Lexikalischer Spracherwerb im Vorschulalter
Eine Annäherung an die Komposition des Lexikons von deutschsprachigen Kindern im Alter von drei bis fünf Jahren
Erstellung eines Korpuslexikons auf der Basis spontaner Sprachdaten

vorgelegt von
Julia Schulze M.A. phil.
geb. in Neuruppin

von der Fakultät I - Geistes- und Bildungswissenschaften
an der Technischen Universität Berlin
zur Erlangung des akademischen Grades

Doktorin der Philosophie
- Dr. phil. -

genehmigte Dissertation

Promotionsausschuss:
Vorsitzender: Prof. Dr. Stefan Weinzierl
Gutachter: Prof. Dr. em. Peter Erdmann
Gutachter: PD Dr. See Young-Cho


Berlin 2017
Inhaltsverzeichnis

1 Einleitung 9

2 Das Lexikon in der Linguistik 13
   2.1 Das Lexikon - Definition . . . . . . . . . . . . . . . . . . . . . . . 14
   2.2 Das Lexikon als Modell . . . . . . . . . . . . . . . . . . . . . . . 16
      2.2.1 Modulares Stufenmodell (Levelt) . . . . . . . . . . . . . 17
      2.2.2 Interaktives Aktivierungsmodell (Dell) . . . . . . . . . . . 19
      2.2.3 Welches ist das richtige Modell? . . . . . . . . . . . . . 20
   2.3 Das Lexikon bis zu einem Alter von drei Jahren . . . . . . . . . . 22
      2.3.1 Von Konstruktionsgrammatiken und anderen Theorien zum Erwerb des Lexikons . . . . . . . . . . . . . . . . . . . . . . . . . 23
      2.3.2 Der Erwerb des Lexikons - allgemeiner Verlauf . . . . . . 29
      2.3.3 Der frühe Wortschatzerwerb nach Kauschke . . . . . . . 33

3 Wortarten in der Linguistik 37
   3.1 Wort und Wortart - Definition . . . . . . . . . . . . . . . . . . . . 37
   3.2 Die Kategorisierung von Wortarten . . . . . . . . . . . . . . . . . 42
      3.2.1 Kategorisierung nach Kauschke . . . . . . . . . . . . . . 42
      3.2.2 Stuttgart-Tübingen-Tagsets (STTS) . . . . . . . . . . . . . 44
      3.2.3 Die Wahl einer geeigneten Kategorisierung - Auswahl und Begründung . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 47

4 Beobachtungsstudie - Ziele und Methodik 49
   4.1 Methodik . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 50
      4.1.1 Sprachaufnahmen . . . . . . . . . . . . . . . . . . . . . . . . 59
      4.1.2 Die Probandinnen und Probanden . . . . . . . . . . . . . . 63
      4.1.3 Diktiergerät . . . . . . . . . . . . . . . . . . . . . . . . . . 66
      4.1.4 Transkription der Daten mit FOLKER . . . . . . . . . . . . 66
      4.1.5 Bearbeitung der Audio-Dateien mit AUDACITY . . . . . . . 70
      4.1.6 Bearbeitung der Transkripte mit dem EXMARaLDA Partitur-Editor . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 71
INHALTSVERZEICHNIS

4.1.7 Kriterien für das Taggen der Wortarten nach den Richtlinien der STTS .............................................. 74
4.1.8 Kriterien für die Lemmatisierung der syntaktischen Wörter 84

5 Die Arbeit mit dem CorpusAnalyser 89

5.1 Die Inhalte im Überblick ............................................. 90
5.2 Die Funktionen des CorpusAnalysers ................................. 97
  5.2.1 01_wave_convert ................................................. 97
  5.2.2 02_create_database ............................................. 98
  5.2.3 03_import_data ................................................. 100
  5.2.4 04_query_db .................................................. 101
    5.2.4.1 01_tag_count ............................................. 101
    5.2.4.2 02_typetoken ............................................. 106
    5.2.4.3 03_typetoken_lemma .................................... 110
    5.2.4.4 04 lemma_tag ............................................ 117
    5.2.4.5 05_lemma_bedeutung .................................... 122
    5.2.4.6 06_typetoken_bedeutung ................................ 123
  5.3 Möglichkeiten für die Arbeit mit dem CorpusAnalyser ............. 126

6 Analyse der Wortarten 129

6.1 Gruppierung der POS-Tags .......................................... 130
6.2 Altersgruppen im Vergleich ......................................... 133
  6.2.1 Verteilung der Tokens ......................................... 133
  6.2.2 Verteilung der Types .......................................... 138
  6.2.3 Das Verhältnis der Types zu den Tokens ......................... 140
  6.2.4 Zusammenfassung .............................................. 143
6.3 Individuelle Unterschiede in den Altersgruppen ..................... 144
  6.3.1 Gruppe 1281 bis 1495 .......................................... 146
  6.3.2 Gruppe 1496 bis 1708 .......................................... 148
  6.3.3 Gruppe 1709 bis 1983 .......................................... 150
  6.3.4 Vergleich der individuellen Verläufe mit den Altersgruppen 151
  6.3.5 Zusammenfassung .............................................. 164
6.4 Analyse der ersten einhundert Wörter je Kind ....................... 166
6.5 Die Wortartenverteilung bei Erwachsenen .......................... 172

7 Der Grundwortschatz semantisch betrachtet 177

7.1 Die Bedeutungskategorien .......................................... 179
  7.1.1 Adjektive ................................................... 180
  7.1.2 Adverbien ................................................... 181
  7.1.3 Präpositionen ............................................... 183
INHALTSVERZEICHNIS

7.1.4 Artikel ............................................................... 184
7.1.5 Fremdsprachliches Material ................................. 184
7.1.6 Interjektionen ..................................................... 184
7.1.7 Konjunktionen und Subjunktionen ......................... 185
7.1.8 Substantive ......................................................... 187
7.1.9 Pronomen ............................................................ 191
7.1.10 Partikeln ............................................................ 192
7.1.11 Verben ............................................................... 195
7.2 Tendenzen in der Verwendung von Inhaltswörtern .......... 198
  7.2.1 Nomen (n-abstr-eig, n-abstr-hdlg, n-abstr-maß, n-abstr-vorg, n-abstr-vorst, n-abstr-zust, n-belebt, n-unbelebt, ne) ........................................ 198
  7.2.2 Verben (v-aux, v-kop, v-mod, v-hdlg, v-vorgang, v-zustand) ........................................ 203
  7.2.3 Adjektive (ad-part, ad-qual, ad-quant, ad-rel) ......... 206
  7.2.4 Adverbien (adv-kaus, adv-komm, adv-lok, adv-mod, adv-temp) ........................................... 211
7.3 Tendenzen in der Verwendung von Funktionswörtern ..... 212
  7.3.1 Interjektionen (itj) ................................................ 212
  7.3.2 Konjunktionen (kon-add, kon-adv, kon-alt, kon-kaus, kon-spez, kon-temp, kon-vgl) und Subjunktionen (sub-fin, sub-kaus, sub-kond, sub-konz, sub-mod-instr, sub-neutr, sub-temp) ........................................ 213
  7.3.3 Pronomen (pav, pdat, pds, piat, pidat, pis, pper, ppos, prels, prf, pwat, pwav, pws) ........................................ 214
  7.3.4 Präpositionen (pr-kaus, pr-lok, pr-mod, pr-neutr, pr-temp) ........................................ 215
  7.3.5 Partikeln (ptk-abt, ptk-ant, ptk-fok, ptk-gespr, ptk-grad, ptk-neg, ptkvz, ptkzu) ........................................ 215
7.4 Hinweise zur Verwendung des Lexikons ...................... 216
  7.4.1 Beispielhafte Ausgabedateien ................................ 217
7.5 Zusammenfassung .................................................... 219

8 Diskussion .............................................................. 223

9 Ausblick ................................................................. 245

Abbildungsverzeichnis ............................................. 249

Tabellenverzeichnis .................................................. 251

Literaturverzeichnis .................................................. 253
Vorwort

Für Richard - meine Inspiration

Danke Alex

Die vorliegende Arbeit enstand durch die Mitwirkung und Unterstützung zahlreicher Personen, denen ich an dieser Stelle meinen Dank aussprechen möchte.

Mein besonderer Dank gilt Herrn Prof. Dr. Peter Erdmann und Herrn PD Dr. habil. See-Young Cho, die mir zu jeder Zeit mit kompetentem fachlichen Rat zur Seite standen. Ich verdanke beiden hilfreiche Unterstützung und viele anregende Gespräche. Sie begleiteten mich während der gesamten Entstehungszeit dieser Arbeit mit vielen kreativen Ideen und wertvoller professioneller Hilfe.


Nicht zu vergessen sind meine lieben Kommilitoninnen und Unimädels, die die vorliegende Arbeit Korrektur lasen und mich in meinem Vorhaben immer bestärkten.
Kapitel 1

Einleitung


„Although aspects of the sound, communicative, morphological, and syntactic systems continue to develop after the age of 5 years, the acquisition of words exhibits the most significant improvements after the preschool years. If the language development of an individual would was mysteriously halted at the age of 5 years, this individual would possess most of the sound, morphological, syntactic, and communicative systems of her native language. However, her vocabulary, at best, would be 25% of that of a normal adult.“ ([Kuczaj (1999)], S. 134)


In Kapitel 3 wird definiert, was genau als Wort und als Wortart verstanden wird und wie die Begriffe hier verwendet werden. In Anlehnung an die spätere Durchführung und Analyse der Sprachdaten muss zudem eine geeignete Klassifikation der Wortarten vorgenommen werden. Hierzu werden verschiedene Klassifikationssysteme gesichtet und zwei von ihnen vorgestellt. Insbesondere im Hinblick darauf, dass sich die Sprache des Kindes von jener des Erwachsenen unterscheidet, muss das Klassifikationssystem mehreren Anforderungen genügen. Es sollte alle Wortarten enthalten, die im Lexikon
von Kindern im Vorschulalter vertreten sind. Im Idealfall enthält es auch alle Wortarten, die im erwachsenen Lexikon vertreten sind, um so einen späteren Vergleich zu ermöglichen. Ferner ist es wichtig, ein System auszuwählen, das allgemein zugänglich ist und den Anspruch einer gewissen Popularität erfüllt.

