subjekt, als bevorzugter Satzinitiator, meist thematisch ist.

- (41) a. Der Präsident schlägt den Kanzler vor. Das Parlament wählt den Kanzler.
- b. Der Präsident schlägt den Kanzler vor. Der Kanzler wird vom Parlament gewählt.
- c. Der Präsident schlägt den Kanzler vor. Den Kanzler wählt das Parlament.

(41a) gilt demnach als *inkohärent*, da *den Kanzler* rhematisiert wird, obwohl dieser im Vorsatz bereits erwähnt wurde, also ein erwartetes Thema darstellt. (41b) und (41c) stellen beide einen kohärenten Nachfolgesatz dar, da das Thema in beiden Fällen einem im Vorsatz eingeführten Referenten korrespondiert.

„Der Passivsatz erlaubt es also, dieselbe semantische Rolle zu thematisieren wie der Aktivsatz mit vorangestelltem Objekt“ (Eisenberg, 1999, S. 133/134). So ist es möglich, sowohl im Aktiv als auch im Passiv das Subjekt zu *rhematisieren*. Im Passiv ist diese (fakultative) Agens-Information meist rhematisiert, im Aktiv nicht.

Eine mögliche Abgrenzung des Vorgangspassiv vom Rezipientenpassiv bildet das Gewicht der *von*-Phrase im gesamten Rhema. Die *von*-Phrase kann im Passivsatz selbst das Rhema bilden oder gemeinsam mit dem Verb (42):

- (42) a. Der Karl bekommt von der Karla das Formular ausgefüllt.
- b. Das Formular wird dem Karl von der Karla ausgefüllt.

In (42a) ist die *von*-Phrase wichtigstes Element des Rhemas, in (42b) ist sie Teil des Rhemas mit dem Verb *ausgefüllt* als wichtigstes Element. Ein Indikator hierfür ist der unterschiedliche Satzakzent in beiden Fällen (Eisenberg, 1999, S. 134).

Ob diese grundlegende Unterscheidung der Wichtigkeit der *von*-Phrase im Rhema zu einer tendenziell häufigeren Agens-Verwendung im Rezipientenpassiv führt, wird im Laufe dieser Arbeit geklärt werden.

Eine weitere Funktion, so führt Diedrichsen (2004) an, ist bei der Koordinationsreduktion zu finden (43a). Hier kann das Subjekt in beiden Satzkonjunkten weitergeführt werden, einmal als Thema im Vorgangspassiv und einmal als Rezipient im Rezipientenpassiv. Fände kein Wechsel der Diathesen statt, so wäre die Konstruktion ohne Reduktion zu realisieren (43b, 43c), was zu einem informationsstrukturell inkohärenten Ergebnis führte.

- (43) a. Er_i fuhr zu schnell, pro_i wurde von der Polizei angehalten und pro_i bekam den Führerschein entzogen
- b. Er_i fuhr zu schnell und die Polizei hielt ihn $_i$ an
- c. Er_i fuhr zu schnell und die Polizei $_j$ entzog ihm $_i$ den Führerschein

3 Problemstellung

Dieses Kapitel befasst sich mit der dieser Arbeit zugrundeliegenden Fragestellung: Ist es möglich, mithilfe der Auswertung bestimmter grammatischer Merkmale die Verwendung des Vorgangs- und Rezipientenpassivs korrekt vorherzusagen? Lassen sich die in Kapitel 2 beschriebenen Kennzeichen der Diathesen entsprechend modellieren? Motiviert wurde die Arbeit durch zwei ähnliche Studien von Joan Bresnan, die in (3.1) beschrieben werden.

In (3.2) wird die Fragestellung mithilfe von Bresnans Erkenntnissen konkretisiert und weitere Fragen aufgeworfen, die mittels des im nachfolgenden Kapitel beschriebenen Experiments zu beantworten versucht werden. In (3.3) wird die Generierung aus semantischen Repräsentationen als eine mögliche Anwendung des Ergebnisses der Fragestellung vorgestellt.

3.1 Verwandte Arbeiten

3.1.1 Soft-Constraints spiegeln Hard-Constraints wider

In (Bresnan et al., 2001) wird beschrieben, inwiefern sich Hard-Constraints einer Sprache als Soft-Constraints in einer anderen Sprache widerspiegeln, anhand der Sprachen Englisch und Lummi (British Columbia). Unter einem Hard-Constraint ist eine strikte grammatische Vorschrift zu verstehen, bei deren Missachtung ein Satz ungrammatisch wird. Bei einem Soft-Constraint handelt es sich jedoch um kein grammatisches Gebot, dessen Verletzung zu einem ungrammatischen Ergebnis führt, sondern um eine statistische Präferenz, die auf großen Datenmengen evaluiert durchaus sichtbar wird, jedoch in vielen Fällen auch nicht in Erscheinung tritt.

Bresnan nennt als repräsentatives Beispiel die Personen-Hierarchie (eine „hierarchische Anordnung der verbalen Kategorien der Person, die in manchen Sprachen für die Grammatik von Bedeutung ist“ (Bussmann, 2002, S. 572)) in Lummi und im Englischen. Für die hier zur Diskussion stehende Differenzierung ist die in Formel (3.1) dargestellte Anordnung der Personen-Werte anzunehmen. Sie besagt, dass die 1. und 2. Person prominenter ist als die 3. Person.

$$1. \text{ Person} \cup 2. \text{ Person} > 3. \text{ Person} \quad (3.1)$$

In Lummi ist die Personen-Hierarchie ein grundsätzlicher Faktor, auf welchem ein Hard-Constraint basiert. Dieser besagt, dass die prominente grammatische Funktion Subjekt nicht mit einer bzgl. der Personen-Hierarchie weniger prominenten Person assoziiert sein kann als das Objekt.

Ist das Objekt mit einer Person der prominenteren Personengruppe (1. oder 2. Person) assoziiert, so kann das Subjekt auch nur mit der 1. oder 2. Person realisiert werden. Ist das Objekt

durch eine mit der 3. Person markierten Phrase ausgedrückt, so kann das Subjekt mit allen Person-Werten beschrieben werden.

Die folgenden Beispiele zeigen mögliche Verletzungen dieses Hard-Constraints. Sie sind in deutscher Sprache verfasst, deren Grammatikalitätsbeurteilung ist jedoch auf Lummi ausgelegt.

- (44) a. *_L Der Mann_{3ps} sieht uns_{1ps}
 b. Wir_{1ps} werden von dem Mann_{3ps} gesehen
 c. *_L Der Mann_{3ps} wird von uns_{1ps} gesehen
 d. Wir_{1ps} sehen den Mann_{3ps}

Der Aktivsatz in (44a) ist ungrammatisch, da das Subjekt als *common noun* realisiert den Personen-Wert 3 trägt und das direkte Objekt in der 1. Person auftaucht. Die in (44a) beschriebene Proposition (der „Satzinhalt“ (Bussmann, 2002, S. 616)) kann jedoch durch einen Passivsatz wie in (44b) grammatisch ausgedrückt werden. Hier ist das Subjekt in der 1. Person codiert und die *von*-Phrase in der 3. Person. Es zeigt sich, dass die Verletzung des Hard-Constraints dadurch vermieden werden kann, dass ein Diathesen-Wechsel erfolgt von Aktiv zu Passiv ((44a) → (44b)) oder umgekehrt ((44c) → (44d)).

Diese harmonische Beziehung zwischen prominenter Person und prominenter Funktion wird in allen Sprachen als existent betrachtet, nur kann in manchen Sprachen dieser Constraint durch andere überlagert sein, deren Priorität sprachspezifisch höher liegt.

Im Englischen ist diese Harmonie nicht als Hard-Constraint vorhanden. Die Sätze (45a) und (45b) sind beide grammatisch.

- (45) a. The man_{3ps} sees us_{1ps}
 b. We_{1ps} are seen by the man_{3ps}

Bresnan untersuchte geparte Sätze aus dem SWITCHBOARD-Korpus, einem Korpus bestehend aus spontanen Telefon-Konversationen von über 500 weiblichen und männlichen Amerikanern.

Trotz der Grammatikalität von Sätzen wie (45a) hat Bresnan eine leichte Tendenz zur Passivierung im Falle der Verletzung dieses Constraints entdeckt. Verglichen mit der Passivierungsrate bei sowohl Agens und Patiens (/Thema) in 3. Person von 1,2 % ist eine Passivierung bei Agenten in 1. oder 2. Person und Patienten in 3. Person praktisch nicht vorhanden (0,0 %) und bei Agenten in 3. Person und Patienten in 1. oder 2. Person deutlich erhöht (2,9 %). Dies illustriert auch Tabelle 3.1:

Aktion:	# Aktiv	# Passiv	% Aktiv	% Passiv
1./2. → 1./2.	179	0	100,0%	0,0 %
1./2. → 3.	6246	0	100,0%	0,0 %
3. → 3.	3110	39	98,8%	1,2 %
3. → 1./2.	472	14	97,1%	2,9 %

Tabelle 3.1: Passivierungsraten im Englischen bzgl. Person/Rolle; nach (Bresnan et al., 2001)

Nun stellt sich die Frage: Ist mithilfe dieses Constraints im Deutschen auch die Verwendung des Rezipientenpassivs gegenüber dem Vorgangspassiv erklärbar? Bresnan beschreibt

hierdurch die Aktiv-Passiv-Alternation. Dass auch die VP-RP-Alternation durch die Personen-Hierarchie, zumindest partiell, zu begründen ist, scheint durchaus denkbar.

Gibt es zudem noch weitere Soft-Constraints, auf welchen die Alternation basiert?

3.1.2 Vorhersage der Dativ-Alternation

In (Bresnan et al., 2005) zeigt Bresnan anhand der Dativ-Alternation im Englischen, dass es linguistische Phänomene gibt, die auf traditionelle Weise, etwa über die Bedeutung des durch das Verb beschriebenen Ereignisses, nicht erklärbar, aber mithilfe statistischer Mittel bis zu einem sehr hohen Prozentsatz vorhersagbar sind.

Im Englischen ist es möglich, eine Proposition beschrieben durch (46a) mit zwei Objekt-Nominalphrasen zu realisieren (46b), man spricht hier auch von der *double-object*-Konstruktion, oder den dem Rezipienten zugewiesenen Referenten als Präpositionalphrase satzfinal zu implementieren (46c).

- (46) a. give(susan, the children, toys)
- b. Susan gives the children toys
- c. Susan gives toys to the children

Die Meinung vieler Linguisten hierzu ist, dass linguistische Probleme dieser Art zu komplex seien. Linguistische Theorie brauche eine fundamentale Idealisierung. Der Slogan „Grammar is grammar and usage is usage“ wirft diese und ähnliche Probleme aus dem Rahmen der linguistischen Theorie. Bresnan widerlegt im Laufe ihrer Arbeit diese Annahme.

Ein traditioneller Ansatz zur Behandlung dieser Problematik ist die Rückführung der Dativ-Struktur auf die Bedeutung des Ereignisses. Die Sätze in (47) beschreiben unterschiedliche Bedeutungen bzgl. des *geben*-Ereignisses. Diese Bedeutungen lassen sich als *Bewegungsergebnis* (47a) und als *Zustandswechsel* (47b) generalisieren.

- (47) a. I pushed the box to John
- b. The lighting here gives me a headache

Allerdings zeigen bereits (46b) und viele weitere Gegenbeispiele aus dem Internet (48), dass die Erklärung der Dativ-Struktur auf alleiniger Basis der Ereignis-Bedeutung nicht ausreicht.

- (48) a. ...Stories like these must **give the creeps to people whose idea of heaven is a world without religion...**
- b. Karen spoke with Gretchen about the procedure for registering a complaint, and **hand-carried her a form**, but Gretchen never completed it

In Satz (48a) handelt es sich trotz einer den Rezipienten codierenden Präpositionalphrase um kein Ereignis der Bewegung. Ein Grund für seine Existenz könnte das „Gesetz der wachsenden Glieder“ sein, eines der von O. Behaghel formulierten Grundprinzipien der Wort- und Satzgliedstellung (Bussmann, 2002). Diesem zufolge geht von zwei Gliedern das kürzere dem längeren voraus. Daraus lässt sich schließen, dass idiomatische Wendungen wie *give the creeps* trotz der Tendenz, in einer *double-object*-Konstruktion vorzukommen, den Gesetzen von Behaghel unterliegen.

In Satz (48b) handelt es sich trotz der zwei Objekt-Nominalphrasen um ein Bewegungsergebnis. Hier fallen jedoch folgende Faktoren auf. Der Rezipient, *her*, ist im selben Satz vorerwähnt, trägt also eine sehr hohe Salienz. Er ist definit, pronominal und kürzer als die nachfolgende Nominalphrase. Daraus lässt sich schließen, dass Bewegungsverbren zwar tendenziell in Konstruktionen mit präpositionalen Dativen auftauchen, jedoch ein Abgleich des Rezipienten mit dem Thema bzgl. Definitheit, Pronominalität, Diskurs-Gegebenheit und Länge zu einer Konstruktion mit zwei Objekt-Nominalphrasen führen kann. Dies ist auf eine Studie von Collins (1995) zurückzuführen.