Im weiteren Verlauf werden in Kapitel 4 die Grundlagen für das Durchführen der Beobachtungsstudie zur Erlangung von spontanen Sprachdaten gelegt. Die Probandinnen und Probanden werden in Bezug auf das Geburtsdatum, das Geschlecht sowie auf weitere soziolinguistische Merkmale vorgestellt. Anschließend erfolgt eine Vorstellung und Erläuterung aller Softwareprogramme, die in der vorliegenden Arbeit Anwendung finden. Im Verlauf der Analysetätigkeit häufte sich eine enorme Datenmenge an, weshalb zur Verarbeitung selbiger ein elektronisches Korpus erstellt wurde. Durch dessen funktionele Basis als Datenbank können vielfältige Aspekte der Sprache untersucht werden, was später gezeigt wird.

Nachfolgend soll in Kapitel 5 eine spezielle Software erläutert werden, die zunächst eigens für die Zwecke der vorliegenden Arbeit konzipiert wurde und ohne welche die Ergebnisse dieser Arbeit nicht in der Form vorliegen würden, wie sie es jetzt tun. Da sich durch die ständige und interaktive Weiterentwicklung des Programms zahlreiche unerwartete Möglichkeiten für die Arbeit mit der Software auftaten, wird dieser ein eigenes Kapitel gewidmet. Diesbezüglich werde ich inspirierend weitere Möglichkeiten für linguistische Arbeiten aufzeigen, die dieses Programm bietet.

In Kapitel 6 wird in einer Analyse der aus der Beobachtung erhaltenen Sprachdaten die Wortartenverteilung unter unterschiedlichen Aspekten untersucht, unter anderem im Vergleich zwischen zuvor festgelegten Altersgruppen sowie unter Beachtung individueller Unterschiede zwischen den Kindern in unterschiedlichen Altersgruppen.

hier insbesondere der Inhaltswörter vorgenommen, um die zuvor ermittelten Daten besser in die bisherige Forschungslandschaft einordnen zu können.

In der anschließenden Diskussion (Kapitel 8) sollen die gewonnenen Ergebnisse in Bezug zu den bisherigen Erkenntnissen bezüglich der Komposition des Lexikons gesetzt werden. Ferner wird versucht, aus den ermittelten Daten ein konsistentes Bild der Wortartenverteilung von Kindern im Alter zwischen 3;5 und 5;5 Jahren zu schaffen, das den Ansprüchen neuer wissenschaftlicher Erkenntnisse gerecht wird und als Grundlage für weitere linguistische Arbeiten dienen kann. Diese möglichen weiterführenden Arbeiten sollen in Kapitel 9 aufgezeigt werden. Selbstverständlich sind weitere als dort genannte anschließende Arbeiten möglich und erwünscht.
Kapitel 2

Das Lexikon in der Linguistik

KAPITEL 2. DAS LEXIKON IN DER LINGUISTIK

deshalb nicht gänzlich unbeachtet bleiben. Zudem können sie insbesondere für weiterführende Arbeiten von Interesse sein.

2.1 Das Lexikon - Definition

"Sprache funktioniert so, indem das Gehirn jedes Menschen ein Lexikon mit Wörtern und den Konzepten, für die sie stehen (also ein mentales Lexikon), enthält sowie eine Menge an Regeln, nach denen die Wörter kombiniert werden, um Beziehungen zwischen den Konzepten zu bezeichnen (also eine mentale Grammatik)." ([Pinker (1996)], S. 99)


- Skier
  - a) one who skis
2.1. DAS LEXIKON - DEFINITION

- b) Noun countable
- c) stem + -er
- d) /skir/


Einen weiteren Grund dafür, dass das mentale Lexikon keinesfalls im Stile eines Wörterbuches organisiert sein kann, sieht Aitchison (2003) in der Tatsache, dass Menschen ständig neue Wörter zum mentalen Lexikon


Aitchison hat mit ihrer Argumentation die Unterschiede zwischen dem mental lexicon und dem dictionary deutlich gemacht. Da in dieser Arbeit das mentale Vokabular und dessen Komposition im Lexikon von Kindern im Vorschulalter thematisiert wird, soll auch im Folgenden der Begriff (mentales) Lexikon im Sinne von Aitchisons Ausführungen verwendet werden.

### 2.2 Das Lexikon als Modell

2.2. DAS LEXIKON ALS MODELL

des Lexikons in der Sprache geht und sollen in dieser Arbeit in Bezug zur Komposition des Lexikons im fortgeschrittenen Spracherwerb gesetzt werden.

2.2.1 Modulares Stufenmodell (Levelt)


So ausgefeilt dieses Modell des Lexikons auch erscheint, ergeben sich doch einige Fragen. Angesichts der Tatsache, dass es sich um eine serielle Verarbeitung aller hier aufgeführten Schritte handelt, bei der es quasi kein *Zurück* mehr gibt, erscheint gerade das Phänomen der Versprecher meines Erachtens nicht damit erklärt werden zu können. Es kann zwar erklären, weshalb bei Versprechern oft form- oder bedeutungssähnliche Lexeme erreicht werden, nämlich aufgrund ihrer *ortsnahen* Speicherung im Lexikon, wodurch es zu einer Aktivierung zweier oder mehrerer Lexeme kommen kann.
erklärt jedoch nicht, weshalb Versprecher, oft noch bevor sie vollständig
gäußert werden, anscheinend mental korrigiert und letztendlich als korrektes
Lexem hervorgebracht werden können. Dies spräche meines Erachtens für
eine interaktive Verarbeitung (Abschnitt 2.2.2), bei welcher auch Rückschritte
zu vorhergehenden Stufen erlaubt sind.

2.2.2 Interaktives Aktivierungsmodell (Dell)

Nach Dell (1992) besteht kein Zweifel daran, dass lexikalischer Zugang
die Abbildung zwischen einer konzeptuellen Repräsentation und der
phonologischen Form eines Wortes beinhaltet ([Dell und O’Séaghdha (1992)]).
Dells Hypothese ist, dass der lexikalische Zugang, der bei der
Sprachproduktion eine Rolle spielt, in zwei Schritte aufgeteilt werden
can: den Lemmatazugang (1), bei dem eine Abbildung erreicht wird
zwischen einem zu lexikalisierenden Konzept und einem Lemma sowie dem
phonologischen Zugang (2), bei dem ein Lemma in seine phonologische
Form übersetzt wird. Eine Frage, die Dell diesbezüglich beschäftigt,
Ansicht überlappen sich diese nicht, sondern erfolgen nacheinander
([Levelt (1992)]). Dell erklärt, dass es bereits während des Lemmazugangs
t zu einer Aktivierung der phonologischen Information kommt sowie zu einer
Aktivierung der semantischen Information während des phonologischen
Zugangs. Formal sieht Dells Interaktives Modell zahlreiche Einheiten vor,
die in einem Netzwerk organisiert sind. In diesem Netzwerk erlauben
die Verbindungen eine bidirektionale Aktivierung zwischen den Einheiten
an den angrenzenden Stufen. Formal erfolgen in diesem interaktiven
Lexikonmodell sechs Schritte beim Zugang zum Lexikon in der Produktion
(nach ([Dell und O’Séaghdha (1992)], S. 295):

1. Die semantischen Einheiten des zu lexikalisierenden Konzeptes erhalten
   externen Input.

2. Die Aktivierung breitet sich im Netzwerk aus, bestimmt durch die
   activation-update-function.

3. Die am meisten aktivierte Worteinheit wird ausgewählt.

4. Wenn ein Wort bereit ist für die phonologische Enkodierung, erhält es
   einen auslösenden Anstoß zur Aktivierung.

5. Die Aktivierung fährt wie zuvor fort mit der Ausbreitung, aber die
   passende phonologische Einheit wird zusätzlich signifikant aktiviert.

Im Gegensatz zu Levelts Modell des Lexikons sind im Interaktiven Aktivierungsmodell interaktive Prozesse zwischen den beteiligten Komponenten möglich. Dadurch ist augenscheinlich gegeben, dass beispielsweise bei einer irrtümlichen Wahl eines Lemmas zum entsprechenden Konzept auch ein Weg zurück möglich ist, um Korrekturen vornehmen zu können.

2.2.3 Welches ist das richtige Modell?


### 2.3 Das Lexikon bis zu einem Alter von drei Jahren

2.3. DAS LEXIKON BIS ZU EINEM ALTER VON DREI JAHREN

2.3.1 Von Konstruktionsgrammatiken und anderen Theorien zum Erwerb des Lexikons


- general nominals: *apple, shoe*
- specific nominals: *Sarah, Mommy*
- action words: *throw, dance*
- personal social words: *bye-bye, thank you*
- modifiers: *cold, wet*
- functors: *of, and*


\(^1\)Ich möchte den Begriff *item* in diesem Zusammenhang nicht ins Deutsche übersetzen, da er folglich nur schwer wiedergeben würde, was gemeint ist. *Item* meint eigentlich *Einheit* und diese Übersetzung passt auch hier sehr gut.
KAPITEL 2. DAS LEXIKON IN DER LINGUISTIK

an, dass unser Wissen über Sprache ein wesentlicher Bestandteil unserer genetischen Ausstattung ist ([Chomsky (1967)])

In diesem fest integrierten Bestandteil existieren sogenannte Universalien, die wiederum angeboren sind und von denen zwei Arten beschrieben werden: 1) sogenannte Zutaten wie Parts-of-speech (Nomen, Verben, Adjektive) oder grammatische Relationen wie Subjekt oder Objekt sowie verschiedene Arten von Lauten (Vokale, Konsonanten) und 2) Regeln, mit denen die Zutaten angewendet werden, um eine Sprache zu konstruieren ([Evans (2014)], S. 68).


1. **Joint attention**: entsteht in Situationen, in denen das Kind und seine Bezugsperson ihre Aufmerksamkeit gleichzeitig auf ein drittes Objekt richten und sich dieser gemeinsamen Aufmerksamkeit bewusst sind.

2. **Intention reading**: die Fähigkeit von Babies bzw. Kindern, die Handlungen anderer Akteure als sinnvoll und zielgerichtet zu erachten.