Nach Collins Korpus-Studie über die Dativ-Konstruktionen im australischen Englisch wird in den *double-object*-Konstruktionen der Rezipient auf den Skalen der Salienz, Definitheit, Pronominalität und Länge prominenter realisiert als das Thema. Die Dativ-Struktur würde folglich so gewählt werden, dass gegebene Referenten den nicht-gegebenen vorangehen, Pronomen vor Nicht-Pronomen stehen, definite Ausdrücke vor indefiniten positioniert werden und kürzere Phrasen vor längeren (vgl. „Gesetz der wachsenden Glieder“).

Bresnan versucht, die Dativ-Alternation durch die oben diskutierten Faktoren mithilfe eines logistischen Regressions-Modells vorherzusagen. Neben diesen Faktoren fügt Bresnan noch weitere Variablen in das Modell ein: Numerus- und Personen-Werte sowie eine semantische Klassifizierung der Verbverwendung (*abstrakt, Besitzwechsel, zukünftiger Besitzwechsel, Verhinderung von Besitz, Kommunikation*).

Ihr konstruiertes Modell erreichte eine Accuracy von 94 %. Bresnans Ergebnis in dieser Studie: Linguistische Daten sind probabilistischer als bisher in weiten Kreisen angenommen. Eine Vorhersage der Dativ-Alternation ist somit keineswegs zu komplex, sondern mit statistischen Mitteln sehr zuverlässig.

Nun stellt sich die Frage: Ist mit derselben Methode die Alternation zwischen Vorgangs- und Rezipientenpassiv vorhersagbar? Welche Merkmale können dafür aus dieser Studie übernommen werden?

3.2 Fragestellung

Bresnans Erfolg in (Bresnan et al., 2001, 2005) motiviert, die Alternation zwischen Vorgangs- und Rezipientenpassiv auf ähnliche Weise vorherzusagen. Die Studien werfen nun folgende Fragen auf:

- Ist die Alternation zwischen den zur Diskussion stehenden Passiv-Varianten auf einen Soft-Constraint oder eine Kombination aus mehreren Soft-Constraints zurückzuführen?
- Handelt es sich bei dieser Alternation um ein linguistisches Phänomen, welches mittels statistischer Auswertung vorhersagbar ist?
- Spielen harmonische Beziehungen bzgl. Definitheit, Salienz, Pronominalität oder relativer Länge ähnlich der Studie von Collins (1995) auch hier eine Rolle?

Aus den bisherigen Erkenntnissen tauchen noch folgende elementare Fragen auf, die die Studie leiten sollen und deren Antworten neben der automatischen Daten-Analyse mittels eines Klassifikators weitere Schlüsse auf die Verwendung der Passiv-Formen geben könnten.

- Anders als beim Vorgangspassiv ist im Rezipientenpassiv keine unpersönliche Form möglich. Spiegelt sich diese Obligatorik des Rezipienten in den Daten wider?
- Beim Rezipientenpassiv trägt das Agens mehr Bedeutung im Satz-Rhema als beim Vorgangspassiv (vgl. (42)). Ist dieses Merkmal durch eine häufigere Agens-Nennung im RP erkennbar?
- Wie oft taucht der Rezipient im VP auf? Ist es generell möglich, die Wahl der Diathesen auf ein Vorhandensein von fakultativen Ergänzungen zurückzuführen?
- Gibt es Verben, die häufiger im Rezipientenpassiv realisiert werden als im Vorgangspassiv? Welches Verb kommt am häufigsten im Rezipientenpassiv vor? Hängt die Wahl der Passiv-Variante etwa von der lexikalischen Semantik des Vollverbs ab?
- Lässt sich die informationsstrukturelle Thema-Rhema-Aufteilung modellieren? Ist der in unmarkierten RP-Sätzen topikalisierte Rezipient im Kontext bekannt? Da zur Analyse nur einzelne Sätze herangezogen werden, ist eine Klassifizierung *bekannt/unbekannt* nur durch weitere grammatische Merkmale approximierbar: Wird der Rezipient mittels einer definiten Phrase oder durch ein Pronomen beschrieben, so kann er als im Kontext bekannt betrachtet werden.
- In welcher Reihenfolge tauchen Rezipient und Thema in den Passiv-Varianten auf?

3.3 Mögliche Anwendung: Generierung aus semantischen Repräsentationen

Im Falle eines ähnlichen experimentellen Erfolgs wie in (Bresnan et al., 2001, 2005) sind für die korrekte Wahl der Passiv-Form einige Anwendungsmöglichkeiten denkbar. So kann eine zuverlässige Klassifikation auch zur Text-Generierung aus semantischen Repräsentationen verwendet werden:

Sei p die semantische Repräsentation (Formel 3.2) abstrahiert aus dem Satz *Der Sohn bekommt von der Mutter ein Auto geschenkt*.

$$p : \text{schenken}(\text{die Mutter}, \text{der Sohn}, \text{ein Auto}); \text{präsens, indikativ} \quad (3.2)$$

p ist zwar temporal und modal verankert, trägt jedoch keine Informationen über das Genus Verbi mehr. Es wären also u.a. die folgenden 3 Sätze generierbar:

- (49) a. Die Mutter schenkt dem Sohn ein Auto
 b. Ein Auto wird dem Sohn von der Mutter geschenkt
 c. Der Sohn bekommt von der Mutter ein Auto geschenkt

Sei nun die Vorgabe gegeben, dass der Satz nicht aktiv realisiert werden soll, so blieben noch (49b) und (49c) übrig. Wie lässt sich nun (49c) als der richtige Satz erkennen? Entsprechend dem Klassifikator dieser Arbeit ist es möglich, einen ähnlichen Klassifizierungsmechanismus zu erzeugen, der die generierten Sätze analysiert und mit der realisierten Diathese evaluiert.

Somit bekommt jeder Satz einen Score zugewiesen, nach welchem er gerankt werden kann. Der am höchsten gerankte Satz wird als korrekte Text-Generierung deklariert. Ist die Klassifikation dieser Arbeit erfolgreich, so wird das Generierungsranking in gleicher Weise funktionieren.

4 Das Experiment

Das Hauptziel des Experiments ist es, eine Antwort auf die in (3.2) beschriebene Fragestellung zu finden: Gibt es linguistische Faktoren, die die Auswahl der richtigen Passiv-Variante beeinflussen und kann man diese Faktoren so modellieren, dass eine korrekte Vorhersage möglich ist? Als Zwischenschritt wird versucht, einen Klassifikator zu implementieren, der eine Reihe von Merkmalen, ein *Feature-Set*, auf die wahrscheinlichste Diathese (*Aktiv*, *Vorgangspassiv*, *Rezipientenpassiv*) abbildet. Hierzu werden Daten beschafft und für eine Klassifikation aufbereitet.

Desweiteren wird versucht durch eine manuelle Auswertung dieser Daten viele der in (3.2) aufgeführten Fragen zu beantworten.

Aus dem Erfolg oder Misserfolg der Klassifikation sowie der manuellen Analyse wird ein Ergebnis für die Fragestellung geschlossen. Dieses wird in Kapitel 5 näher beschrieben.

Abbildung 4.1 illustriert den Aufbau des Experiments. Die Kernpunkte sind:

1. Daten-Beschaffung (4.1)
 - a) Extraktion von Sätzen im Rezipientenpassiv (4.1.1)
 - b) Extraktion möglicher Rezipientenpassiv-Verben (4.1.2)
 - c) Extraktion von Sätzen mit den häufigsten Rezipientenpassiv-Verben (4.1.3)
2. Daten-Aufbereitung (4.2)
 - a) Parsing extrahierter Sätze (4.2.1)
 - b) Transfer der Features aus den Parses (4.2.2)
3. Daten-Analyse (4.3)
 - a) Erstellung von Feature-Sets aus den feature-Dateien (4.3.1)
 - b) Auswertung der Feature-Sets mittels eines Klassifikators (4.3.2)
 - c) Manuelle Auswertung der Feature-Sets (4.3.3)

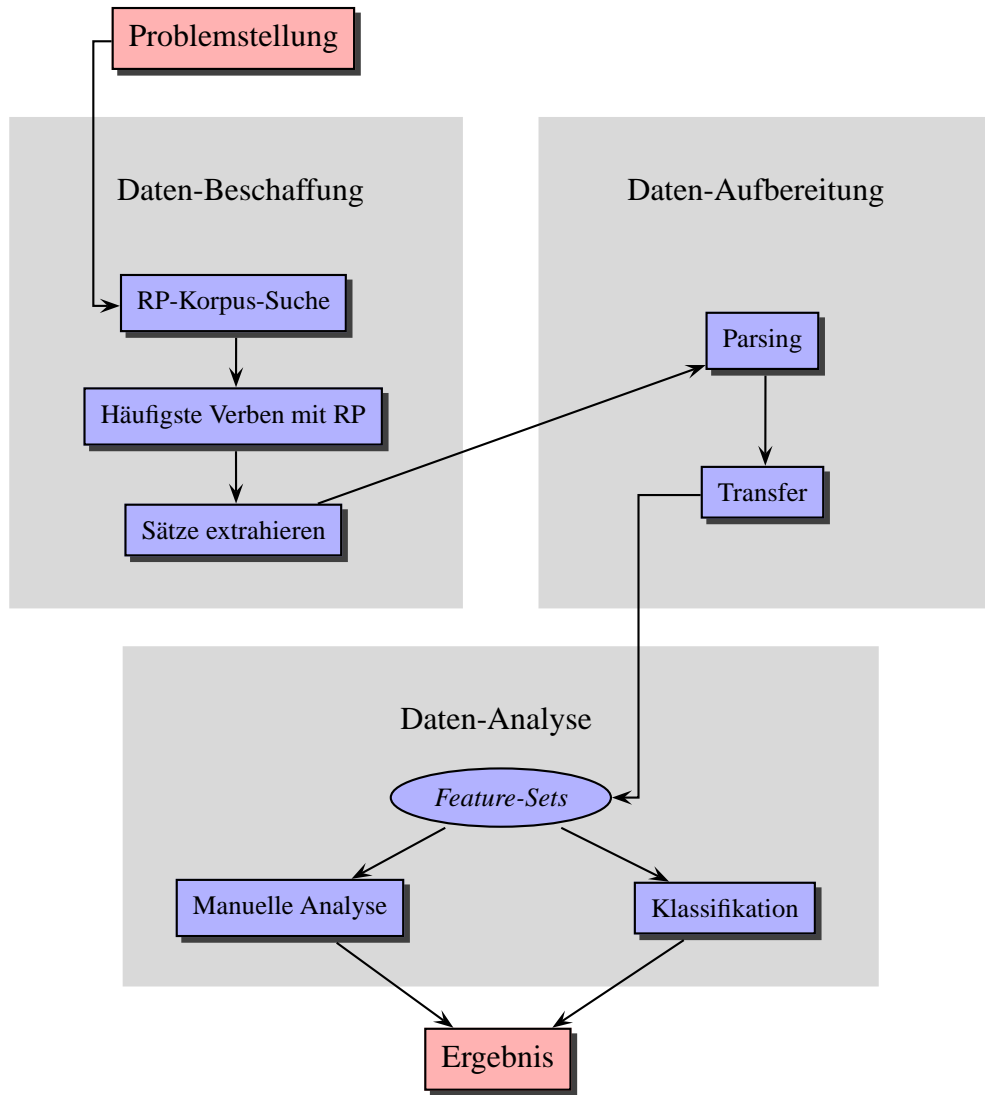


Abbildung 4.1: Aufbau des Experiments

4.1 Die Korpora

„Korpora sind Sammlungen linguistisch aufbereitete(r) Texte in geschriebener oder gesprochener Sprache, die elektronisch gespeichert vorliegen.“ (Carstensen et al., 2009)

Zunächst stellt sich die Frage nach der Datenquelle, die zur Auswertung von Sätzen im Rezipienten- und Vorgangspassiv in Betracht kommt.

Als Korpus-Werkzeug wird das Tool CQP (*Corpus-Query-Processor*) verwendet, welches als eine Komponente der IMS Corpus Workbench (*CWB*) fungiert. Der Vorteil dieses Tools ist, dass eine schnelle Korpus-Suche mittels regulärer Ausdrücke möglich ist (Evert, 2005).

Die Suchanfragen bei CQP sind token-basiert, d.h. jedes Element der Query (die Eingabe der Suchanfrage) beschreibt ein Token. So findet CQP mit (50a) alle Tokens, die mit dem Lemma *bekommen* annotiert sind, z.B.: *bekomme*, *bekommst*, *bekommt*, *bekommen*, *bekam*, *bekäme*, und mit (50b) solche, die mit dem *Stuttgart-Tübingen-Part-of-Speech-Tag* *VVPP* (Vollverb im Partizip-Perfekt) annotiert sind, z.B.: *geschenkt*, *überreicht*, *ausgehändigt*. Durch (50c) wird in CQP nach Ausdrücken bestehend aus diesen beiden Tokens in Folge gesucht.

- ```
(50) a. [lemma="bekommen"];
 b. [stts="VVPP"];
 c. [stts="VVPP"][lemma="bekommen"];
 d. [lemma="(werden | sein | bekommen | erhalten |
 kriegen)"][]* [stts="VVPP"];
```

Sowohl die Tokens als auch die sie beschreibenden Strings können mittels regulärer Ausdrücke modifiziert oder generalisiert werden. So lassen sich Beispielsweise durch (50d) Ausdrücke finden, die aus einem Token bestehen, das mit dem Lemma *werden*, *sein*, *bekommen*, *erhalten* oder *kriegen* annotiert ist, gefolgt von beliebig vielen weiteren willkürlichen Tokens und final einem Token annotiert mit dem Part-of-Speech-Tag *VVPP*.

- ```
(51) ... verurteilt < worden . Javiera und Pablo hatte er Fußbälle geschenkt >, ...
```

Mit dem Query-Zusatz *within s* ermöglicht CQP die Einschränkung der Suchanfrage auf einen Satz. Damit werden Suchergebnisse über Satzgrenzen hinweg verhindert (51).

Mittels dieser Funktionalität lässt sich ein Muster aufbauen, welches eine Rezipientenpassiv-Konstruktion beschreibt.

CQP umfasst Korpora aus verschiedenen Domänen und Sprachen. Für die Extraktion von Rezipientenpassiven sind nur deutsche Korpora zu berücksichtigen. Zur Extraktion dieser Instanzen müssen Verbformen zu *bekommen* sowie Partizipformen von Vollverben erkannt werden. Die Korpora sollten folglich mit Lemma- und Part-of-Speech-Informationen annotiert sein (vgl. (50)). Tabelle 4.1 zeigt die aufgrund dieser Restriktionen in Betracht gezogenen Korpora.

Korpus	Beschreibung
HGC	das „Huge German Corpus“ setzt sich aus mehreren kleinen Korpora zusammen: Frankfurter Rundschau (FR), Stuttgarter Zeitung (STZ), VDI-Nachrichten (VDI), tageszeitung (TAZ), German Law Corpus (GL), Donaukurier (DK) und Computerzeitung (CZ) Größe: 204.813.118 Tokens, 12.223.281 Sätze
SDeWaC-1	dependenzgeparter Ausschnitt aus dem deutschen Web-Korpus DeWaC Teil 1 - Größe: 232.945.959 Tokens, 9.536.124 Sätze
SDeWaC-2	Teil 2 - Größe: 233.332.317 Tokens, 9.635.628 Sätze
SDeWaC-3	Teil 3 - Größe: 200.584.356 Tokens, 12.206.222 Sätze
GUTENBERG-DE	Klassische Literatur in deutscher Sprache in der 9. Ausgabe mit Texten von mehr als 550 Autoren, mehr als 2000 Romanen, Erzählungen und Novellen, tausenden von Gedichten, Fabeln, Märchen und Sagen. Größe: 137.336.680 Tokens, 6.560.627 Sätze
KRENN2000	Größe: 7.723.833 Tokens, 397.210 Sätze
BUNDESTAG	Bundestag-Korpus Debatten aus der Protokollabteilung des Bundestags 1994-1997 Größe: 5.756.188 Tokens, 278.160 Sätze
VERBRECHEN	Größe: 5.749.053 Tokens, 275.013 Sätze
MLCC-DE	Debatten des Europaeischen Parlaments (1992-1994) Größe: 8.945.152 Tokens, 389.793 Sätze
KS-TIGER	Größe: 641.650 Tokens, 36.475 Sätze
PSYNDEX-D	Abstracts von Fachartikeln in Psychologie von der Universität Ulm, Psychologische Beratungsgespräche Größe: 2.546.525 Tokens, 109.008 Sätze
IMSEARCH-LOCAL-DE	Beinhaltet öffentliche sowie lokale Webseiten auf Deutsch. Größe: 519.672 Tokens, 74.279 Sätze
Gesamtgröße aller Korpora	1.040.894.503 Tokens, 51.721.820 Sätzen

Tabelle 4.1: Verwendete Korpora aus CQP; Beschreibung nach IMS-CORPORA (2010)

4.1.1 Extraktion von Sätzen im Rezipientenpassiv

Für die nachfolgende Extraktion wurde auf die Auxiliare *erhalten* und *kriegen* verzichtet und nur das für beide Hilfsverben substituierbare *bekommen* verwendet (vgl. 2.4.2).

Die generelle Form eines Rezipientenpassivs besteht aus dem Auxiliar *bekommen* und aus der partizipialen Form eines Vollverbs. Diese Informationen lassen sich wie in (50) dargestellt mittels der Lemma- und Part-of-Speech-Annotation codieren.

Um Fehler wie in (51) zu vermeiden, muss das korrekte Rezipientenpassiv-Muster die Restriktion *within s* beinhalten. Dies führt zu folgender Query (52):

(52) `[lemma="bekommen"][*][stts="VVPP"] within s;`

Diese Query erlaubt jedoch noch keine Präzedenz des Partizips vor dem Auxiliär, wie es in Verbletz-Sätzen, bei komplexeren periphrastischen Konstruktionen mit infinitem *bekommen* oder bei topikalisierten Verbalphrasen der Fall ist (53).

- (53) a. weil er sein Auto gewaschen bekam
 b. Er hat/wird sein Auto gewaschen bekommen
 c. Gewaschen bekam er sein Auto

Somit wird eine weitere Query für die Fälle in (53) erstellt:

(54) `[stts="VVPP"][*][lemma="bekommen"] within s;`

Eine letzte Problematik, die die Restriktion *within s* nicht erfassen kann, sind unter- oder nebensubordinierte Teilsätze, die mit Konjunktionen, Kommata, Semikolons, Doppelpunkten oder Gedankenstrichen voneinander abgetrennt sind und die jeweils einen der beiden relevanten Elemente (Auxiliär oder Partizip) enthalten (55).

- (55) a. Sie < **bekommt eine schriftliche Einladung - falls sie eingeladen** > wird .
 b. Dort < **bekomme ich mein Eselsdiplom , in dem mir bescheinigt** > wird , dass ...

Um solche Fälle auszuschließen, werden zwischen dem Auxiliär und dem Partizip keines dieser Teilsätze trennenden Zeichen erlaubt. Die daraus resultierenden finalen Queries (56) schließen damit zwar auch RP-Instanzen aus, die korrekt sind wie in (57), jedoch ist in diesem Schritt des Experiments keine absolute Abdeckung erforderlich.

- (56) a. `RP1 = [lemma="bekommen"][stts!="(\$,) | (KO.*) "& word!="-"][*][stts="VVPP"] within s;`
 b. `RP2 = [stts="VVPP"][stts!="(\$,) | (KO.*) "& word!="-"][*][lemma="bekommen"] within s;`
 c. `RP = union RP1 RP2;`

- (57) Er bekommt ein Auto, ein Fahrrad und einen Golfschläger geschenkt

Das Part-of-Speech-Tag „*KO.**“ matcht alle Konjunktionen und das Tag „*\\$,*“ beschreibt satzinterne Interpunktion. Mithilfe von (56c) werden die Ergebnisse beider Queries zu einem Gesamtergebnis fusioniert¹.

Führt man die CQP-Suche mit den Queries in (56) auf den oben genannten Korpora durch, so erhält man die in Tabelle 4.2 aufgelisteten Anzahlen von Vorkommen von Rezipientenpassiven.

Dies verdeutlicht, wie selten die Rezipientenpassiv-Diathese ist. Im Korpus SDEWAC-1 tritt sie mit 0,088 % am häufigsten auf.

¹Da manche Korpora für die Part-of-Speech-Information anstelle des *stts*-Attributs ein *pos*-Attribut verwenden, müssen die Queries hierfür entsprechend modifiziert werden.

Korpus	Anzahl der RPs	Anzahl der Sätze	Anteil
HGC	4.664	12.223.281	0,038%
SDEWAC-1	8.431	9.536.124	0,088%
SDEWAC-2	7.413	9.635.628	0,077%
SDEWAC-3	7.391	12.206.222	0,061%
GUTENBERG-DE	966	6.560.627	0,015%
KRENN2000	189	397.210	0,048%
BUNDESTAG	160	278.160	0,058%
VERBRECHEN	95	275.013	0,035%
MLCC-DE	58	389.793	0,015%
KS-TIGER	9	36.475	0,025%
PSYINDEX-D	4	109.008	0,004%
IMSEARCH-LOCAL-DE	2	74.279	0,003%
Gesamtanzahl	29.382	51.721.820	0,057%

Tabelle 4.2: Anzahl der gefundenen Rezipientenpassive je Korpora

4.1.2 Extraktion möglicher Rezipientenpassiv-Verben

Der nächste Schritt besteht darin, die Vollverben, die partizipial in den gefundenen Rezipientenpassiven vorkommen, als Lemma zu extrahieren und zu zählen, um so die Verben zu finden, die am häufigsten den Rezipientenpassiv realisieren. Alle hierfür benötigten Werkzeuge sind bereits in CQP enthalten.

Zunächst müssen in den Queries in (56) die Vollverben mit sogenannten *Targets* markiert werden, indem vor das Token, das das partizipiale Vollverb repräsentiert, ein @-Zeichen gesetzt wird (58).

- (58) a. `RP1 = [lemma="bekommen"][stts!="(\$,) | (KO.*)"& word!="-"]*@[stts="VVPP"] within s;`
b. `RP2 = @[stts="VVPP"][stts!="(\$,) | (KO.*)"& word!="-"]*[lemma="bekommen"] within s;`

Die in dem fusionierten Ergebnis (56c) enthaltenen Vollverben können nun mithilfe der *Targets* extrahiert und als Lemmata gezählt werden:

- (59) `count RP by lemma on target;`

Tabelle 4.3 zeigt die 24 häufigsten Vollverbanzahlen über die 12 Korpora verteilt. Hierbei fällt auf, dass das häufigste Verb im Rezipientenpassiv, *schenken*, fast 3 mal so häufig vorkommt als das zweithäufigste Verb, *stellen*. Dieser frequentive Unterschied zwischen dem häufigsten und den übrigen Verben bringt *schenken* in einen Sonderstatus. Gründe hierfür könnten die prototypischen Eigenschaften eines dreiwertigen, agentiven Verbs sein sowie die Semantik von *schenken*, die die Konverse zu *bekommen*, *geben*, näher spezifiziert, weshalb der Satz (60a) als ungrammatisch betrachtet wird, da er sich semantisch äquivalent zu Satz (60b) verhält. Hier zeigt sich deutlich, wie stark die Eigensemantik der Auxiliare noch vorhanden ist.

- (60) a. * Er bekommt ein Auto gegeben
 b. Er bekommt ein Auto

Verb	RP-Anzahl	Verb	RP-Anzahl
schenken	2.129	zuteilen	445
stellen	829	servieren	426
bieten gebieten	826	bezahlen	419
anbieten	698	überreichen	408
zuweisen	653	verpassen	394
verleihen	622	vorsetzen	377
vermitteln	618	präsentieren	376
zusprechen	568	zeigen	344
zuschicken	545	drücken	323
sagen	529	auszahlen	316
erstatten	513	anzeigen	312
liefern	513	aushändigen	289

Tabelle 4.3: Anzahl der gefundenen Rezipientenpassive je Vollverb

4.1.3 Extraktion von Sätzen mit den häufigsten Rezipientenpassiv-Verben

Der letzte Schritt der Datenerfassung besteht aus der Extraktion aller Sätze, die eines der häufigsten Vollverben mit Rezipientenpassiv enthalten. Zunächst sollte jedoch berücksichtigt werden, dass es Verben gibt, die per se häufiger auftreten als andere. So führe etwa eine Extraktion aller Sätze mit dem Verb *sagen* zu 1.013.324 Sätzen, wovon lediglich 529 im Rezipientenpassiv realisiert sind. Dies hätte zur Folge, dass die Aufbereitung dieser Daten zeitlich nicht durchführbar und desweiteren kein aussagekräftiges Ergebnis zu erwarten wäre, da dieses Verb mit 0,052 % RP-Anteil kein prototypisches Verb für das Rezipientenpassiv darstellt.

Daher muss zunächst der prozentuale Anteil für jedes der in Tabelle 4.3 aufgestellten Verben berechnet werden (Tabelle 4.4).

Die Extraktion von Sätzen mit allen Verben $\$VERB$ aus Tabelle 4.4 ist in CQP durch (61) ohne weitere Schritte durchführbar.

```
(61) [lemma = "$VERB"] > "saetze.$VERB.txt";
```

4.2 Datenaufbereitung

Nachdem die Daten erfolgreich aus CQP extrahiert wurden, müssen sie zur weiteren Analyse noch aufbereitet werden. Zunächst werden die Sätze geparkt und danach mittels eines Transfer-Skripts in Feature-Repräsentationen übertragen, aus denen die für die Differenzierung relevanten Merkmale extrahiert und zu Feature-Sets zusammengefasst werden können.

\$VERB	RP-Anzahl	Verb-Anzahl	RP-Anteil
zuschicken	545	2.854	19,1%
vorsetzen	377	2.247	16,8%
zuteilen	445	3.789	11,7%
zusprechen	568	5.925	9,6%
aushändigen	289	3.286	8,8%
zuweisen	653	7.831	8,3%
schenken	2.129	40.371	5,3%
servieren	426	8.373	5,1%
erstatten	513	10.200	5,0%
überreichen	408	9.616	4,2%
verleihen	622	33.162	1,9%
Gesamtanzahl	6.975	127.654	5,5%

Tabelle 4.4: Die relativ häufigsten RP-Vollverben

4.2.1 Parsing extrahierter Sätze

Für den Parsing-Schritt ist eine Grammatik notwendig, die die drei zur Diskussion stehenden Diathesen bereits erkennt. Als Parser wurde das XLE-Tool verwendet (Xerox Linguistic Environment). Hierbei handelt es sich um eine Kombination aus einem Tcl-Benutzer-Interface und linguistischen Werkzeugen, die in PARC (*Palo Alto Research Center*) and Grenoble XRCE (*Xerox Research Centre Europe*) entwickelt wurden (XLE-User-Dokumentation, 2010).