3. **Schematisierung**: erlaubt Kindern Ähnlichkeiten zwischen Phrasen wie *more juice*, *more apple*, *more milk* zu erkennen und daraus ein Muster zu abstrahieren, wie *more X*, das eine freie Stelle für andere linguistische Elemente bereit hält.

4. **Rollentausch und Imitation**: beide Faktoren spielen eine entscheidende Rolle für das triadische Muster der Joint attention. Die Fähigkeit, Laute, die von anderen Personen geäußert werden, zu imitieren, ist essentiell, um Sprache zu erlernen. Umgekehrt muss das Kind lernen, dass es selbst in die Rolle der Laute produzierenden Person schlüpfen kann und so zum Sender wird.

5. **Musterwiedererkennung**: betrifft die Fähigkeit statistische Regularitäten (Muster) in der Sprache zu erkennen.

Einige dieser Fähigkeiten (zum Beispiel Schematisierung und Musterwiedererkennung) sind keine typisch menschlichen, während die Fähigkeiten Joint...

1. **Voraussetzende Prozesse**: segmenting speech, conceptualizing referents
2. **Grundlegende Prozesse**: joint-attention, intention-reading, cultural learning
3. **Vereinfachende Prozesse**: lexical contrast, linguistic context ([Tomasello (2000a)], S. 58)

Demnach stellen sich Kinder während ihres ersten Lebensjahres auf die Sprachunterschiede und Muster in ihrer Muttersprache ein. Das Erlernen eines neuen Verbs sei dabei besonders schwierig. Die Gründe dafür sind, dass die unterschiedlichen Zustände eines Verbs kurzlebig sind, wodurch die referentielle Situation oft nicht greifbar ist, wenn ein Verb geäußert wird. Tomasello et al. (1992) fanden zum Beispiel heraus, dass Kinder Verben am besten in direkt bevorstehenden Situationen erlernen, am zweitbesten durch gerade abgeschlossene Aktionen und am schlechtesten in Situationen, die gerade im Gange sind. Für den Erwerb neuer Wörter spielt außerdem der lexikalische Kontrast eine entscheidende Rolle. Dieser hilft Kindern, die jeweiligen Referenten in den unmittelbaren Situationen zu identifizieren. Mit 3 oder 4 Jahren besitzen Kinder ausreichend syntaktisches Wissen, um damit neue Wörter zu erwerben. Das bedeutet also, dass

„In social-pragmatic view, then, children acquire linguistic symbols as a kind of by-product of social interactions with adults, in much the same way they learn many other cultural conventions. ([Tomasello (2000a)], S. 90)“

Teile derselben übergreifenden Prozesse sind ([Tomasello (2000a)], S. 93).


Ob und welche der hier erwähnten Ansätze den Lexikonerwerb korrekt wiedergeben, kann an dieser Stelle nicht beurteilt werden. Es macht jedoch den Anschein, dass das Erlernen von Wörtern, Wortkategorien und damit verbunden der Aufbau des mentalen Lexikons nicht losgelöst vom allgemeinen Prozess des Spracherwerbs betrachtet werden kann.

2.3.2 Der Erwerb des Lexikons - allgemeiner Verlauf


„Das Kind wählt naturgemäß pädozentrisch; seine Umgebung, seine Interessen: Eltern, Geschwister, Wärterin, Spielsachen und Tiere, Eßbares und Trinkbares, Tönendes und Bewegtes, bilden das ausschließliche Material für seinen ersten Wortschatz.“ ([Stern und Stern (1965)], S. 195)


„Viel mehr geht mit dem starken Aufschwung der Gegenstandsbezeichnungen schon ein leises Einsetzen der Tätigkeitswörter Hand in Hand; das „Aktionsstadium“ bereitet sich vor.“ ([Stern und Stern (1965)], S. 196)

In der darauffolgenden Phase kommt es zur lexikalischen Strukturierung der Wörter in Wortfelder; die zwischen 3;0 und 12;0 Jahren angesiedelt ist und unter anderem durch das Auftreten polarer Adjektive und relational deﬁniert er Verwandtschaftswörter gekennzeichnet ist ([Wode (1988)], S. 144). Dittmann (2006) konstatiert eine hohe Differenzierung des Wortschatzes mit etwa 3;0 Jahren, welche als Ausdruck von innerpsychischen Prozessen verstanden werden kann ([Dittmann (2006)], S. 49). Demzufolge verwenden Kinder in dieser Phase zunehmend Nomen für Gefühle und mentale Prozesse (Idee), Verben, die sich auf emotionale Vorgänge beziehen (mögen, brauchen, dürfen, weinen, helfen) und Verben, die mentale Prozesse beschreiben


Eine Vielzahl der Untersuchungen umfasst die produktive Lexikonentwicklung, also das expressive Vorkommen von Nomen und Verben im Wortschatz ([Kauschke (2007)], S. 59). Die vorwiegende Untersuchung von Nomen und Verben hängt vermutlich damit zusammen, dass diese Wortkategorien mit einem großen Anteil am Wortschatz vorkommen und
2.3. DAS LEXIKON BIS ZU EINEM ALTER VON DREI JAHREN


2.3.3 Der frühe Wortschatzerwerb nach Kauschke


diente als Vergleichspunkt. Die Aufnahmesituation der Daten bestand in der freien Interaktion zwischen Mutter und Kind in einem Untersuchungsraum. Es wurden Videonaufnahmen erzeugt, mittels welcher im Anschluss Transkripte über zehn Minuten erstellt wurden. Festgehalten wurden alle verbalen, vokalen und paraverbalen Äußerungen von Kind und Mutter. Für die Analyse wurde jedes einzelne Wort extrahiert und in seiner zielsprachlichen Form notiert. Anschließend wurde jedes auf diese Weise erhaltene Wort in eine Datenbank überführt und die Type-Token-Relation ermittelt. Danach wurden die Wörter folgenden Wortartenkategorien zugeordnet (vgl. 3.2.1):

- Nomen
- Verben
- Adjektive
- personal-social words
- relationale Wörter
- Pronomen
- Funktionswörter
- Onomatopöien
- Sonstige

Die Datenbank umfasste laut Kauschke (1999) insgesamt 751 verschiedene Wörter, die von allen Kindern 3440 mal als Types und 9115 mal als Tokens geäußert wurden. Den Ergebnissen nach zu urteilen nimmt die Anzahl der Wörter im Laufe der Zeit zu. Kauschke zufolge ist ein Type-Zuwachs zu verzeichnen, was bedeutet, dass die Anzahl unterschiedlicher Wörter in Abhängigkeit vom Alter ansteigt. Demnach konnte ein exponentielles Wachstum im zweiten Lebensjahr festgestellt werden (Kauschke (1999) S. 141). Erst im dritten Lebensjahr würde es zu einer Abnahme und einem anschließenden linearen Verlauf der Types kommen. Hinsichtlich der Verwendungshäufigkeit der Wörter (Tokens) ist das Wachstumsmuster mit dem der Types vergleichbar, so Kauschke. Mit zunehmendem Alter sind keine bedeut samen Veränderungen mehr feststellbar, was darauf schließen lässt, dass das Verhältnis der Types zu den Tokens gleich bleibt. Hinsichtlich der Komposition des Lexikons fand Kauschke heraus, dass relationale Wörter sowie personal-social-words mit über zwei Dritteln anfangs vorherrschend
2.3. DAS LEXIKON BIS ZU EINEM ALTER VON DREI JAHREN

Kapitel 3

Wortarten in der Linguistik


3.1 Wort und Wortart - Definition

„Wortarten sind Mengen bestimmter Art, und keine Eigenschaften (Merkmale o. ä.): Wortartbegriffe wie „Verb“ sollen zur Bezeichnung von Mengen dienen, und nicht zur Bezeichnung von Eigenschaften, durch die diese Mengen festgelegt werden.“ ([Budde (2000)], S. 4)


1) Was wird klassifiziert? (Lexeme, Wortformen, syntaktische Wörter)

2) Nach welchen Kriterien wird klassifiziert? (nach grammatisch-semantischen, nach morphologischen, nach syntaktischen, mit einer Kombination aus den Kriterien)

3) Zu welchem Zwecke wird klassifiziert? (zur Beschreibung der Muttersprache, für die maschinelle Sprachbearbeitung, für die linguistische Theorie, für die Modellierung des kindlichen Spracherwerbs)

([Knobloch und Schaeder (2009)] S. 22 ff.).


„Ein syntaktisches Wort ist eine abgeschlossene morphologische Einheit mit bestimmten *formalen* Merkmalen (=Signifiant) sowie
3.1. WORT UND WORTART - DEFINITION

bestimmten *grammatischen* und/oder *inhaltlichen* Merkmalen (=Signifik), die eine Position in einer syntaktischen Struktur einnehmen kann." ([Gallmann (1991)], S. 2)

Dieser Definition zufolge verfügen alle syntaktischen Wörter über Wortartmerkmale und es gibt kein syntaktisches Wort, das nicht hinsichtlich der Wortart spezifiziert werden kann (siehe auch Abschnitt 3.2.2). Geht es um die Abbildung lexikalischen Wissens in Form von Wörterbüchern oder Lexika, dann werden ihre Grundeinheiten als Lemmata abgebildet ([Gallmann (1991)] und siehe Abschnitt 4.1.8).

3.2 Die Kategorisierung von Wortarten

In diesem Abschnitt sollen zwei Möglichkeiten der Wortartenkategorisierung vorgestellt werden. Dabei wird zunächst nicht unterschieden nach syntaktisch, semantisch, morphologisch, phonologisch oder lexikalisch motivierten Klassifikationen. Vielmehr wird im Anschluss an die vorgestellten Klassifikationssysteme eine Bewertung in Bezug auf das hiesige Vorhaben vorgenommen und bei Bedarf die Vor- und Nachteile erläutert.