Als Grammatik wird die deutsche LFG-Grammatik (*Lexikalisch-Funktionale-Grammatik*, (Falk, 2001)), die in (Rohrer u. Forst, 2006) beschrieben wird, verwendet. Diese wurde im Rahmen des ParGram-Projekts entwickelt (Butt et al., 2002).

Der XLE-Parser überführt jeden Satz in eine C-Struktur, eine Konstituentenstruktur, und eine F-Struktur, eine Attribut-Wert-Matrix, wobei ein Attribut einem Feature oder dem Namen einer grammatischen Funktion entspricht (Falk, 2001).

Desweiteren wird ein Desambiguierungsmodell verwendet, um aus der Menge von möglichen Satzanalysen die plausibelste zu wählen. Hierbei ist ein Fehler aufgetreten, da einige Sätze falsche bzw. unwahrscheinliche Interpretationen zugeordnet bekamen. Satz (62) ist syntaktisch ambig. Sowohl „*Petrus*“ als auch „*eine Leitungsgabe*“ können in dieser Form im Nominativ wie im Akkusativ modelliert sein, also sowohl das Subjekt als auch das Objekt beschreiben. Aus semantischen Gründen ist „*Petrus*“ als Subjekt zu bevorzugen, jedoch weist XLE in diesem und ähnlichen Fällen dem belebten Aktanten die Objekt-Funktion und damit, im Falle des Rezipientenpassivs, die Thema-Rolle zu.

(62) Petrus bekommt eine Leitungsgabe zugesprochen

Diese Analyse ist zwar nicht zufriedenstellend, jedoch sind hiervon lediglich ein Teil der ambigen Sätze betroffen. Ein alternatives Desambiguierungsmodell stand für das Experiment nicht zur Verfügung. Der Vorteil der automatischen gegenüber der manuellen Daten-Analyse ist, dass eine große Menge an Daten evaluiert werden kann, jedoch sind Fehler dieser Art dabei meist unvermeidlich.

Aus den 12 Korpora wurden die 11 häufigsten Verben ermittelt (Tabelle 4.4). Dies ergab 127.654 Sätze, unterteilt in 4 XLE-*testfiles* (für den XLE-Parser lesbare Dateien) mit je ca. 30.000 Sätzen. Diese werden mittels der Grammatik durch XLE geparkt und jeder analysierte Satz wird in eine durchnummerierte *f-struktur*-Datei (nachstehend *fs*-Datei genannt) gespeichert, welche in Prolog (eine Programmiersprache; Name vom Französischen: **P**rogrammation en **L**ogique) codiert ist und als Input für den nachfolgenden Transfer dient.

(63) Er bekommt von seiner Mutter ein Auto geschenkt

Der Baum in Abbildung 4.2 stellt eine C-Struktur dar, die XLE für den Satz (63) ausgibt. Die Matrix in Abbildung 4.3 ist die durch XLE erstellte zugehörige F-Struktur. Der Rezipient wird als pronominales Subjekt in der 3. Person, maskulin, singular deklariert. Das Agens, die präpositionale *von*-Phrase, wird als Wert der Funktion OBL-AG zusammengefasst und das Thema als Akkusativ-Objekt festgehalten. Die Information der Diathese ist in der F-Struktur unter dem Eintrag „PASSIVE bekommen“ vermerkt. Die Grammatik hat das Rezipientenpassiv erkannt.

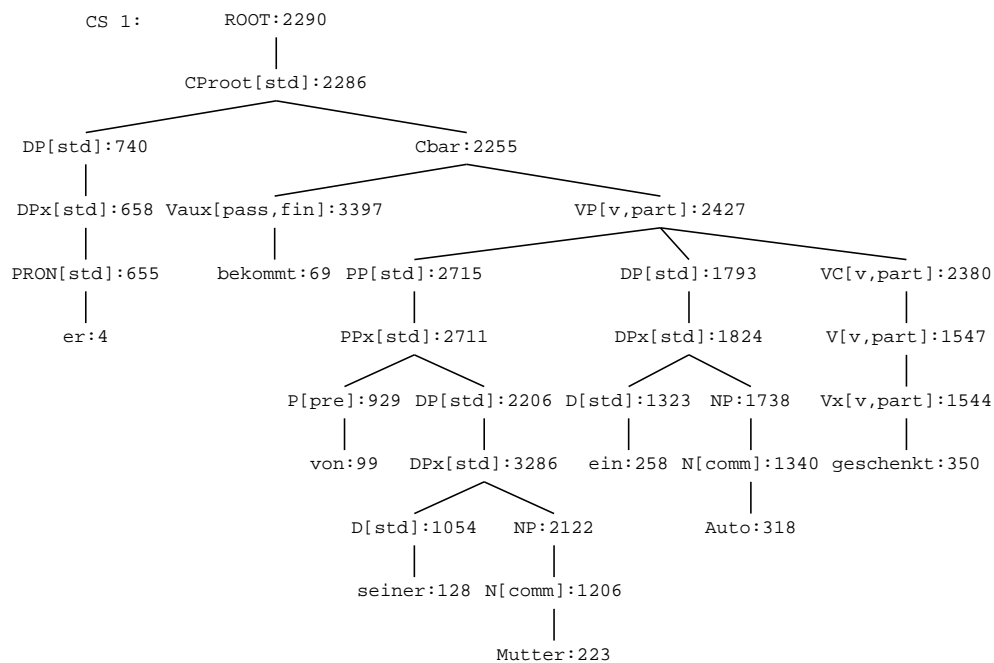


Abbildung 4.2: Beispiel: C-Struktur

Ein Problem bereiteten auch viele Sätze, die aufgrund der komplexen Satzstruktur von der Grammatik nur partiell geparkt werden konnten. Von 127.654 Sätzen wurden nur 57.934 Sätze problemlos als grammatisch analysiert. Demzufolge waren knapp 55% aller Sätze für die LFG-Grammatik „ungrammatisch“. Allerdings stand einerseits durch diese Datenlage eine

"Er bekommt von seiner Mutter ein Auto geschenkt"

	PRED	'schenken<[99:von], [258:Auto], [4:pro]>'	
		4[PRED 'pro']	
	SUBJ	655 NTYPE [NSYN pronoun]	
		658 CASE nom, GEND masc, NUM sg, PERS 3, PRON-FORM sie, PRON-TYPE pers	
		740	
		[PRED 'von<[128:Mutter]>'	
		[PRED 'Mutter'	
		CHECK [_SPEC-TYPE [_COUNT +, _DEF +, _DET attr]]	
		[_INFL strong-det]	
		223	
	OBL-AG	OBJ 1206 NTYPE [NSEM [COMMON count]]	
		2122 NTYPE [NSYN common]	
		128	
		99 1054 SPEC [POSS [PRED 'pro'	
		929 3286 [GEND masc, NUM sg, PERS 3, PRON-FORM ihre]]	
		2711 2206 CASE dat, GEND fem, NUM sg, PERS 3	
		2715 PSEM dir, PTYPE sem	
		[PRED 'Auto'	
		CHECK [_SPEC-TYPE [_COUNT +, _DET attr]]	
		318	
	OBJ	1340 NTYPE [NSEM [COMMON count]]	
		1738 NTYPE [NSYN common]	
		258	
		1323 SPEC [DET [PRED 'eine']	
		1824 [DET-TYPE indef]]	
		1793 CASE acc, GEND neut, NUM sg, PERS 3	
350			
1544			
1547			
2380			
2427	CHECK	[_AUX-FORM {bekommen-pas}]	
69		_VLEX [_AUX-SELECT haben]	
3397		_VMORPH [_PARTICIPLE perfect]	
2255	TNS-ASP	[MOOD indicative, PASS-ASP dynamic, TENSE pres]	
2286	TOPIC	[4:pro]	
2290	CLAUSE-TYPE	decl, PASSIVE bekommen, VTYPE main	

Abbildung 4.3: Beispiel: F-Struktur

relativ große Satzmenge zur Untersuchung, so dass die große Menge an „ungrammatischen“ Sätzen vertretbar war und andererseits waren einige dieser Parse, da partiell analysierbar, noch für die weitere Verwendung zu gebrauchen.

4.2.2 Transfer der Features aus den Parses

Für die weitere Aufbereitung der aus den Korpora extrahierten und geparsten Sätze wird ein XLE-Transfer-Skript verwendet, bestehend aus Feature-Templates, welche Teile der F-Struktur (in einer *fs*-Datei codiert) in eine besser zu extrahierende Form bringen und zudem gewisse Funktions- und Merkmals-Kombinationen generalisieren. Für diese Aufgabe wird ein großer Teil der Templates aus (Forst, 2007) und (Cahill u. Riester, 2009) verwendet.

"Translation of: Er bekommt von seiner Mutter ein Auto geschenkt."

PRED		[[= ['schenken<[99: von], [258: Auto], [1: pro]>']]]
		[['schenken<[128: Mutter], [258: Auto], [1: pro]>']]
		[sverb-schenken_voice-bepassive_feat_arg_AGENT]
		[sverb-schenken_voice-bepassive_feat_arg_RECIPIENT]
		[sverb-schenken_voice-bepassive_feat_arg_THEME]
		[sverb-schenken_voice-bepassive_feat_isCommon_AGENT]
		[sverb-schenken_voice-bepassive_feat_isCommon_THEME]
		[sverb-schenken_voice-bepassive_feat_isIndef_THEME]
		[sverb-schenken_voice-bepassive_feat_isNum_sg_AGENT]
		[sverb-schenken_voice-bepassive_feat_isNum_sg_RECIPIENT]
		[sverb-schenken_voice-bepassive_feat_isNum_sg_THEME]
		[sverb-schenken_voice-bepassive_feat_isPers_3_AGENT]
		[sverb-schenken_voice-bepassive_feat_isPers_3_RECIPIENT]
ADD-PROP		[sverb-schenken_voice-bepassive_feat_isPers_3_THEME]
		[sverb-schenken_voice-bepassive_feat_isPossDef_AGENT]
		[sverb-schenken_voice-bepassive_feat_isPronPersRecipRefl_RECIPIENT]
		[sverb-schenken_voice-bepassive_feat_isPronPers_RECIPIENT]
		[sverb-schenken_voice-bepassive_feat_isPron_RECIPIENT]
		[sverb-schenken_voice-bepassive_feat_isReentrant_AGENT]
		[sverb-schenken_voice-bepassive_feat_isSimpleIndef_THEME]
		[sverb-schenken_voice-bepassive_feat_precedes_AGENT_THEME]
		[sverb-schenken_voice-bepassive_feat_precedes_RECIPIENT_AGENT]
		[sverb-schenken_voice-bepassive_feat_precedes_RECIPIENT_THEME]
		[sverb-schenken_voice-bepassive_feat_subcat_V_AGENT_THEME_RECIPIENT]
		[PRED 'Mutter']
		[CHECK [_SPEC-TYPE [_COUNT +, _DEF +, _DET attr]]]
		[[_INFL strong-det]]
AGENT	223	[NSEM [COMMON count]]
	1221	[NTYPE]
	2140	[NSYN common]
	128	[POSS [PRED 'pro']]
	1069	[[GEND masc, NUM sg, PERS 3, PRON-FORM ihre]]]
	3513	[]]
	2224	[CASE dat, GEND fem, IS_POSS_DEF 16_29, NUM sg, PERS 3]]
		[_AUX-FORM (bekommen-pas)]]
CHECK		[_VLEX [_AUX-SELECT haben]]]
		[_VMORPH [PARTICIPLE perfect]]]
		[PRED 'pro']]
RECIPIENT	1	[CHECK [MORPH [_CAPITAL +]]]
	670	[NTYPE [NSYN pronoun]]]
	673	[]]
	755	[CASE nom, GEND masc, NUM sg, PERS 3, PRON-FORM sie, PRON-TYPE pers]]
		[PRED 'Auto']]
		[CHECK [_SPEC-TYPE [_COUNT +, _DET attr]]]]
		[]]
THEME	318	[NSEM [COMMON count]]]
	1562	[NTYPE]]
	1565	[NSYN common]]
	2607	[]]
	2654	[DET [PRED 'eine']]]
	69	[[DET-TYPE indef]]]]
	3619	[CASE acc, GEND neut, IS_SIMPLE_INDEF 30_38, NUM sg, PERS 3]]]
	2273	[]]
TNS-ASP		[MOOD indicative, PASS-ASP dynamic-, TENSE pres]]
2304		[]]
2311		[CLAUSE-TYPE decl, PASSIVE bekommen, VTYPE main, subcat V_AGENT_THEME_RECIPIENT]]]
	99	[PRED 'von<[128: Mutter]>']]]
	944	[PSEM dir, PTYPE sem]]]
	2938	[]]]
	2942	[]]]

Abbildung 4.4: Beispiel: Transferierte F-Struktur

Um einige der Fragen in (3.2) beantworten zu können, müssen hierzu zunächst grammatische und semantische Features festgelegt werden, die über das Transfer-Skript aus den F- und C-Struktur-Informationen extrahiert werden können. Der Wert eines solchen Features stellt in aller Regel eine semantische Rolle dar.