3.2.1 Kategorisierung nach Kauschke

3.2. DIE KATEGORISIERUNG VON WORTARTEN

<table>
<thead>
<tr>
<th>Wortart</th>
<th>Untergruppen und Beispiele</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nomen</td>
<td>specific: Eigennamen (Holger), specific: Personen und spezifische Objekte (Mama), general: belebte und unbelebte Objekte (Ball), abstrakte Nomen (Idee), internal state-Nomen (Angst)</td>
</tr>
<tr>
<td>Verben (action words)</td>
<td>objektbezogene Handlungen und Tätigkeiten (suchen), nicht objektbez. Handlungen: Events, Bewegung, Zustand (pullern), innere Handlungen oder Tätigkeiten z. B.: mentale, emotionale, volitionale „internal state“-Verben (weinen, glauben)</td>
</tr>
<tr>
<td>Adjektive (modifiers)</td>
<td>modifizierte Elemente wie Attribute, Eigenschaften und Merkmale (heiß, vier), innerpsychische Zustände, internal state-Adjektive (böse, toll)</td>
</tr>
<tr>
<td>Personal-social words</td>
<td>Interaktive und expressive Wörter (ja, nein), social-expressives: Floskeln, Grüße (hallo, danke), Gesprächssignale (hm), attention getting devices (guck), auf das persönliche Erleben bezogene expressive Äußerungen, internal state-Interjektionen (aua)</td>
</tr>
<tr>
<td>Relationale Wörter</td>
<td>Relationen zwischen Handlungen/Objekten: Auftauchen/Verschwinden (da, weg), Ort (oben) Funktionen von/mit Objekten (ran, auf), zeitliche Durchführung von Handlungen (wieder)</td>
</tr>
<tr>
<td>Pronomen</td>
<td>Personalpronomen (du), Demonstrativpronomen (dies), Possessivpronomen (sein) u. a.</td>
</tr>
<tr>
<td>Funktionswörter</td>
<td>Präpositionen (aus), Hilfsverben (haben), Artikel (ein), Konjunktionen (weil), Fragewörter (warum) u. a.</td>
</tr>
<tr>
<td>Onomatopöien</td>
<td>lautmalerische Äußerungen (brumm, tatütata)</td>
</tr>
<tr>
<td>Sonstige</td>
<td>Partikeln (eben, denn), nicht klassifizierbare Wörter</td>
</tr>
</tbody>
</table>

Tabelle 3.1: Wortartenklassifikation nach Kauschke ([Kauschke (1999)], S. 140)


3.2.2 Stuttgart-Tübingen-Tagsets (STTS)

Im Folgenden werden die Stuttgart-Tübingen Tagsets (STTS) vorgestellt, die am Institut für maschinelle Sprachverarbeitung der Universität Stuttgart und am Seminar für Sprachwissenschaft der Universität Tübingen entwickelt wurden. Die STTS enthalten insgesamt 54 Tags (Wortartenbezeichner) für deutsche Textkorpora. Die unten stehende Tabelle stellt alle 54 Tags vor:

<table>
<thead>
<tr>
<th>POS-Tag</th>
<th>DESCRIPTION</th>
<th>EXAMPLES</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADJA</td>
<td>attributives Adjektiv</td>
<td>[das] große [Haus]</td>
</tr>
<tr>
<td>ADJD</td>
<td>adverbiales oder prädikatives Adjektiv</td>
<td>[er fährt] schnell, [er ist] schnell</td>
</tr>
<tr>
<td>ADV</td>
<td>Adverb</td>
<td>schon, bald, doch</td>
</tr>
<tr>
<td>APPR</td>
<td>Präposition, Zirkumposition links</td>
<td>in [der Stadt], ohne [mich]</td>
</tr>
<tr>
<td>APPRART</td>
<td>Präposition mit Artikel</td>
<td>im [Haus], zur [Sache]</td>
</tr>
<tr>
<td>APPO</td>
<td>Postposition</td>
<td>[ihm] zufolge, [der Sache] wegen</td>
</tr>
</tbody>
</table>


2Für weitere Informationen verweise ich auf die folgende Webseite: http://www.ims.uni-stuttgart.de/forschung/ressourcen/lexika/TagSets/stts-table.html; siehe auch [Schiller, Teufel, Stöckert und Thielen (2009)], S. 6-7)
### 3.2. DIE KATEGORISIERUNG VON WORTARTEN

<table>
<thead>
<tr>
<th>Code</th>
<th>Definition</th>
<th>Beispiel</th>
</tr>
</thead>
<tbody>
<tr>
<td>APZR</td>
<td>Zirkumposition rechts</td>
<td>[von jetzt] an</td>
</tr>
<tr>
<td>ART</td>
<td>bestimmter oder unbestimmter Artikel</td>
<td>der, die, das, ein, eine</td>
</tr>
<tr>
<td>CARD</td>
<td>Kardinalzahl</td>
<td>zwei [Männer], [im Jahre] 1994</td>
</tr>
<tr>
<td>FM</td>
<td>fremdsprachliches Material</td>
<td>[Er hat das mil] A big fish [übersetzt]</td>
</tr>
<tr>
<td>ITJ</td>
<td>Interjektion</td>
<td>mhm, ach, tja</td>
</tr>
<tr>
<td>KOUS</td>
<td>unterordnende Konjunktion mit Satz</td>
<td>weil, dass, damit, wenn, ob</td>
</tr>
<tr>
<td>KON</td>
<td>nebenordnende Konjunktion</td>
<td>und, oder, aber</td>
</tr>
<tr>
<td>KOKOM</td>
<td>Vergleichskonjunktion</td>
<td>als, wie</td>
</tr>
<tr>
<td>NN</td>
<td>normales Nomen</td>
<td>Tisch, Herr, [das] Reisen</td>
</tr>
<tr>
<td>NE</td>
<td>Eigennamen</td>
<td>Hans, Hamburg, HSV</td>
</tr>
<tr>
<td>PDS</td>
<td>substituierendes Demonstrativpronomen</td>
<td>dieser, jener</td>
</tr>
<tr>
<td>PDAT</td>
<td>attribuierendes Demonstrativpronomen</td>
<td>jener [Mensch]</td>
</tr>
<tr>
<td>PIS</td>
<td>substituierendes Indefinitpronomen</td>
<td>keiner, viele, man, niemand</td>
</tr>
<tr>
<td>PIAT</td>
<td>attribuierendes Indefinitpronomen ohne Determiner</td>
<td>kein [Mensch], irgendein [Glas]</td>
</tr>
<tr>
<td>PIDAT</td>
<td>attribuierendes Indefinitpronomen mit Determiner</td>
<td>[ein] wenig [Wasser], [die] beiden [Brüder]</td>
</tr>
<tr>
<td>PPER</td>
<td>irreflexives Personalpronomen</td>
<td>ich, er, ihm, mich, dir</td>
</tr>
<tr>
<td>PPOSS</td>
<td>substituierendes Possessivpronomen</td>
<td>meins, deiner</td>
</tr>
<tr>
<td>PPOSAT</td>
<td>attribuierendes Possessivpronomen</td>
<td>mein [Buch], deine [Mutter]</td>
</tr>
<tr>
<td>PRELS</td>
<td>substituierendes Relativpronomen</td>
<td>[der Hund], der</td>
</tr>
<tr>
<td>PRELAT</td>
<td>attribuierendes Relativpronomen</td>
<td>[der Mann], dessen [Hund]</td>
</tr>
<tr>
<td>PRF</td>
<td>reflexives Personalpronomen</td>
<td>sich, einander, dich, mir</td>
</tr>
<tr>
<td>PWS</td>
<td>substituierendes Interrogativpronomen</td>
<td>wer, was</td>
</tr>
<tr>
<td>PWAT</td>
<td>attribuierendes Interrogativpronomen</td>
<td>welche [Farbe], wessen [Hut]</td>
</tr>
<tr>
<td>PWAV</td>
<td>adverbiales Interrogativ- oder Relativpronomen</td>
<td>warum, wo, wann, worüber, wobei</td>
</tr>
<tr>
<td>PAV</td>
<td>Pronominaladverb</td>
<td>dafür, dabei, deswegen, trotzdem</td>
</tr>
<tr>
<td>PTKZU</td>
<td>„zu“ vor Infinitiv</td>
<td>zu [gehen]</td>
</tr>
<tr>
<td>PTKNEG</td>
<td>Negationspartikel</td>
<td>nicht</td>
</tr>
<tr>
<td>PTKVZ</td>
<td>abgetrennter Verbzusatz</td>
<td>[er kommt] an, [er fährt] Rad</td>
</tr>
<tr>
<td>PTKANT</td>
<td>Antwortpartikel</td>
<td>ja, nein, bitte, danke</td>
</tr>
<tr>
<td>PTKA</td>
<td>Partikel bei Adjektiv oder Adverb</td>
<td>am [schönsten], zu [schnell]</td>
</tr>
<tr>
<td>TRUNC</td>
<td>Kompositionserstglied</td>
<td>An- [und Abreise]</td>
</tr>
<tr>
<td>VVFIN</td>
<td>finitives Verb, voll</td>
<td>[du] gehst, [wir] kommen [an]</td>
</tr>
</tbody>
</table>

_Inwieweit findet ein adäquater Gebrauch der reflexiven Personalpronomen statt? Stellen Auxiliarverben eine besondere Schwierigkeit im Sprachgebrauch dar und wenn ja, in welchem Alter?_

Diese und weitere Fragen sind sehr leicht beantwortbar, wenn die Daten eines Korpus’ mit Hilfe der STTS getaggt und im besten Falle in eine
3.2. DIE KATEGORISIERUNG VON WORTARTEN


3.2.3 Die Wahl einer geeigneten Kategorisierung - Auswahl und Begründung

In diesem Abschnitt möchte ich die Auswahl der für diese Arbeit gewählten Kategorisierung begründen. Ich beziehe mich dabei nur auf die oben aufgeführte Wortartenklassifikation von Kauschke sowie die STTS –
wohlwissend, dass es viele weitere gibt. Doch es ist an dieser Stelle unmöglich, mich auf alle existierenden Klassifikationsvorschläge zu beziehen.³


Kapitel 4

Beobachtungsstudie - Ziele und Methodik


\(^1\)Genauere Informationen zur Kita finden Sie im Internet unter http://www.studentenwerk-berlin.de/kita/standorte/kita_tu/index.html.

\(^2\)Genauere Informationen zu FOLKER finden Sie im Internet unter http://agd.ids-mannheim.de/folker.shtml
relevante Abfragen getätigt und Ergebnisse ausgegeben werden, die über die Wortartenverteilung in verschiedenen Altersclustern sowie über die Type-Token-Verteilung (Paragraph 4.1) Aufschluss geben. Ferner erfolgte eine Klassifizierung aller Lemmata hinsichtlich ihrer Bedeutungskategorie und eine anschließende Zusammenstellung dieser in einem *Lexikon des Vorschulalters* (Kapitel 7; Anhang).

Im Verlauf der Arbeit sollen vor allem folgende Fragen überprüft werden:

1. *Gibt es im Alter von 3;5 bis 5;5 Jahren individuelle Unterschiede in der Verteilung der Wortarten?*

2. *Ist hinsichtlich der Verwendung der Wortarten sowie in der inhaltlichen Entwicklung des Lexikons ein Trend erkennbar (individuell oder allgemein) oder bleibt die Komposition des Lexikons zwischen 3;5 und 5;5 Jahren konstant?*

3. *Wie ist das Lexikon der hier untersuchten Kinder zwischen 3;5 und 5;5 Jahren komponiert?*

### 4.1 Methodik


Ergebnisse, die aber einen ebenso umfangreichen Einsatz der Eltern verlangen.


Hinsichtlich der zeitlichen Aspekte können zum einen Längsschnittstudien, zum anderen Querschnittstudien durchgeführt werden. Bei einer Längsschnittstudie werden einzelne Kinder über einen längeren Zeitraum beobachtet, wodurch sich die Entwicklung der sprachlichen Fortschritte gut verfolgen lässt. Eine Querschnittstudie wird zu einem gegebenen Zeitpunkt mit mehreren Kindern durchgeführt und soll - wie der Name schon sagt - einen Querschnitt über die sprachlichen Fähigkeiten aller Kinder aufzeigen. In der


Die Type-Token-Relation  In diesem und den folgenden Paragraphen soll das Verhältnis der Types (alle verschiedenen Wörter im gegebenen Korpus) zu den Tokens (alle Wörter im gegebenen Korpus) unter allgemeinen Aspekten

„The type-token relation is understood as the ration of the number of different words to all words in the text, or with other words, the ration of vocabulary richness to the text length.“ ([Wimmer (2005)], S. 361)

Wimmer erläutert sehr übersichtlich die unterschiedlichen Interpretationen, Methoden zur Berechnung der Type-Token-Ratio sowie auch Möglichkeiten zur Modellierung. Da der Artikel der Autorin sehr umfassend und übersichtlich gestaltet ist und die für die vorliegende Arbeit nötigen Grundlagen zum Thema Type-Token-Ratio enthält, möchte ich mich im Folgenden darauf stützen ([Wimmer (2005)]). Wimmer geht auf die Vor- und Nachteile jeder Methode im Hinblick auf die jeweils verfolgten Zwecke ein. Demzufolge sei dieses Forschungsinteresse heutzutage nicht nur sehr umfassend, sondern auch äußerst uneinheitlich, da jedes neue Forschungsfeld auch neue Unregelmäßigkeiten mit sich bringe. Dieser Mangel an Uneinheitlichkeit ist auf folgende Aspekte zurückzuführen:

1. Identifikation der Types
2. Interpretation
3. Messung
4. Entstehung und theoretische Annäherung
5. Statistische Prozesse (Stichproben etc.)

Ich möchte nachfolgend nicht auf jeden einzelnen Aspekt eingehen, sondern nur jene betrachten, die für die Zwecke der vorliegenden Arbeit relevant sein können. Deshalb gehe ich vor allem auf die Punkte 1 bis 3 ein und in Annäherung auf Punkt 4.

**Identifikation** Laut Wimmer gibt es zwei Möglichkeiten, die Types in einem Text zu identifizieren.

ii) Es werden unterschiedliche Lexeme betrachtet. Diese Methode führt zu komplexen qualitativen Problemen. Die Daten müssten zuvor adäquat vorbereitet werden (von Hand oder mit Hilfe von Programmen). Diese Vorbereitung aber hängt von der linguistischen Schule ab, die die untersuchende Person genossen hat. Demzufolge können Probleme folgender Art entstehen:

a) Im Deutschen sind die unbestimmten Artikel *ein, eine, einer, eines* vereinigt in dem Lexem *ein*. Wie verhält es sich dann mit den bestimmten Artikeln *der, die, das* oder den dazugehörigen Pronomen *er, sie, es*?


b) Wie geht man um mit analytischen Konstruktionen wie *am schönsten*, was in diesem Fall den Superlativ eines Adjektivs darstellt, dessen andere Formen synthetisch sind?

In der vorliegenden Arbeit kann dieses Problem ebenfalls durch die Struktur der STTS gelöst werden. Für eine solche Konstruktion gibt es klar abgegrenzte POS-Tags. So wird das Tag PTKA für Partikeln in Verbindung mit Adjektiven im Superlativ genutzt, also in diesem Falle für *am*. Handelt es sich bei *am* um eine Präposition, würde es mit APPRART getaggt werden. Zudem behält *am* (PTKA) nach der Lemmatisierung die Wortform *am*, während die Präposition *am zu an* würde, weil es sich in diesem Fall um die Präposition *an* + Artikel handelt. Der Superlativ *schönsten* wird nach der Lemmatisierung in seiner Grundform *schön* notiert und erhält das POS-Tag (ADJA bzw. ADJD).
c) Wie kann das Problem von Verbformen wie in „ich werde gelobt werden“. gelöst werden? Handelt es sich hierbei um zwei Lemmata, drei oder sogar vier?

Diese Verbform tritt in den zu analysierenden Daten dieser Arbeit nicht auf, wodurch das oben genannte Problem irrelevant ist. Für den hypothetischen Fall, dass die aufgeführte Verbform mit Hilfe der STTS getaggt werden würde, ergäbe sich folgende Kennzeichnung: ich (PPER) werde (VAFIN) gelobt (VVPP) werden (VAINF/VVINF). Die dazugehörigen Lemmata wären: ich (ICH) werde (WERDEN) gelobt (LOBEN) werden (WERDEN). Würde an dieser Stelle nur lemmatisiert werden, könnte dies durchaus ein Problem bei der späteren Zuordnung der Lemmata zu den entsprechenden Wortformen geben. Da die jeweiligen POS-Tags jedoch immer mit den Wörtern und den Lemmata in der Datenbank (Kapitel 5) verknüpft sind, ist eine spätere Zuordnung möglich. In der Datenbank wäre leicht erkennbar, dass das erste Lemma WERDEN das Tag VAFIN besitzt und das letzte Lemma WERDEN das Tag VAINF bzw. VVINF.

d) Wie geht man mit dem Problem der Homonymität um, welches durch ein Programm leicht verdeckt bzw. nicht erfasst werden kann, z. B. der Artikel ein, die Zahl ein und das Präfix ein (ich trat ein).

Auch dieses Problem kann unter der Verwendung der STTS und seinen POS-Tags nicht auftreten. Der Artikel wird getaggt mit ART, die Zahl wird getaggt mit CARD und für das Präfix gibt es das Tag PTKVZ. Das Taggen erfolgt selbstverständlich immer von Hand. Würde es automatisiert werden können, wäre eine solch genaue Bezeichnung sicher (noch) nicht möglich. Diese Methode erfolgt - zumindest unter diesen Umständen - auf einer syntaktischen Basis, was im Grunde den Zielen dieser Arbeit widerspricht. Da jedoch die STTS die Basis der Wortartenbestimmung bilden, muss in diesen Fällen von einer rein lexikalisch basierten Wortartenbestimmung abgesehen werden.

e) Wie kann das Problem femininer Formen³ von Nomen gelöst werden, z. B. beau und belle oder actore und actrice?

Die Verwendung von Nomen in der femininen respektive maskulinen Form stellt auch in der vorliegenden Arbeit zum Teil ein Problem dar. In fast allen

Fällen handelt es sich dabei um Nomen, die entweder nur in der einen oder in der anderen Variante verwendet wurden. Es wurde das Wort *Erzieherin* geäußert, nicht aber *Erzieher*. Oder es kam zu Äußerungen wie *Krieger*, nicht aber *Kriegerin*. In diesen Fällen, in denen also die Verwendung der femininen oder maskulinen Form möglich ist, aber nur eine Form genutzt wurde, wurde diese Form auch nach der Lemmatisierung beibehalten.

**Interpretation** Hinsichtlich der Interpretation der Type-Token-Ratio suggeriert Wimmer zwei Möglichkeiten:

i) Das Verhältnis ist charakteristisch für die Vielfalt des Wortschatzes eines Textes.

ii) Das Verhältnis ist ein Modell des Informationsflusses im Text.

Type-Token-Ratio als Messung der Wortschatzvielfalt in einem Text fungieren ([Wimmer (2005)], S. 362).


**Messung** Hinsichtlich der Messung der Type-Token-Relation zeigt Wimmer (2005) drei Methoden auf:

i) Der Position jedes Tokens (x) wird die Anzahl der Types (y) bis zu dieser Position zugeordnet.

Dies sei die gebräuchlichste Methode, die wiederum einen monoton steigenden Graphen generiert (<token, type>).

ii) Das Dividieren der Anzahl der Types bis zur Position x durch die Position selbst.

Dies ergebe eine monoton fallende Funktion (<token, type/token>).

iii) Die Köhler-Galle-Methode ([Köhler und Galle (1993)])

\[
TTRx = \frac{tx + T - \frac{xT}{N}}{N}
\]

wobei: \(x\) = Position im Text (Anzahl von Tokens bis zu dieser Position x),
tx = Anzahl von Types bis zur Position x (inkl.), T = Anzahl von Types im gesamten Text, N = Textlänge (Anzahl von Tokens im gesamten Text); <x, TTRx>.

Methode i) ist für die zu analysierenden Daten dieser Arbeit nicht praktikabel, da es sich um annähernd 30.000 Tokens handelt. Auch Methode iii) erschien für die hier vorliegenden Daten nicht passend zu sein, da es sich um gesprochene Daten handelt, die nur schwer in eine bestimmte Textlänge umgerechnet werden können. Vielmehr wurde aus praktikablen Gründen Methode ii), wenn auch in abgewandelter Form, durchgeführt.

Entwicklung und Modellierung Wimmer beschreibt drei Annäherungen, um die Type-Token-Ratio eines Textes zu beschreiben:


2) Entwurf eines Graphen, der den Informationsfluss eines Textes erfasst.