Das erste Feature beschreibt das Vorkommen eines Argumentes. Die drei für die Untersuchung relevanten Argumente (*Agens*, *Rezipient*, *Thema*) werden hierzu aus den F-Struktur-Informationen herausgefiltert. Ein Subjekt im Aktiv oder ein obliquer Agens (OBL-AG) in einer Passiv-Diathese wird als Agens klassifiziert. Analog wird für Thema und Rezipient verfahren.

Die nächsten Features basieren auf der in (3.1.1) beschriebenen Personen-Hierarchie. So wird jede Person-Beschreibung aus der F-Struktur extrahiert und der jeweiligen semantischen Rolle zugewiesen.

Ob es sich bei dem Argument um ein *Common Noun* (eine Unterklasse von Substantiven, komplementär zu Eigennamen) handelt, könnte auch ein Indikator für die Wahl der Diathese sein. Im Kontrast hierzu wird auch noch das Feature eines Eigennamens verwendet.

Die Frage nach dem Informationsstatus eines Arguments lässt sich durch die Analyse eines einzelnen Satzes nur abschätzen. Da bekannte Nominalphrasen häufig definit oder pronominal auftauchen und unbekannte Nominalphrasen meist nur indefinit beschrieben werden können, sind die Features bzgl. der Definitheit sowie der Pronominalität womöglich sehr hilfreiche Faktoren der Diathesen-Differenzierung.

In (Bresnan et al., 2001) spielt für die Aktiv-Passiv-Alternation das „Behaghel’sche Gesetz der wachsenden Glieder“ eine entscheidende Rolle. Um diese Tendenz auch für die Alternation der Passiv-Varianten auszunützen, wird als weiteres Feature die Komplexität einer Nominalphrase verwendet. Enthält eine Nominalphrase ein Adjektiv oder einen Relativsatz, so wird diese als komplex bewertet.

Eine Nominalphrase kann auch quantifiziert auftreten (vgl. *Alle Männer bekommen mit 18 Jahren ein Auto geschenkt*). Um diese Wirkung auf das Genus Verbi zu erforschen, wird auch hierfür ein Feature beschrieben.

In (2.5) wurde bereits als mögliche Funktion des Rezipientenpassivs die Koordinationsreduktion erwähnt, die in Sätzen wie (43a), wiederholt in (64), auftaucht.

- (64) Er_i fuhr zu schnell, pro_i wurde von der Polizei angehalten und pro_i bekam den Führerschein entzogen

Dieses Phänomen der Koindizierung fällt in der LFG auch unter den Begriff „Reentrance“. Hierfür wird ein weiteres Feature angesetzt.

Ein weiteres Merkmal beschreibt die Reihenfolge, in welchen die Argumente linear im Satz auftauchen. Dieses Feature hat als einziges keine einzelne semantische Rolle, sondern ein Tupel zweier semantischer Rollen, von welchen die erste der zweiten vorangeht.

Der XLE-Transfer bildet mittels des Transfer-Skripts, abermals in Prolog codiert als *voice-properties.pl*, jede *fs*-Datei auf eine *feature*-Datei (nachfolgend *feat*-Datei genannt) ab, die die transferierte Satzanalyse enthält. Hierbei werden für die weitere Bearbeitung die erstellten Features durch den String „*sverb*“ markiert und unter einem gesonderten Attribut ADD-PROP zusammengefasst. Abbildung 4.4 zeigt solch eine modifizierte F-Struktur. Anstelle der grammatischen Funktionen tauchen nun auch die ihnen zugehörigen Argumente als Attribute auf.

Der Transfer-Prozess wird außerhalb der XLE-Umgebung durch (65) gestartet.

```
(65) transfer --inStem [Präfix der fs-Dateien] --inMode fs_file
      --outStem [Präfix der feat-Dateien] --outMode fs_file
      --rules [Transfer-Skript, hier: voice-properties.pl]
```

Damit wäre die Aufbereitung der aus CQP extrahierten Daten abgeschlossen.

4.3 Daten-Analyse

Im letzten Schritt werden die Merkmale aus den *feat*-Dateien extrahiert und statistisch ausgewertet. Wie oft tauchen welche Merkmale bei welcher Diathese auf? Die Auswertung dieser Daten erfolgt auf zwei Wegen: mittels eines Klassifikators (4.3.2) und durch eine manuelle Analyse (4.3.3).

4.3.1 Erstellung von Feature-Sets aus den feature-Dateien

Zunächst werden die relevanten Feature-Informationen aus den *feat*-Dateien extrahiert und zu *Feature-Sets* zusammengefügt. Ein Feature-Set ist ein Tupel bestehend aus einer Menge von Features (grammatischen Merkmalen und deren Werten) und einer Klasse, die der Feature-Menge zugewiesen wird. Entsprechend der Problemstellung werden 3 Klassen definiert: *Aktiv*, *Vorgangspassiv* und *Rezipientenpassiv*.

Aus den *feat*-Dateien werden mithilfe eines *grep*-Befehls (66) alle durch das Transfer-Skript mit dem String „*sverb*“ markierten Zeilen extrahiert und in eine Datei (nachstehend *grep*-Datei genannt) gespeichert.

```
(66) grep -r "sverb-.*_voice-" featFiles/* > grepDatei
```

Diese *grep*-Datei dient als Eingabe für ein Skript, welches aus allen Zeilen die wichtigen Features herausfiltert. Da viele Sätze aus neben- und untergeordneten Teilsätzen bestehen, sind in den Sätzen häufig mehrere Propositionen vorhanden, welche sich durch die F-Struktur-Variable unterscheiden, eine Variable, die als Einbettungsstufe des Satzes verwendet werden kann. So ist jeder Teilsatz einer anderen F-Struktur zugewiesen, welche eine andere Variable trägt, und bildet somit einen für diese Untersuchung eigenständigen Satz. Dies führt zu dem Resultat, dass die endgültige Anzahl an „Sätzen“ ohne Restriktionen höher liegt als die ursprüngliche Gesamtanzahl von 127.654 Sätzen. Die aus der *grep*-Datei extrahierten Features werden zu einem Feature-Set je Satz zusammengefasst und in einer Liste gespeichert.

Für die Klassifikation sind nur Sätze interessant, deren Argument-Rahmen in allen drei Diathesen denkbar sind, da einer bestimmten Diathese fehlende Argumente deren Klassifikation ausschliesse. Für das *Aktiv* ist ein Agens obligatorisch, für das *Vorgangspassiv* ein Thema (abgesehen von seltenen unpersönlichen Passiven) und für das *Rezipientenpassiv* ein Rezipient. Daher werden nur Sätze betrachtet, die alle drei Rollen realisieren. Dadurch wird bei der Erstellung der Feature-Sets ein Filter verwendet, der Sätze herausfiltert, die nicht alle drei Argumente enthalten. Dieser Filter wird optional eingesetzt, da für die Frage nach der Anzahl von fakultativen Ergänzungen (vgl. (3.2)) in der manuellen Analyse keine Restriktionen auf die Argument-Rahmen der Sätze vorhanden sein dürfen.

4.3.2 Auswertung der Feature-Sets mittels eines Klassifikators

Als automatisches Analyseverfahren wird ein Maximum-Entropie-Klassifikator verwendet. Die Grundaufgabe der Klassifikation ist es, aus einer gegebenen Beobachtung Features zu extrahieren, die diese Beobachtung beschreiben, und auf Basis dieser Features die Beobachtung einer Klasse zuzuordnen.

Die Maximum-Entropie-Klassifikatoren gehören zur Gruppe der *exponentiellen* oder *log-linearen* Klassifikatoren. Sie extrahieren aus einer Eingabe eine Menge von Features, kombinieren diese linear durch unterschiedliche Gewichtungen und verwenden diese Summe als Exponenten. Die Formel 4.1 beschreibt diese Berechnung, wobei Z als Normalisierungsfaktor dient (Jurafsky u. Martin, 2008).

$$p(c|x) = \frac{1}{Z} e^{\sum_i w_i f_i} \quad (4.1)$$

Die Klasse c , deren Wahrscheinlichkeit $p(c|x)$ am größten ist, wird als die der Beobachtung x zugehörige Klasse ausgegeben. Weitere Details zur Funktionsweise dieses Klassifikators und seiner Grundlagen (*lineare/logistische Regression*) liefern Jurafsky u. Martin (2008).

Der Klassifikator ist aus der NLTK-Collection (*Natural Language Toolkit*) entnommen, einer Sammlung von Python-Modulen, mit welchen natürliche Sprachen verarbeitet werden können; beschrieben in (Bird et al., 2009). Als Eingabe für den Klassifikator kann das erste Element eines Feature-Sets aus den aufbereiteten Daten verwendet werden. Die Ausgabe des Klassifikators ist die Diathese, welche dem zweiten Element des Feature-Sets entspricht bzw. entsprechen sollte.

Der Klassifikator wird mit einer Menge von Feature-Sets (dem sogenannten *Trainings-Set*) trainiert und mit einer dem *Trainings-Set* disjunkten Menge von Feature-Sets (dem sogenannten *Test-Set*) getestet, indem die Feature-Mengen durch den trainierten Klassifikator klassifiziert werden und mit den tatsächlichen Klassen des *Test-Sets* abgeglichen werden.

Der für die Untersuchung verwendete Klassifikator wird mit 90% der Feature-Sets trainiert und auf 10% der Feature-Sets getestet. Vor der Aufteilung in *Trainings-Set* und *Test-Set* müssen die Feature-Sets mittels eines Zufallsgenerators gemischt werden, da sie durch die *XLE-testfiles*, die *fs*-Dateien, die daraus transferierten *feat*-Dateien und die über die *grep*-Datei aus ihnen extrahierten Features nach Verben sortiert sind.

Die Klassifikation wird in zwei Varianten durchgeführt. Einmal werden die Feature-Sets von allen drei Diathesen als Eingabe verwendet und einmal wird eine Passiv-Klassifikation durchgeführt, bei welcher nur Feature-Sets im VP und RP verwendet werden. Letztere Klassifikation ist in solchen Fällen denkbar, in welchen bereits das Aktiv ausgeschlossen werden kann (vgl. (3.1.1)).

Der Test des Klassifikators, der Abgleich der klassifizierten Diathesen mit den tatsächlichen im *Test-Set*, lässt sich mittels diverser Evaluations-Maße auswerten und interpretieren. Neben der *Accuracy* (die Anzahl korrekter Klassifikationen dividiert durch die Anzahl aller durchgeführten Klassifikationen), welche aufgrund der hohen Baseline (der Anteil der größten Klasse) von Aktiv-Vorhersagen nur bedingt aussagekräftig ist, werden noch die *Precision*, der *Recall* und das *F-Maß* berechnet.

Die Maße *Precision* und *Recall* basieren auf den vier Kategorien *True-Positive (TP)*, *False-Positive (FP)*, *False-Negative (FN)* und *True-Negative (TN)*. Diese Kategorien bestehen aus

zwei Elementen. Das linke Element (*True/False*) beschreibt die Bewertung der Klassifikation. *True* bewertet die Klassifikation als richtig. *False* weist auf eine falsche Klassifikation hin. Das rechte Element (*Positive/Negative*) beschreibt die Ausgabe des Klassifikators. Hierbei fällt auf, dass nur zwei Werte möglich sind. In Bereichen wie Information Retrieval reichen diese zwei Klassen aus (relevantes vs. irrelevantes Dokument). Im Fall der Diathesen-Klassifikation muss die Drei-Diathesen-Klassifikation jedoch auf zwei Werte heruntergebrochen werden. Dies wird dadurch realisiert, dass die 4 Kategorien relativ zu einer Diathese X betrachtet werden. *Positive* relativ zu einer Diathese X meint nun, dass die jeweilige Diathese X klassifiziert wurde. *Negative* meint, dass eine der anderen zwei Diathesen (Y oder Z) klassifiziert wurde. Damit sind auch die Evaluationsmaße *Precision* und *Recall* relativ zur Diathese X berechnet. *Precision* ist die Anzahl aller richtigen X -Klassifikationen dividiert durch die Anzahl aller getätigten X -Klassifikationen (Formel 4.2) und misst, wie präzise der Klassifikator „ins Schwarze trifft“. *Recall* ist die Anzahl aller richtigen X -Klassifikationen dividiert durch die Anzahl aller X -Diathesen im Test-Set (Formel 4.3) und misst, wieviel der X -Diathesen vom Klassifikator „erwischt“ wurden. Ein drittes Maß, das *F-Maß* oder der *F-Score*, kombiniert *Precision* und *Recall* zu einem gemeinsamen Wert, der das harmonische Mittel der beiden Maße darstellt (Formel 4.4).

$$\text{Precision} = \frac{TP}{TP + FP} \quad (4.2)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (4.3)$$

$$\text{F-Maß} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4.4)$$

Eine nützliche Darstellungsweise bietet die *Confusion-Matrix*. Bei drei und mehr Klassen, so auch bei der Drei-Diathesen-Klassifikation, stellt sich häufig die Frage, womit eine Klasse am häufigsten verwechselt wurde. Um herauszufinden, welche Diathese die größte Verwechslungsgefahr für das Rezipientenpassiv birgt, wird einem weiteren Python-Modul aus NLTK zwei geordnete Listen von Labels (Diathesen) übergeben, wovon eine als „Gold-Standard“ alle korrekten Labels enthält und die andere die Ausgabe des Klassifikators.

Die *Confusion-Matrix* ist eine $N \times N$ -Matrix für N Klassen, wobei auf der Diagonalen die Anteile an korrekten Klassifizierungen stehen und abseits der Diagonalen die Fehler.

4.3.3 Manuelle Auswertung der Feature-Sets

Neben der automatischen Datenauswertung durch einen Klassifikator werden die Daten noch manuell analysiert. Nur hierdurch können alle in (3.2) aufgeworfenen Fragen beantwortet werden. Ziel ist es, interessante Werte zu finden, die weitere Hinweise auf die Diathesen-Alternation geben. Hierfür wurde ein weiteres Python-Skript geschrieben, welches als Eingabe, analog zum Klassifikator, eine Menge an Feature-Sets nimmt und die unterschiedlichen Feature-Value-Paare relativ zur zugehörigen Diathese zählt. Es werden zwei Feature-Set-Mengen als Eingabe verwendet. Einmal werden alle Feature-Sets berücksichtigt, so dass etwa der Anteil an fakultativen Ergänzungen je Diathese berechnet werden kann, und einmal werde nur Feature-Sets berücksichtigt, deren verbaler Argumentrahmen nur aus den Rollen

Agens, *Rezipient* und *Thema* besteht, so dass etwa die Präzedenz eines Rezipienten vor einem Thema auch nur bei den Sätzen evaluiert wird, die diese Argumente besitzen.

Die manuelle Auswertung erfolgt über zwei Tabellen für absolute (Formel 4.5) und relative (Formel 4.6) Häufigkeit der Feature-Value-Paar-Vorkommen gegeben eine Diathese.

$$f((feature, value) \cap diathese) \quad (4.5)$$

$$P((feature, value) | diathese) = \frac{f((feature, value) \cap diathese)}{f(diathese)} \quad (4.6)$$

Für die nachfolgende Beschreibung der Ergebnisse werden lediglich die relativen Häufigkeiten vorgestellt. Desweiteren werden Schlussfolgerungen daraus gezogen: Ist das Experiment gelungen und wurde die Problemstellung gelöst?

5 Ergebnisse und deren Evaluation

In diesem Kapitel werden zunächst die Ergebnisse des Klassifikations-Experimentes vorgestellt und erläutert. Dieses wurde in zwei Varianten durchgeführt. Einmal wurden alle Diathesen klassifiziert („Drei-Diathesen-Klassifikation“) und einmal nur die zwei zu untersuchenden Passiv-Varianten („Passiv-Klassifikation“). Anschließend werden weitere Fragen aus der Fragestellung (3.2) auf Basis der manuellen Datenanalyse beantwortet. Abschließend werden aus den Ergebnissen und Antworten Schlussfolgerungen auf die Problemstellung gezogen: Inwiefern ist die Problemstellung gelöst worden? Welche positiven und welche negativen Ergebnisse wurden festgestellt?

5.1 Ergebnisse der Auswertung

Ergebnisse der Drei-Diathesen-Klassifikation

Verwendung	Diathese	Feature-Set-Anzahl
Training (90 %)	Aktiv	51.652
	Vorgangspassiv	1.649
	Rezipientenpassiv	1.080
	Alle Diathesen	54.381
Testing (10 %)	Aktiv	5.716
	Vorgangspassiv	198
	Rezipientenpassiv	128
	Alle Diathesen	6.042
Insgesamt (100 %)	Aktiv	57.368
	Vorgangspassiv	1.847
	Rezipientenpassiv	1.208
	Alle Diathesen	60.423

Tabelle 5.1: Anzahl an Feature-Sets für die Klassifikation

Tabelle 5.1 zeigt die Anzahl an Feature-Sets, die für die Klassifikation verwendet wurden, unterteilt in Trainings-Set, Test-Set und gesamte Feature-Set-Menge. Bereits diese Zahlen zei-

gen, wie privilegiert die Aktiv-Diathese gegenüber den beiden Passiv-Formen ist. Dies hat allerdings zur Folge, dass die Baseline für Aktiv-Vorhersagen mit 94.6% bereits relativ hoch liegt.

Abbildung 5.1 illustriert die Diathesen-Verteilung in den Feature-Sets.

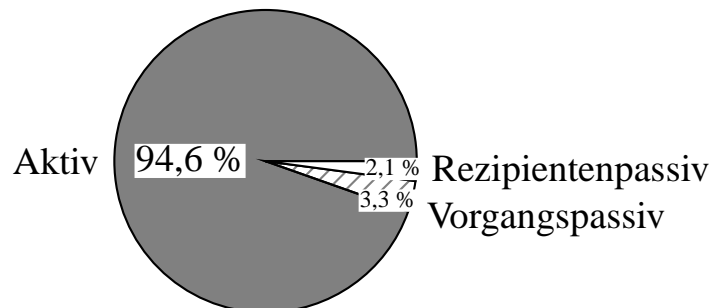


Abbildung 5.1: Diathesen-Verteilung bei *Agens-Rezipient-Thema*-Beschränkung

Die Accuracy (67) zeigt trotz dieser hohen „Mindestqualität“ eine deutliche Verbesserung in der Vorhersage der Diathesen.

(67) **Accuracy: 98,05%** (*Baseline* + 3,45%)

Die Klassifikation ergab noch folgende Evaluationsmaße:

Evaluationsmaß	Diathese	Score
Precision	Aktiv	99,4%
	Vorgangspassiv	73,7%
	Rezipientenpassiv	71,8%
Recall	Aktiv	99,7%
	Vorgangspassiv	76,3%
	Rezipientenpassiv	57,8%
F-Maß	Aktiv	99,5%
	Vorgangspassiv	74,9%
	Rezipientenpassiv	64,1%

Tabelle 5.2: Precision, Recall und F-Maß der Drei-Diathesen-Klassifikation

Diese Werte zeigen einen durchaus positiven Trend. So werden über die Hälfte aller RP-Instanzen im Test-Set erkannt und nur 28,2% der RP-Klassifikationen sind falsch (71,8% *Precision*). Dies deutet bereits auf eine gute Modellierung hin. Betrachtet man die F-Maße, so wird deutlich, dass das Rezipientenpassiv als seltenstes Phänomen auch den niedrigsten Wert

aufweist. Der Recall-Wert von 57% konstatiert jedoch auch eine hohe Verwechslungsrate des RPs. Womit das Rezipientenpassiv am häufigsten verwechselt wurde, zeigt die Confusion-Matrix.

Die Confusion-Matrix zur Drei-Diathesen-Klassifikation ist in Abbildung 5.2 dargestellt.

$$\begin{pmatrix} & \textit{Aktiv} & \textit{VP} & \textit{RP} \\ \textit{Aktiv} & < 94,3\% > & 0,2\% & 0,1\% \\ \textit{VP} & 0,4\% & < 2,5\% > & 0,4\% \\ \textit{RP} & 0,2\% & 0,7\% & < 1,2\% > \end{pmatrix}$$

Abbildung 5.2: Confusion-Matrix der Drei-Diathesen-Klassifikation

Aus dieser Confusion-Matrix lassen sich folgende Informationen ablesen. Nur 0,3% der Aktiv-Sätze werden falsch klassifiziert und zwar 0,2% als Vorgangspassiv und 0,1% als Rezipientenpassiv. Eine Verwechslung des Aktivs mit dem RP ist daher nahezu ausgeschlossen.

3,3% der Sätze im Test-Set sind im VP realisiert, wobei 0,8% falsch klassifiziert werden, zu jeweils 0,4% als Aktiv und RP. Eine stärkere Neigung, das VP mit dem RP anstelle des Aktivs zu verwechseln, ist hier nicht zu erkennen.

Entsprechend des Recalls für den Rezipientenpassiv sind etwas mehr als die Hälfte der RP-Instanzen korrekt (1,2% zu 0,9%). Hierbei fällt auf, dass nur 0,2% als Aktiv klassifiziert werden, jedoch 0,7% als Vorgangspassiv. Diese starke Verwechslungs-Tendenz des Rezipientenpassivs mit dem Vorgangspassiv wird in nachfolgender Passiv-Klassifikation fokussiert.

Ergebnisse der Passiv-Klassifikation

Obige Ergebnisse gehen von einer möglichen Aktiv-Klassifikation aus. Diese könnte jedoch bereits durch einen Soft-Constraint wie z.B. in (Bresnan et al., 2001) ausgeschlossen sein. Desweiteren hat die Drei-Diathesen-Klassifikation eine starke Tendenz zur Interpretation der RP-Instanzen als Vorgangspassive gezeigt und keine sehr große Ähnlichkeit zwischen Rezipientenpassiv und Aktiv offengelegt. Um die beiden Passiv-Diathesen direkt gegenüberzustellen, wird eine binäre Passiv-Klassifikation durchgeführt. Hierfür werden die Feature-Sets für das Training und den Test auf Vorgangs- und Rezipientenpassive beschränkt. Die Einschränkung auf Sätze mit *Agens*, *Rezipient* und *Thema* bleibt zunächst erhalten.

Die Accuracy (68) wirkt im Vergleich zur Accuracy der Drei-Diathesen-Klassifikation (67) zunächst schwächer, jedoch liegt die Baseline der Vorgangspassiv-Vorhersagen mit 61 % deutlich tiefer als die Baseline der Klassifikation aller drei Diathesen.

(68) **Passiv-Accuracy: 81,31%** (*Baseline* + 20,31%)

Ein gutes Ergebnis liefern auch *Precision*, *Recall* und *F-Maß*. Diese liegen in der reinen Passiv-Klassifikation alle über 80% bzw. 70%:

Wie auch bei der Drei-Diathesen-Klassifikation ist hier der Recall des Rezipientenpassivs am schwächsten. Allerdings liegt dieser nun bei 73,1%, also werden 26,9% der RP-Instanzen als VP klassifiziert. Diese Klassifikation bringt folglich eine Verminderung der RP-als-VP-Verwechslungen um rund 6,4% (vgl. $\frac{0,7\%}{2,1\%}$ in Abbildung 5.2). Ist eine Klassifikation als Aktiv also nicht möglich, so werden mehr RP-Instanzen korrekt klassifiziert. Dies lässt allerdings

Evaluationsmaß	Diathese	Score
Precision	Vorgangspassiv	83,4%
	Rezipientenpassiv	77,7%
Recall	Vorgangspassiv	86,6%
	Rezipientenpassiv	73,1%
F-Maß	Vorgangspassiv	85%
	Rezipientenpassiv	75,3%

Tabelle 5.3: Precision, Recall und F-Maß der Passiv-Klassifikation

nicht den Eindruck schwinden, dass der Klassifikator eine starke Tendenz aufweist, viele Rezipientenpassive als Vorgangspassive zu klassifizieren. Viele Merkmalskombinationen, die der Klassifikator dem VP zugeordnet hat, gehören also zur Klasse der Rezipientenpassive.

Die Confusion-Matrix zur Passiv-Klassifikation ist in Abbildung 5.3 dargestellt. Sie bringt, da es sich hierbei um eine binäre Klassifikation handelt, zwar keine neuen Informationen, stellt aber die *Accuracy*-, *Precision*- und *Recall*-Werte aus einem anderen Blickwinkel dar.

$$\begin{pmatrix} & VP & RP \\ VP & < 52,8\% > & 8,2\% \\ RP & 10,5\% & < 28,5\% > \end{pmatrix}$$

Abbildung 5.3: Confusion-Matrix der Passiv-Klassifikation

So ergibt sich aus der Diagonalen die Aufteilung des Accuracy-Werts und aus dem Verhältnis des Diagonal-Wertes zur Summe der Spalte bzw. Zeile der *Precision*- bzw. *Recall*-Wert.

Die Summen der Zeilen ergeben die jeweiligen Passiv-Anteile im Test-Set. Dies zeigt auch, dass das Vorgangspassiv mit 61% deutlich häufiger auftritt als das Rezipientenpassiv und somit mehr Kombinationsmuster für das VP als für das RP trainiert wurden. Eine Verwechslung des Rezipientenpassivs mit dem Vorgangspassiv ist daher auch wahrscheinlicher als bei gleicher Häufigkeit der beiden Diathesen.

Löste man nun die Einschränkung auf die Subkategorisierung, etwa durch den Wegfall eines obligatorischen Agens, so würde sich zwar eine Abgrenzung der Passiv-Varianten bzgl. des Vorhandenseins eines Agens bieten, jedoch nimmt die Anzahl an Vorgangspassiven deutlich stärker zu. Ist der VP-Anteil bei der *Agens-Rezipient-Thema*-Restriktion noch bei 61%, so liegt dieser bei einer *Rezipient-Thema*-Restriktion bereits bei 74,7% und ganz ohne Argument-Restriktionen bei 84%. Und dies führt wiederum zu geringeren Recall-Werten des Rezipientenpassivs. Tabelle 5.4 zeigt überdies, dass trotz schwindenden Recalls der Precision-Wert relativ konstant bleibt. Das heißt, die bisherigen Merkmalskombinationen sind für das Rezipientenpassiv (zu rund 75%) charakteristisch, jedoch könnte es noch mehr Merkmale oder Merkmals-Muster für das Rezipientenpassiv geben, welche der Klassifikator noch nicht erfasst hat und somit diese Feature-Sets fälschlicherweise dem Vorgangspassiv zuordnet.