3) Ausgehend von einem statistischen Prozess werden neue Wörter generiert und die Ergebnisse für unterschiedliche Zwecke verwendet (Wortschatzreichtum, Prognosen, Type-Token-Ratio, Häufigkeitsverteilungen).

Tatsache sei, dass all diese Hinweise, Graphen, Verteilungen etc. akzeptabel sind. Problematisch ist das, was wir erfassen wollen und welche theoretische Tiefe wir erreichen möchten ([Wimmer (2005)], S. 364).


---

des jeweiligen Kindes unterworfen, wodurch es zu Einbußen bei der Aussagegenauigkeit kommen kann.

Trotz aller Kritik an der Berechnung der TTR soll diese in der vorliegenden Arbeit im weiteren Verlauf zu einem kleinen Teil ermittelt werden (Abschnitt 6.2.3). Zum einen aus purer Neugier meinerseits, welche Ergebnisse zu Tage treten; zum anderen bin ich daran interessiert, ob sich bei den Kindern in den drei Altersgruppen Unterschiede auftun.


4.1.1 Sprachaufnahmen


<table>
<thead>
<tr>
<th>Nummer der Aufnahme</th>
<th>Beschreibung des Kontextes zum Zeitpunkt der Aufnahme</th>
<th>Datum der Aufnahme</th>
</tr>
</thead>
<tbody>
<tr>
<td>009_AV, 009_MM</td>
<td>Gespräch mit AV und MM</td>
<td>15.09.2009</td>
</tr>
<tr>
<td>011_MM</td>
<td>Karten spielen mit MM</td>
<td>15.09.2009</td>
</tr>
<tr>
<td>013_JS, 013_RD</td>
<td>Im Sandkasten mit JS und RD</td>
<td>15.09.2009</td>
</tr>
<tr>
<td>015_MK, 016_MK, 017_MK</td>
<td>Im Garten mit MK</td>
<td>15.09.2009</td>
</tr>
<tr>
<td>019_JS, 019_LEO, 019_MK, 019_RD</td>
<td>Im Garten mit JS, RD, MK und LEO</td>
<td>15.09.2009</td>
</tr>
<tr>
<td>Nummer der Aufnahme</td>
<td>Beschreibung des Kontextes zum Zeitpunkt der Aufnahme</td>
<td>Datum der Aufnahme</td>
</tr>
<tr>
<td>---------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------</td>
<td>--------------------</td>
</tr>
<tr>
<td>022_LAR, 022_MM</td>
<td>Gespräch mit LAR und IV</td>
<td>22.09.2009</td>
</tr>
<tr>
<td>023_MK, 023_JS, 024 LUA, 024_MK, 025 LUA</td>
<td>MK und LUA spielen Bauernhof; spielen mit Steinen; spielen Ritterburg</td>
<td>29.09.2009</td>
</tr>
<tr>
<td>026_LUA, 026_MK, 027_AV, 027 LUA, 030 LUA, 030_MK</td>
<td>MK spielt Bauernhof; später Imbiss</td>
<td>06.10.2009</td>
</tr>
<tr>
<td>038_JK, 038_LL, 038_MK</td>
<td>MM, LL spielen Memory. Danach Versteckspiel mit MM und LL. Später spielen andere Kinder Autobahn.</td>
<td>01.12.2009</td>
</tr>
<tr>
<td>041_AV, 041_JS, 041_LAR, 041_LEO, 041_MK, 041_MM, 041_RD, 041_SO</td>
<td>Kinder malen mit Glitzerfarbe und Kreide. MM hat sich gestoßen. Spielen mit LAR, MM und AV (Memory). Danach Malen. Vorlesen mit MK und LEO.</td>
<td>05.01.2010</td>
</tr>
<tr>
<td>Nummer der Aufnahme</td>
<td>Beschreibung des Kontextes zum Zeitpunkt der Aufnahme</td>
<td>Datum der Aufnahme</td>
</tr>
<tr>
<td>---------------------</td>
<td>------------------------------------------------------------------------------------------------------------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>042_JS, 042_LEO, 042_LUA, 042_MK, 042_MM</td>
<td>MK erzählt, welches Tier er gerade ist. LEO und MK spielen Kap'n Sharky. Uno spielen mit MK. Danach Spiel mit Magneten.</td>
<td>12.01.2010</td>
</tr>
<tr>
<td>051_LUA, 051_MK</td>
<td>Spielen mit MK, JK, RD und JS. Andere Kinder toben umher. Türme bauen mit JK.</td>
<td>24.02.2010</td>
</tr>
<tr>
<td>052_AV, 052_JS, 052_LEO, 052_LUA, 052_MK, 052_MM, 052_RD, 052_SO</td>
<td>MM, LL, SO und LAR basteln Perlenketten. Memory spielen mit MM. Malen mit AV. Buch ansehen mit MK, RD, JS und LEO. Zwischendurch spricht LUA.</td>
<td>03.03.2010</td>
</tr>
<tr>
<td>053_AV, 053_JS, 053_LEO, 053_LUA, 053_MK, 053_RD</td>
<td>MM, SO und LAR basteln Perlenketten. Memory spielen mit MM. Zwischendurch spricht LUA. Später malen mit AV. Imbiss: AV, MK und LUA erzählen.</td>
<td>10.03.2010</td>
</tr>
<tr>
<td>Nummer der Aufnahme</td>
<td>Beschreibung des Kontextes zum Zeitpunkt der Aufnahme</td>
<td>Datum der Aufnahme</td>
</tr>
<tr>
<td>---------------------</td>
<td>------------------------------------------------------</td>
<td>-------------------</td>
</tr>
</tbody>
</table>

**4.1.2 Die Probandinnen und Probanden**

Tabelle 4.2 veranschaulicht die Zusammensetzung der Gruppe der Testpersonen (untersuchte Kinder).


Neben dem sozioökonomischen Status können relativ genaue Angaben über die Familienstruktur vorgenommen werden, insbesondere was die Geschwisterkinder angeht. Bis auf zwei Kinder haben alle Kinder mindestens ein Geschwisterkind (ein Kind bleibt ohne Angabe). Der Einfluss der Größe der Familie auf den Spracherwerb wurde schon früh von Stern&Stern (1928, 1965) untersucht, die ermittelten, dass das Vorhandensein älterer Geschwister die sprachliche Entwicklung vorantreibt ([Stern und Stern (1928)]; [Stern und Stern (1965)], S. 293 f.). McCarthy (1954) hingegen fand heraus, dass Kinder, die vorwiegend mit Erwachsenen verkehren, die schnellste sprachliche Entwicklung vorweisen können. Einzelkinder (insbesondere Mädchen) seien dabei in allen Aspekten am weitesten fortgeschritten.
In der vorliegenden Arbeit wurden derartige Einflüsse nicht untersucht. Ein Grund dafür war, dass das Sprachverhalten aller Kinder keinen Anlass zu einer Untersuchung des Einflusses der Geschwisterkinder auf die Sprachentwicklung gab. Zudem befinden sich alle Kinder in einem weitaus fortgeschrittenem Stadium des Spracherwerbs, so dass ein Urteil nicht eindeutig wäre.

soll deshalb nicht durchgeführt werden.

4.1.3 Diktiergerät


4.1.4 Transkription der Daten mit FOLKER


Bindestrich obligatorisch ist, als zwei Wörter transkribiert (z. B. baden württemberg). Wörter, bei denen der Bindestrich optional ist, werden als ein Wort transkribiert (z. B. bildzeitung).

- Verzögerungssignale: werden ausgeschrieben (z. B. äh ähm).

- Lachen, Weinen: wird entweder als Beschreibung (lacht), (weint) oder bei silbigem Lachen entsprechend der verwendeten Silben als hahaha oder hihi etc. notiert.

- Rezeptionssignale: kommen in den Daten der vorliegenden Arbeiten gelegentlich vor. Sie werden in Anlehnung an die Regeln der FOLKER-Transkription als hm, nee, nee_e, ja_a etc. notiert. Bei der späteren Lemmatisierung (Abschnitt 4.1.8) werden zweisilbige Signale wieder in ihrer einsilbigen Form notiert.

- Pausen: werden in den FOLKER-Dateien mit (.) für eine Mikropause (ca. 0,2 Sekunden Dauer) bzw. als längere geschätzte Pause durch eine konkrete Angabe notiert (1.5). Im FOLKER-Transkriptionshandbuch sind weitere Pausenarten angegeben; dazu gehören eine kurze geschätzte Pause von 0,2 bis 0,5 Sekunden Dauer (-), eine mittlere geschätzte Pause von 0,5 bis 0,8 Sekunden Dauer (- -) und eine längere geschätzte Pause von 0,8 bis 1,0 Sekunden Dauer (- - -). Diese Pausenarten wurden in den Daten der vorliegenden Arbeit nicht verwendet, da sie zum einen laut dem Transkriptionshandbuch nicht vorzugsweise verwendet werden sollten und zum anderen für die Zwecke dieser Arbeit nicht in erster Linie relevant waren. Hier werden vordergründig die syntaktischen Wörter an sich betrachtet und diese werden zum Erreichen der Ziele anschließend in den EXMARaLDA-Partitur-Editor (Abschnitt 4.1.6) importiert und weiter verarbeitet. Alle Pausen sind ab diesem Moment hinfällig, weshalb sie in erster Linie nur nebenbei gesetzt wurden.

- Nonverbales: wird nach den cGAT-Konventionen durch doppelte runde Klammern gekennzeichnet, z. B. ((hustet)) oder ((unverständlich)). In Anbetracht der Wichtigkeit der festgelegten Wörter, nicht aber nebensprachlicher Erscheinungen, wurden in dieser Arbeit lediglich unverständliche Wörter mit ((unverständlich)) gekennzeichnet.

- Unverständliches: Ein unverständliches Wort kann nicht nur durch doppelte runde Klammern gekennzeichnet werden. Insofern die Silbenanzahl erkennbar ist, kann diese durch die Zeichenkette +++ (für eine Silbe) ++++++ (für zwei Silben) etc. dargestellt werden.