Subkategorisierungs-Filter	Precision	Recall
<i>Agens-Rezipient-Thema</i>	77,7%	73,1%
<i>Rezipient-Thema</i>	75,7%	42,3%
Keine Restriktion	74,1%	39%

Tabelle 5.4: Precision, Recall des RPs relativ zum Subkat-Filter

Ergebnisse der manuellen Analyse

Nachfolgend werden die Ergebnisse der manuellen Analyse (Tabelle 5.5 und 5.6) betrachtet und die in (3.2) aufgestellten Fragen beantwortet.

Hierbei bleibt zu beachten, dass die *Argument*-Features auf allen in (4.3.1) erstellten 140.449 Feature-Sets ausgewertet wurden. Die übrigen Features beziehen sich auf 60.423 Feature-Sets, die auf den Subkategorisierungsrahmen *<Agens, Rezipient, Thema>* beschränkt sind.

Bei den Werten des *Argument*-Features zeigt sich, dass das *Agens* mehr als doppelt so häufig im RP realisiert wird als im VP, obwohl beide *Passiv*-Varianten das *Agens* nur fakultativ nennen. Dies belegt, wie zu Beispiel (42), wiederholt in (69), erklärt wird, dass die *Agens*-Phrase im Rhema des Rezipientenpassivs eine größere Bedeutung einnimmt als im Rhema des Vorgangspassivs. Intuitiv wirkt ein *agensloser* Satz für viele Sprecher im Vorgangspassiv weniger ungrammatisch als im Rezipientenpassiv (70).

- (69) a. Der Karl bekommt von der Karla das Formular ausgefüllt.
b. Das Formular wird dem Karl von der Karla ausgefüllt.
- (70) a. Ihm wird ~~von seiner Mutter~~ ein Auto geschenkt
b. ? Er bekommt ~~von seiner Mutter~~ ein Auto geschenkt

Interessant scheint auch, dass der Rezipient in allen drei Diathesen sehr häufig vorkommt. Betrachtet man jedoch die Verben in den Feature-Sets (vgl. Tabelle 4.4), so handelt es sich ausschließlich um ditransitive Verben, welche per se ein *Dativ*-Objekt regieren. Der Rezipient taucht aufgrund der unprominenten Funktion im VP seltener auf als im RP. Dort taucht er quasi immer auf. Dies liegt daran, dass er das Subjekt konstituiert und kein unpersönliches Rezipientenpassiv möglich ist. Dieses ist jedoch als *Vorgangspassiv* möglich, weshalb das Thema nur in 89,2% der VP-Fälle vorkommt. Analog zum Rezipient kommt auch das Thema stets im RP vor. Seltene themalose Konstruktionen wie (71) sind in den Daten offensichtlich nicht vorgekommen.

- (71) Er bekommt vom Publikum applaudiert

Bei den *Personen*-Merkmale wurde das Thema nicht betrachtet, da es im Falle der 1. und 2. Person fast immer bei 0% lag und im Falle der 3. Person bei 100%. Dies liegt darin begründet, dass das Thema höchst selten belebt ist und damit nicht in der 1. oder 2. Person auftritt (72).

- (72) ? Er bekommt mich/dich geschenkt

Das Feature zum *Common Noun* zeigt, dass das Subjekt der jeweiligen Diathese am seltensten ein *Common Noun* ist, wobei diese Tendenz für das Thema im VP am schwächsten ausfällt. Eine ähnliche Tendenz zeigt das *Eigennamen*-Feature. Jedoch taucht der Rezipient als Eigenname am seltensten im Aktiv auf (und nicht im RP).

Feature	Rolle	Aktiv	VP	RP
Argument	Agens	81,8%	9,6%	21%
	Rezipient	55,9%	54,3%	97,8%
	Thema	76,3%	89,2%	97,8%
1./2. Person	Agens	9,4%	0,4%	3,6%
	Rezipient	13,9%	7,9%	16,3%
3. Person	Agens	73,7%	95,1%	95,3%
	Rezipient	84,1%	89,9%	74,5%
Common Noun	Agens	30,1%	74,2%	70,9%
	Rezipient	44,9%	43,9%	29,6%
	Thema	67,3%	60,1%	65%
Definit	Agens	15,9%	49,3%	38,2%
	Rezipient	26,6%	31,1%	13,7%
	Thema	18,4%	31,9%	22,2%
Indefinit	Agens	3,8%	3,5%	8,1%
	Rezipient	2,2%	2,2%	6%
	Thema	16,7%	9,9%	23,3%
Ohne Def.-Merkmal	Agens	7,8%	18,3%	10,3%
	Rezipient	7,5%	9,6%	7%
	Thema	22,6%	11,9%	12,9%

Tabelle 5.5: Wichtigste Features bei manueller Auswertung der Daten (1)

Überraschend ist die Distribution der definiten und indefiniten Rezipienten. Hier zeigt sich, dass der Rezipient im Rezipientenpassiv verglichen mit Aktiv und VP häufiger indefinit auftaucht und seltener definit. Diese Tendenz zeichnet einen leichten Widerspruch ab zur Thema-Rhema-Struktur, nach der das Rezipienten-Subjekt in unmarkierter Vorfeld-Position das informationsstrukturelle Thema darstellt und somit im Kontext bekannt ist. So sollte der Rezipient auch am häufigsten im RP bekannt und u.a. als definit codiert sein. Wie in (4.2.1) über das Desambiguierungsmodell von XLE bereits beschrieben, kann dies jedoch auch auf die mangelhafte Desambiguierung bei ambigen Sätzen zurückzuführen sein.

Eine weitere Codierungsmöglichkeit für bekannte Referenten stellt das Pronomen dar. Das *Pronomen*-Feature untermauert die informationsstrukturelle Theorie, nach der die kanonische Subjektposition in V2-Sätzen für das informationsstrukturelle Thema vorgesehen ist. Alle Subjekte der drei Diathesen weisen den höchsten Pronomen-Anteil auf. Hierbei zeigt sich, dass fast die Hälfte aller Rezipienten im RP pronominal sind. Beim VP sind nur knapp ein Drittel der Rezipienten Pronomen.

In (73) sind hierzu drei Beispiele von pronominalen Rezipienten-Subjekten und deren Vorwähnungen aufgelistet.

- (73) a. SDeWaC-1 : *Viele haben mich gefragt, wann es eine Neuauflage geben wird, berichtete Albrecht Mayer unmittelbar nach Beendigung des Rennens gegenüber dem Gäubote. Von den Fans bekam er zum Abschied ein Tablett mit Kuchen in Form eines Handschuhs mit Baseball darin und eines Schlägers überreicht.*
- b. VERBRECHEN : *Erst der englische Seefahrer Sir Walter Raleigh trug mit zur Trendwende von der Heilpflanze zur Rauchkultur bei. 1586 bekam er von Freunden Tabak aus Virginia geschenkt...*
- c. SDeWaC-3 : *Der Beklagte berief sich darauf, dass er die beanspruchten Werke vom Künstler gekauft habe bzw. geschenkt bekommen habe.*

Das Komplexitäts-Merkmal reflektiert Behaghels „Gesetz der wachsenden Glieder“, indem die Argumente, die in der jeweiligen Diathese das Subjekt realisieren, am seltensten komplex sind. So ist der Rezipient im RP um 4,1% seltener komplex als im VP.

Der Rezipient ist im RP häufiger quantifiziert als im VP. Dieses Feature lässt sich ebenfalls gut als Abgrenzungskennzeichen verwenden.

Hinsichtlich der Präzedenz der drei Argumente erkennt man die typische Grundposition des Subjekts vor allen anderen. Im Rezipientenpassiv geht der Rezipient der Thema-Rolle voran, im Vorgangspassiv ist es umgekehrt. Das Agens steht im Aktiv am häufigsten vor den anderen Rollen. Weiter zeigt sich, dass das Agens im RP häufiger vor dem Thema realisiert wird (43,5%) als vor dem Rezipient im VP (26,3%), wie es auch in Beispiel (69) der Fall ist.

Das Reentrant-Merkmal zeigt eindeutig einen Trend dazu, dass der Rezipient am häufigsten im RP bei einer Koindizierung, etwa bei der Koordinationsreduktion, verwendet wird. Bei einer Satzkoordination mit einem Subjekt im 1. (aktiven) Konjunkt und einer koreferenten Rezipienten-Rolle im 2. Konjunkt wird letztere also häufig als Subjekt realisiert und kann damit getilgt werden (vgl. hierzu (2.5)).

In (74) sind aus diversen Korpora weitere Beispiele solcher Koordinationsreduktionen aufgelistet.

- (74) a. SDeWaC-1 : *Volker Jung hatte am 4.4. Geburtstag und bekam von Peer Steinbrück einen großen Blumenstrauß geschenkt.*
- b. SDeWaC-3 : *Die Stadt bestimmt durch eine Registrierung wer in St. Petersburg wohnen darf und eine Unterkunft zugewiesen bekommt.*
- c. SDeWaC-2 : *Thomas holt inzwischen noch ein paar Kekse und bekommt als „Anerkennung“ dafür dieses Amt zugesprochen.*

Tabelle 4.4 zeigt, dass die Verben mit dem größten RP-Anteil alle ditransitive Verben sind, die alle möglichen Facetten des *gebens*, der Konversen zu *bekommen*, beschreiben. Offenbar

handelt es sich bei der lexikalischen Semantik des Vollverbs durchaus um ein ausschlaggebendes Kriterium für die Entscheidung über die Passiv-Variante. Eine mögliche Ausnutzung dieses Faktors, der in dieser Arbeit nicht in Betrachtung gezogen werden kann, könnte die GermaNet-Datenbank oder ähnliche ontologische Systeme darstellen, welche eine Relation des Vollverbs zu *geben* beschreiben könnte. So verwendete Bresnan et al. (2005) auch semantische Verbklassen zur statistischen Analyse.

Feature	Rolle	Aktiv	VP	RP
Eigennamen	Agens	7,9%	12,5%	13,6%
	Rezipient	7,4%	15,1%	10,5 %
	Thema	6,4%	6%	10,3%
Komplex	Agens	14%	22,7%	17,4%
	Rezipient	14,3 %	14,9%	10,8%
	Thema	29,1%	23%	23,3%
Pronomen	Agens	39,6%	3,2%	9,5%
	Rezipient	40%	32,3%	45,7%
	Thema	17,7%	24,9%	17,8%
Quantifiziert	Agens	1,1%	0,9%	0,4%
	Rezipient	1,7%	0,9%	2,6%
	Thema	4%	2,9%	2,9%
Präzedenz	Agens > Rezipient	64,4%	26,3%	18,3%
	Agens > Thema	65,6%	29,5%	43,5%
	Rezipient > Agens	22%	69,2%	75,7%
	Rezipient > Thema	64,7%	39,7%	63,7%
	Thema > Agens	15,5%	50,5%	45,9%
	Thema > Rezipient	22%	37,6%	19,8%
Reentrant	Rezipient	0,8%	0,6%	10,3%
	Thema	1,9%	4,7%	1,7%

Tabelle 5.6: Wichtigste Features bei manueller Auswertung der Daten (2)

5.2 Schlussfolgerungen

Das Rezipientenpassiv und das Vorgangspassiv gehören zur Klasse der Diathesen, d.h. sie beschreiben dieselbe Proposition. Die dieser Arbeit zugrundeliegende Fragestellung war, ob es möglich ist, trotz dieser semantischen Äquivalenz mittels statistischer Methoden eine Vorhersage über die richtige Diathese zu treffen. Gibt es linguistische Merkmale, mittels derer eine Modellierung des Rezipientenpassivs möglich ist?

Die in (5.1) diskutierten Ergebnisse des in Kapitel 4 beschriebenen Experiments zeigten durchweg eine positive Antwort auf diese Fragestellung. Die Drei-Diathesen-Klassifikation erreicht trotz hoher Baseline eine Accuracy von über 98%. Jedoch wird durch einen nur mäßigen Recall-Wert eine große Tendenz zur Verwechslung des Rezipientenpassivs mit dem Vorgangspassiv deutlich. Mit dem Aktiv wird das RP selten verwechselt, was aus der Confusion-Matrix (Abbildung 5.2) hervorgeht. Bei der reinen Passiv-Klassifikation liegt die Accuracy mit über 81% auf einem vielversprechenden Wert, der jedoch auch gewisse Defizite andeutet: Vernachlässigt man die Subkategorisierungsbeschränkungen der Passiv-Klassifikation, so steigt der VP-Anteil und der RP-Anteil sinkt. Dies hat zur Folge, dass der Recall-Wert analog zur steigenden VP-Baseline sinkt. Die Precision bleibt jedoch konstant bei etwa 75%. Dies zeigt, dass die trainierten Merkmale durchaus für eine verlässliche Vorhersage von 75% Genauigkeit verwendet werden können.

Für eine Implementierung zum Zwecke eines Generierungsranking (vgl. (3.3)) sollte durch andere Features, z.B. aus der lexikalischen Semantik, versucht werden, das Ergebnis noch weiter zu verbessern.

Es können auch noch weitere Constraints für die Diathesen-Alternation maßgeblich sein, die in diesem Experiment nicht berücksichtigt worden sind.

Auch ist die Möglichkeit zu berücksichtigen, dass es Fälle gibt, für die kein eindeutiges Ergebnis gefunden werden kann. Dies ließe sich mittels Human-Judgement-Experimenten herausfinden. Gibt es RP-Instanzen, die von Menschen unterschiedlich beurteilt werden? So könnten Menschen auch unterschiedliche stilistische Präferenzen haben, welche sich in ihrer Grammatikalitätsbeurteilung oder spontanen Diathesen-Wahl niederschlagen würden. Stil-Merkmale wären eine weitere Ergänzungsmöglichkeit. Ein Faktor könnte auch sein, wie spontan eine Äußerung ist. Hierfür ließe sich z.B. ein Merkmal einführen, welches gesprochene und geschriebene Sprache differenziert.

6 Zusammenfassung

In dieser Studie wurde die Verwendung des Rezipientenpassivs untersucht. Konstruktionen mit Thema und Rezipient können i.d.R. sowohl als Vorgangspassiv als auch als Rezipientenpassiv realisiert werden. Handelt es sich folglich um reine Willkür oder gibt es Faktoren, die eine Rezipientenpassiv-Verwendung gegenüber der Vorgangspassiv-Verwendung attraktiver machen?

Zunächst wurden die drei häufigsten Passiv-Formen, das Vorgangspassiv, das Zustandspassiv und das Rezipientenpassiv auf Basis von diverser linguistischer Literatur untersucht. Das Vorgangspassiv wird durch das Auxiliar *werden* und ein partizipiales Vollverb gebildet. Die Form der anderen beiden Passiv-Varianten unterscheidet sich primär in der Wahl des Auxiliars (*sein* beim Zustandspassiv und *bekommen* beim Rezipientenpassiv). Die vorgangspassivierbaren Vollverben müssen in aktivischer Form einen Agens fordern. Diese Restriktion gilt auch für das Rezipientenpassiv. Im Vorgangspassiv bildet die Thema-Rolle das Subjekt, welches im Rezipientenpassiv durch den Benefizienten realisiert wird.

Eine mögliche Funktion der unterschiedlichen Passiv-Varianten liegt in der Informationsstruktur. Im Kontext gegebene Referenten werden demnach kohärent satzinitial positioniert, wo in unmarkierten Sätzen das Subjekt steht. Ist der Rezipient einer Proposition also im Kontext gegeben und das Thema nicht, so müsste der Sprecher, eine unmarkierte Phrasenstellung vorausgesetzt, das Rezipientenpassiv bevorzugen.

In Kapitel 3 wurden zwei Arbeiten von Joan Bresnan vorgestellt, die eine ähnliche Problemstellung untersuchten. In Lummi spielt die Personenhierarchie eine wesentliche Rolle in der Wahl des Genus Verbi. Das Subjekt darf nie in 3. Person konstituiert werden, wenn das Objekt in 1. oder 2. Person steht. Sollte eine solche Konstellation auftauchen, wird vom Aktiv ins Passiv oder umgekehrt gewechselt. Dieser Hard-Constraint in Lummi taucht als Soft-Constraint im Englischen auf. Die weitere Untersuchung sollte zeigen, ob solche Soft-Constraints auch für die Verwendung des Rezipientenpassivs verantwortlich sind. Eine ähnliche Problematik stellt die Dativ-Alternation im Englischen dar. Konstruktionen mit Verben wie *give* können zwei Nominalphrasen enthalten (*double-object*-Konstruktion) oder eine Nominal- und eine Präpositionalphrase. Joan Bresnan zeigte, dass die Wahl der Konstruktion durch diverse Faktoren wie Salienz, Definitheit, Pronominalität, Numerus, Person, Semantik und relative Länge gut vorhersagbar ist. Die weitere Untersuchung sollte zeigen, ob diese und ähnliche Faktoren die Alternation der Genera Verbi vorhersagen kann.

Eine mögliche Anwendung einer erfolgreichen Vorhersage des Genus Verbi stellt die Generierung aus semantischen Repräsentationen dar, welche keine Informationen über Genus Verbi enthalten.

Im 4. Kapitel wurde der experimentelle Ablauf beschrieben: Aus CQP-Korpora wurden die häufigsten Vollverben extrahiert und gezählt, die im Rezipientenpassiv auftauchten. Sätze mit den anteilig stärksten Rezipientenpassiv-Verben wurden aus den Korpora extrahiert und durch XLE geparkt und ggf. desambiguiert. Die Parses wurden durch ein XLE-Transfer-Skript

modifiziert, so dass für die Untersuchung relevante Faktoren leicht zu analysieren waren. Zu diesen Faktoren zählen u.a. das Vorhandensein eines Argumentes, die Person, Definitheit oder Pronominalität eines vorhandenen Argumentes, die lineare Abfolge der vorhandenen Argumente im Satz, sowie Informationen über Common Nouns, Eigennamen, Komplexitäten und Quantifikationen.

Die durch das XLE-Transfer-Skript gesammelten Informationen wurden zu Feature-Sets zusammengefasst, welche die Eingabe eines nachfolgenden Maximum-Entropie-Klassifikators und einer manuellen Analyse darstellten.

Der Klassifikator wurde mit 90% der Feature-Sets trainiert und auf 10% getestet. Evaluiert wurde der Klassifikator durch die *Accuracy*, die Maße *Precision*, *Recall* und *F-Score* und eine *Confusion-Matrix*.

Betrachtet man die Klassifikation aller drei Diathesen, so betrug die Accuracy hierbei 98,05%. Zwar lag die Baseline bei 94,6% an Aktiv-Vorhersagen, jedoch zeigten sich deutlich positive Werte bei den Precision- und Recall-Werten des Rezipientenpassivs von 71,8% und 57,8%. Das Rezipientenpassiv wurde mit dem geringsten Recall-Wert aber auch am häufigsten mit dem Vorgangspassiv verwechselt.

Eine binäre Passiv-Klassifikation stellt die beiden Passiv-Formen direkt gegenüber. Hier zeigte sich eine Accuracy von 81,31% relativ zu einer Baseline von 61%. Die Precision- und Recall-Werte beider Diathesen lagen über 70% bzw. 80%.

Trotz dieser Verbesserung des Recall-Wertes ließ sich eine deutliche Tendenz zur Verwechslung des RPs durch das VP erkennen. Dies mag daran liegen, dass der VP-Anteil der Feature-Sets bei 61% lag und somit mehr Merkmalsmuster des VPs trainiert wurden. Diese Theorie wurde dadurch untermauert, dass durch Wegnahme von Subkategorisierungs-Restriktionen der VP-Anteil weiter steigt und analog der Recall-Wert des Rezipientenpassivs sinkt. Der Precision-Wert des RPs blieb jedoch mit rund 75% relativ stabil.

Dies zeigte, dass der Klassifikator so gut trainiert wurde, dass er mit einer Treffsicherheit von 75% die Rezipientenpassive erkennt, es also eindeutige Muster gibt, die dem RP zuzuordnen sind. Nur gibt es noch mehr Muster und Varianten, in denen das Rezipientenpassiv auftauchen kann, diese wurden jedoch vom Klassifikator nicht erkannt.

Die manuelle Analyse ergab, dass im RP häufiger ein Agens realisiert wird als im VP. Die Definitheit überraschte insofern, als dass der Rezipient am häufigsten im RP indefinit ist und am seltensten definit. Dies deutete auf einen kleinen Widerspruch in der Thema-Rhema-Struktur, bei welcher bekannte Referenten durch definite Phrasen modelliert werden. Dies scheint jedoch auf die mangelhafte Desambiguierung von ambigen Sätzen zurückzuführen sein, wie in (4.2.1) beschrieben.

Das Pronomen-Merkmal zeigte hingegen einen deutlichen Beleg für die Informationsstruktur. Alle Subjekte der drei Diathesen wiesen den höchsten Pronomen-Anteil auf.

Es wurde zudem deutlich, dass die Vollverben mit dem häufigsten RP-Anteil solche waren, die eine besondere Form des *gebens* beschrieben, aber nicht *geben* selbst. (vgl. (60)). Ein weiteres mögliches Kriterium für die Abgrenzung der Diathesen könnte daher die Einteilung in semantisch Klassen via GermaNet sein.

Für eine Implementierung zum Zwecke eines Generierungsranking sollten noch solche und weitere Kriterien herangezogen werden. Desweiteren muss man auch die Möglichkeit berücksichtigen, dass es Fälle gibt, für die kein eindeutiges Ergebnis gefunden werden kann.

Literaturverzeichnis

- [Bird et al. 2009] BIRD, Steven ; KLEIN, Ewan ; LOPER, Edward: *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*. Beijing : O'Reilly, 2009 <http://www.nltk.org/book>
- [Bresnan et al. 2005] BRESNAN, Joan ; CUENI, Anna ; NIKITINA, Tatiana ; BAAYEN, R. H.: Predicting the dative alternation. In: *Cognitive foundations of interpretation*, 2005, S. 69–94
- [Bresnan et al. 2001] BRESNAN, Joan ; DINGARE, Shipra ; MANNING, Christopher D.: Soft constraints mirror hard constraints: Voice and person in English and Lummi. In: *Proceedings of the LFG01 Conference*. Stanford, CA : CSLI Publications, 2001, S. 13–32
- [Bussmann 2002] BUSSMANN, Hadumod: *Lexikon der Sprachwissenschaft*. Stuttgart : Kröner, 2002
- [Butt et al. 2002] BUTT, Miriam ; DYVIK, Helge ; KING, Tracy H. ; MASUICHI, Hiroshi ; ROHRER, Christian: The Parallel Grammar project. In: *COLING-02 on Grammar engineering and evaluation*. Morristown, NJ, USA : Association for Computational Linguistics, 2002, S. 1–7
- [Cahill u. Riester 2009] CAHILL, Aoife ; RIESTER, Arndt: Incorporating information status into generation ranking. In: *ACL-IJCNLP '09: Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2*. Morristown, NJ, USA : Association for Computational Linguistics, 2009
- [Carstensen et al. 2009] CARSTENSEN, Kai-Uwe ; EBERT, Christian ; KLABUNDE, Ralf ; JEKAT, Susanne ; LANGER, Hagen ; EBERT, Cornelia: *Computerlinguistik und Sprachtechnologie: Eine Einführung*. 3. Spektrum Akademischer Verlag, 2009
- [Collins 1995] COLLINS, P.: The indirect object construction in English: an informational approach. In: *Linguistics* 33 (1995), S. 35–49
- [Diedrichsen 2004] DIEDRICHSEN, Elke: The german 'bekommen-passive' and RRG. In: *Proceedings of the International Role and Reference Grammar Conference (RRG)*. Dublin, Ireland, October 2004
- [Diewald 1997] DIEWALD, Gabriele M.: *Grammatikalisierung. Eine Einführung in Sein und Werden grammatischer Formen*. Niemeyer, Tübingen, 1997
- [DUDEN 2009] DUDEN: *Duden 04. Die Grammatik: Unentbehrlich für richtiges Deutsch: Band 4*. 8. Bibliographisches Institut, Mannheim, 2009

- [Eisenberg 1999] EISENBERG, Peter: *Grundriss der deutschen Grammatik*. 2. Stuttgart : Metzlersche J.B. Verlagsb, 1999
- [Evert 2005] EVERT, Stefan: *The CQP Query Language Tutorial*. IMS Stuttgart, Juli 2005
- [Falk 2001] FALK, Yehuda N.: *Lexical-Functional Grammar - An Introduction to Parallel Constraint-Based Syntax*. Ctr For Study Of Lang & Info, 2001
- [Forst 2007] FORST, Martin: Filling Statistics with Linguistics – Property Design for the Disambiguation of German LFG Parses. In: *ACL 2007 Workshop on Deep Linguistic Processing*. Prague, Czech Republic : Association for Computational Linguistics, June 2007, S. 17–24
- [Helbig u. Buscha 2005] HELBIG, G. ; BUSCHA, J.: *Deutsche Grammatik: Ein Handbuch für den Ausländerunterricht*. Langenscheidt Verlag, 2005
- [IMS-CORPORA 2010] IMS-CORPORA: *Alphabetische Auflistung aller IMS-Korpora*. <http://www.ims.uni-stuttgart.de/projekte/corplex/local/imscorpora.html>. Version: 4. August 2010, Abruf: 4. August 2010
- [Jurafsky u. Martin 2008] JURAFSKY, Daniel ; MARTIN, James H.: *Speech and Language Processing (2nd Edition) (Prentice Hall Series in Artificial Intelligence)*. 2. Prentice Hall, 2008
- [Pimanyonok 2004] PIMANYONOK, Maxim: *Das bekommen-Passiv im Deutschen*. GRIN Verlag, 2004
- [Pittner u. Berman 2004] PITTNER, Karin ; BERMAN, Judith: *Deutsche Syntax*. Narr Dr. Gunter, 2004
- [Rohrer u. Forst 2006] ROHRER, Christian ; FORST, Martin: Improving coverage and parsing quality of a large-scale LFG for German. In: *Proceedings of LREC-2006*. Genoa, Italy, 2006
- [XLE-User-Documentation 2010] XLE-USER-DOCUMENTATION: *XLE User Documentation*. <http://www2.parc.com/isl/groups/nlitt/xle/doc/xle.html>. Version: 31. Juli 2010, Abruf: 31. Juli 2010