KAPITEL 4. BEOBACHTUNGSSTUDIE - ZIELE UND METHODIK


Abbildung 4.1: Folker-Datei: ic_a_31.flk

FOLKER-Dateien können in unterschiedlichen Html-Formaten ausgegeben sowie zur weiteren Bearbeitung in anderen Programmen genutzt werden. Im Folgenden ist eine beispielhafte Erstellung einer FOLKER-Transkription für diese Arbeit beschrieben:


8In der vorliegenden Arbeit werden die in FOLKER transkribierten Dateien zur weiteren Analyse in den EXMARaLDA Partitur Editor überführt (Abschnitt 4.1.6).
4.1. METHODIK


4.1.5 Bearbeitung der Audio-Dateien mit AUDACITY

Aus Gründen der anschließenden Analyse, aber auch aus Datenschutzgründen, wurde in allen überflüssigen Passagen der Ton entfernt. Ein Ausschneiden oder Löschen der Passagen war nicht möglich, da die FOLKER-Dateien in diesem Fall nicht mehr mit den dazu gehörigen Wave-Dateien kompatibel gewesen wären, was das Abspielen, Ansehen oder Bearbeiten der FOLKER-Dateien unmöglich gemacht hätte. AUDACITY bietet die Möglichkeit, einzelne Passagen mit einem sogenannten *Silence-Modus* zu versehen, wodurch an den ausgewählten Stellen Stille erzeugt wird. Die FOLKER-Datei ic_a_16.flk sieht im Programm Audacity nach der Bearbeitung entsprechender Passagen mit dem Silence-Modus so aus:

![Abbildung 4.2: Folker-Datei: ic_a_16.flk](image)

Die grau markierte Passage (01m06.81s bis 01m26.75s) zeigt einen Bereich, in welchem der Ton entfernt wurde. Dies stellt lediglich ein Beispiel zur Veranschaulichung dar, alle weiteren überflüssigen Bereiche sind im Beispiel aus Gründen der Übersichtlichkeit noch nicht grau markiert. Da sich die Bearbeitung aller FOLKER-Dateien mit AUDACITY als sehr aufwendig herausstellte, konnte ein OpenSource-Programm\textsuperscript{10} genutzt

\textsuperscript{10}Das Programm zur Korpusanalyse wurde für die Ziele dieser Arbeit von Alexander Mack entwickelt und programmiert. Es ist als OpenSource-Programm unter folgender Adresse zu finden: https://github.com/fiedler-mack/CorpusAnalyser. Eine genaue Beschreibung der einzelnen Funktionen

4.1.6 Bearbeitung der Transkripte mit dem EXMARaLDA Partitur-Editor

Nachdem eine geeignete Methode zur Bestimmung der Wortarten gefunden war (Abschnitt 3.2.3), musste ein Bearbeitungsprogramm gefunden werden, das die Bezeichnung der Wortarten (POS-Tags) mit den dazugehörigen Wörtern und später auch Lemmata vereint. FOLKER stellte sich dafür als ungeeignet heraus. Mit dem Programm selbst können zwar Transkripte erstellt und zu Text- und Html-Dateien exportiert werden, eine weitere Bearbeitung ist allerdings nicht möglich. Der EXMARaLDA Partitur-Editor\footnote{Für weitere Informationen und für die Möglichkeit des Downloads verweise ich auf folgende Webseite: \url{http://www.exmaralda.org/partitureditor.html} (letzter Zugriff 22.10.2014)} erwies sich hingegen als geeignet. Laut den Entwicklern ist der Editor nicht nur zum Datenaustausch mit vielen anderen Systemen kompatibel, sondern erlaubt eine Analyse der Wortarten unter unterschiedlichen Transkriptionskonventionen:


Zunächst wurde jede der 39 Dateien in den Editor importiert. Im Editor selbst werden die Tonspuren der jeweils teilnehmenden Kinder untereinander in den entsprechenden Zeilen angezeigt. Pro FOLKER-Datei wurden meistens mehrere Kinder transkribiert. Dies ist logischerweise in jeder FOLKER-Datei der Fall, da während der Aufnahmesituationen in der Kita immer mehrere

Abbildung 4.3: Beispielhafte EXMARaLDA-Datei mit sieben Sprecherinnen und Sprechern

Abbildung 4.4: Separierte Tonspur in einer EXMARaLDA-Datei

Anschließend wurden alle Lücken und unbenutzten Zeitpunkte in den Tonspuren entfernt und die syntaktischen Wörter voneinander separiert. Zusätzlich wurde eine Spur (Annotation) angefügt, in welche später die
entsprechenden POS-Tags geschrieben werden konnten (Abbildung 4.5 und Abschnitt 3.2.2).

Abbildung 4.5: Tonspur eines einzelnen Sprechers (RD) mit Annotationsspur

Auf diese Weise entstanden 155 Einzeldateien von insgesamt 12 Kindern zwischen 3;5 und 5;5 Jahren. Aufgrund der Fülle der Daten wurden mehrere (Korrektur-)Durchläufe vorgenommen. Im Anschluss an die Bezeichnung der syntaktischen Wörter mit den jeweiligen POS-Tags wurde eine dritte Spur in jede der Dateien eingefügt, in die das zum Wort entsprechende Lemma eingetragen wurde (Abbildung 4.6).

Abbildung 4.6: EXMARaLDA-Datei: MM_lem.exb

Eine Lemmatisierung erwies sich als notwendig, weil ohne diese bei der späteren Analyse eine irrtümliche Type-Token-Zuordnung entstehen würde. Eine Relation der POS-Tags zu den jeweiligen syntaktischen Wörtern würde bedeuten, dass zum Beispiel alle Flexionsformen eines Verbs als unterschiedliche Lexeme definiert würden (z. B. *spielen - spielte - gespielt*). Tatsächlich handelt es sich bei diesem Beispiel um nur ein Lemma, nämlich *spielen*. Ebenso würden alle deklinierten Nomen und Artikel als unterschiedlich erkannt, auch wenn sie demselben Lemma angehören. In dem Beispiel „Ich habe der Frau das Geld gegeben“, gehört „der“ zum Lemma „die“ bzw. „d“. Ein Programm, welches alle Wörter nur aufgrund ihrer äußeren Form erkennt und auswertet, erkennt dies nicht. Deshalb mussten im nächsten...

- **das** (syntaktisches Wort)
- **ART** (POS)
- **das** bzw. **d** (Lemma)

Kontextuell betrachtet handelt es sich zwar um ein Demonstrativpronomen - denn es fehlt das darauf folgende Nomen. Doch um dem Anspruch einer lexikalisch basierten Analyse gerecht zu werden, wurde tatsächlich mit ART getaggt sowie mit „das“ bzw. „d“ lemmatisiert (siehe auch Abschnitt 4.1.7 und 4.1.8), da in diesem Fall eine Ellipse vorliegt.

### 4.1.7 Kriterien für das Taggen der Wortarten nach den Richtlinien der STTS

In diesem Abschnitt möchte ich alle Kriterien anführen, nach denen ich mich beim Taggen der syntaktischen Wörter mittels der STTS (Abschnitt 3.2.2) gerichtet habe. Die Stuttgart Tübingen Tagsets bieten zahlreiche Beschreibungen und Hinweise, nach denen ein Wort mit einem bestimmten POS-Tag bezeichnet werden sollte und diese wurden auch hier zu großen Teilen verwendet ([Schiller, Teufel, Stöckert und Thielen (2009)])

geeignet, um alle Wörter im Korpus zu bezeichnen - auch nach überwiegend lexikalischen Kriterien. Durch ein anschließendes Zusammenführen der POS-Tags in größere Klassen (Tabelle 6.1) konnte der syntaktische Aspekt jedoch größtenteils, wenn auch nicht in allen Fällen, umgangen werden. In Anlehnung an die STTS möchte ich im Folgenden alle Tags mit den Definitionen der Dudengrammatik in Einklang bringen. Die Reihenfolge ist an die Abfolge der POS-Tags in der entsprechenden Tabelle (Tabelle 3.2) angelehnt.


\textsuperscript{12}Mir ist bewusst, dass durch die Unterscheidung zwischen ADJA und ADJD ein syntaktisches Kriterium zugrunde gelegt wird. In der späteren Analyse werden die Lemmata jedoch zusammengefasst als ADJ betrachtet, wodurch distributionelle Aspekte wieder vernachlässigt werden.
zwischen Adverb (ADV) und Pronominaladverb (PAV). Beim Taggen
nach den Regeln der STTS unterscheide ich dementsprechend ebenso
nach ADV und PAV, wobei die Pronominaladverbien später zu den
Adverbien gezählt werden. Zur Kategorie ADV zählen demnach alle
nach Duden unterschiedenen Adverbien, bis auf das Pronominaladverb.
Das Pronominaladverb (auch Präpositionaladverb) wird laut Duden
mit einer Präposition als zweitem Bestandteil und mit da(r)-, hier- oder
wo(r)- als erstem Bestandteil gebildet ([Dudenredaktion (2009)], S.
579 sowie Regel 860). Auf Basis dieser Definition wurden die Wörter
im Korpus als PAV identifiziert und mit PAV getaggt. Ebenso wurde
trotzdem und außerdem als PAV getaggt, da es in den STTS als
Pronominaladverb aufgeführt wird. In der Dudengrammatik entspricht
es den Konjunktionaladverbien (Regel 864), zu denen noch weitere
Adverbien gehören.

- Präpositionen (APPR, APPRART, APPO und APZR): Im hier
untersuchten Korpus kommen lediglich Präpositionen (APPR) sowie
Präpositionen mit Artikel (APPRART) vor. Postpositionen (APPO)
und Zirkumpositionen rechts des Bezugswortes (APZR) kommen
nicht vor. Die Bezeichnung der Wörter mit den POS-Tags APPR bzw.
APPRART wurde größtenteils in Anlehnung an die Ausführungen
der Dudengrammatik vorgenommen ([Dudenredaktion (2009)], S.
600-619). Bei der späteren Analyse (Kapitel 6) wurden die Präpositionen
zusammengefasst zum POS-Tag APPR.

- Artikel (ART): Die Dudengrammatik führt den definiten Artikel
([Dudenredaktion (2009)], S. 291-302) getrennt vom indefiniten Artikel
auf ([Dudenredaktion (2009)], S. 330-337). In den STTS werden
der definite und der indefinite Artikel unter dem POS-Tag ART
zusammengefasst. In der vorliegenden Arbeit wurden die Artikel
demnach mit ART bezeichnet, unabhängig davon, ob es sich um einen
definiten oder einen indefiniten Artikel handelt. Nicht nur im vorliegenden
Korpus fiel es nicht immer leicht zwischen einem Artikel oder einem
Pronomen zu unterscheiden. Auch in der Dudengrammatik befindet
sich der Eintrag zum Artikel umrahmt von zahlreichen Einträgen zu
den verschiedenen Pronomen des Deutschen. Im Grunde erfüllen viele
Pronomen (z. B. Indefinitpronomen) oft die Funktion eines indefiniten
Artikels. Und obwohl diese Arbeit die Lemmata auf der Basis der
Wortform bzw. in Bezug auf ihre Form im Lexikon untersuchen soll,
ist es an dieser Stelle nahezu unmöglich, syntaktische Gesichtspunkte
außer Acht zu lassen. In der Dudengrammatik sind alle Formen


- **Fremdsprachliches Material (FM):** Mit dem POS-Tag FM wurden alle Wörter bezeichnet, die nicht in deutscher Sprache geäußert wurden.


4.1. METHODIK


wurde an diesen (wenigen) Stellen kontextuell analysiert. Durch das spätere Zusammenführen der POS-Labels PDS+PDAT+PIDAT+PIS+PIAT zu den PIDATS wurde diese teilweise syntaktische Analyse jedoch wieder vernachlässigt, weil dadurch die einzelne Wortart wieder in den Vordergrund rückte. Das Wort *beide* ist unter der Regel 413 unter Umständen als Kardinalzahl einzustufen. Dies ist in der vorliegenden Arbeit nicht der Fall und *beide* ist Indefinitpronomen. In allen anderen Fällen wurden die Indefinitpronomen analog zu den Regeln der Dudengrammatik als PIS (*substituierendes Indefinitpronomen*), PIAT (*attribuierendes Indefinitpronomen*) oder PIDAT (*attribuierendes Indefinitpronomen mit Determiner*) getaggt.

- **irreflexives Personalpronomen (PPER):** Die *irreflexiven Personalpronomen* wurden analog zu der Regel der Dudengrammatik mit PPER getaggt ([Dudenredaktion (2009)], S. 263-271).


- **Interrogativpronomen** (PWAT, PWAV und PWS): Die *attribuierenden (PWAT) und substituierenden Interrogativpronomen* (PWS) bzw.


\textsuperscript{13}An dieser Stelle wurde die Kategorisierung der STTS-Guidelines ([Schiller, Teufel, Stöckert und Thielen (2009)], S. 54) beibehalten, womit alle POS-Tags der Kategorie PWAV sowohl Interrogativ- als auch Relativpronomen sein können.
Es macht Spaß, ihr vorzulesen → ART VVFIN NN PPER VVIZU). Die Gradpartikeln, Fokuspartikeln und Abtönungspartikeln entsprechen weitestgehend den Partikeln bei einem Adjektiv oder Adverb in den STTS (PTKA). Eine weitere Kategorie stellen die Gesprächspartikeln dar (Regel 800), die je nach Kontext den Antwortpartikeln (PTKANT) entsprechen. In den meisten Fällen wurden sie aber mit PTKA bezeichnet werden, nämlich dann, wenn sie tatsächlich die Funktion eines Füllwortes bzw. einer Partikel übernehmen, wie im folgenden Beispiel: Der (ART) Ball (NN) ist (VVFIN) ja (PTKA) klein (ADJD)!

- Verben (VVFIN, VVIMP, VVINF, VVIZU, VVPP, VAFIN, VAIMP, VAINF, VAPP, VMFIN, VMINF, VMPP): Die Dudengrammatik widmet sich in einem umfassenden Eintrag der Wortart Verb ([Dudenredaktion (2009)], S. 389-566). Dabei unterscheidet sie zunächst nach Bedeutung und Funktion unterschiedlicher Verbarten (ebd., S. 390-429), was für die Zwecke der Wortartenanalyse als Teilziel dieser Arbeit nicht vordergründig relevant ist. Ich möchte an dieser Stelle deshalb nicht näher darauf eingehen und erst in einem späteren Kapitel wieder darauf zu sprechen kommen (Kapitel 7). Auf den Seiten 429 bis 476 beschreibt die Dudengrammatik jene Verbenformen, die in ihrer Kategorisierung durch die STTS dem Korpus dieser Arbeit am ehesten entsprechen. Unter der Regel 596 werden die finiten Kategorien vorgestellt, die in den STTS den finiten Vollverben (VVFIN), finiten Auxiliarverben (VAFIN) und finiten Modalverben (VMFIN) entsprechen. Unter der Regel 597 werden die infiniten Kategorien aufgeführt, die am ehesten mit den Bezeichnungen infinites Vollverb (VVINF), infinites Auxiliarverb (VAINF) und infinites Modalverb (VMINF) gleichzusetzen sind. Die Partizipien VAPP (Partizip Perfekt, auxiliar), VVPP (Partizip Perfekt, voll) sowie VMPP (Partizip Perfekt, modal) finden sich am ehesten in der Regel 598 wieder. Mehrteilige Verbenformen, wie sie in dieser Regel zu finden sind, werden in den STTS mit Hilfe der bisher aufgeführten POS-Tags einzeln getaggt, da es um die Wortart an sich gehen soll, nicht aber um das funktionale Gefüge. Der Konditionalssatz „Sie würde dir meine Adresse nicht geben.“ würde nach den Regeln der STTS mit PPER (sie) VAFIN (würde) PPER (dir) PPOSAT (meine) NN (Adresse) PTKNEG (nicht) VVINF (geben) getaggt werden. In Bezug auf den Infinitiv mit zu könnte ein Tagging-Vorgang folgendermaßen aussehen: PPOSAT (mein) NN (Ziel) VVFIN (ist) PPER (es), VVIZU (anzukommen).

Eine Besonderheit betrifft die Auxiliarverben, wenn diese sich in ihrer infiniten Form befinden. Den STTS zufolge werden diese dann als
4.1. METHODIK


finite Vollverb, also VVFIN. Losgelöst vom Kontext könnte es ebenso ein infinites Vollverb, also VVINF, darstellen. An dieser Stelle wurden alle Verben zunächst unter Berücksichtigung des Kontextes bezeichnet und erst später zu Unterkategorien zusammengefasst, wodurch dem lexikalisch begründeten Anspruch wieder Rechnung getragen wird (Tabelle 6.1). Da ich mich im Vorfeld dazu entschieden habe, die STTS als populäres Tagging-System für meine Analyse zu nutzen, hielt ich es dennoch für sinnvoll, dass auch die Möglichkeit einer späteren syntaktischen Analyse gegeben ist.


- **Auxiliarverb** (VAFIN, (VAINF), VAIMP, (VAPP))
- **Modalverb** (VMFIN, VMINF, VMPP)
- **Vollverb** (VVFIN, VVINF, VVPP, VVIMP, VVIZU)

### 4.1.8 Kriterien für die Lemmatisierung der syntaktischen Wörter

Analog zu den Kriterien für das Bezeichnen der Wortarten mit den POS-Tags in Abschnitt 4.1.7 möchte ich an dieser Stelle die Kriterien anführen, nach denen die geäußerten, transkribierten und voneinander separierten Wörter lemmatisiert wurden, wobei ich insbesondere auf die Sonderfälle eingehen möchte. Bei der Lemmatisierung stütze ich mich vor allem auf die Regeln der eben zitierten Dudengrammatik ([Dudenredaktion (2009)])

Die POS-Tags vor den Klammern bilden die Tags in ihrer gruppierten Form ab (Tabelle 6.1), die POS-Tags in Klammern sind alle hier vorkommenden, nicht zusammengefassten Wortarten (Tabelle 3.2).

- **ADJ** (ADJA und ADJD): Adjektivbezeichnungen sind in ihrer lemmatisierten Form immer im Positiv angegeben. Grenzfälle gab es wie bereits weiter oben beschrieben bei substantivierten Adjektiven,

- **ADV** (ADV und PAV): Adverbien und Pronominaladverbien unterliegen keinen besonderen Grenzfällen. Sie bilden bereits als Wort die lemmatisierte Form ab und wurden dementsprechend auch so übernommen.

- **APPR** (APPR, APPRART, APPO, APZR): Präpositionen ohne Artikel (APPR) behielten ihre Form, die sie als Wort aufweisen, auch als Lemma bei. Präpositionen mit Artikel (APPRART) verloren in der lemmatisierten Form ihren Artikel (beim → bei, ans → an). Die Präpositionen APPO und APZR kommen im Korpus nicht vor.

- **ART**: Die bestimmten Artikel (der, die, das, die (Pl.)) wurden als d (im Singular) bzw. d-p (im Plural) lemmatisiert. Die unbestimmten Artikel (ein, eine) wurden zu ein.

- **CARD**: Da die Kardinalzahlen häufig einen adjektivischen Charakter aufweisen, wurden sie in ihrer Form als Lemma ebenso nach den Regeln der Adjektivlemmatisierung lemmatisiert. Die Ordnungszahlen veränderten sich folgendermaßen: erste → erst, dritte → dritt, wobei die Grundzahlen ihre Form als Wort auch als Lemma beibehielten: eins → eins, zweihundert → zweihundert.

- **FM**: Fremdsprachliches Material kommt im gesamten Korpus nur sehr selten vor und blieb in der lemmatisierten Form in seiner fremdsprachlichen Wortform erhalten.

- **ITJ**: Interjektionen erfuhren keine Änderung in der lemmatisierten Form.

- **KON** (KOUI, KOUS, KOKOM, KON): Konjunktionen erfuhren keine Änderung in der lemmatisierten Form.

- **N** (NE und NN): Nomen (NN) und Eigennamen (NE) wurden als Lemma in der Nominativform im Singular angeführt.

• **PPER**: Die *irreflexiven Personalpronomen* blieben immer im Nominativ (Singular oder Plural): z. B. *ich → ich, ihr (2. P. Pl.) → ihr, ihm → er, uns → wir*.

• **PPOS** (PPOSS und PPOSAT): Die *substituierenden* (PPOSS) und *attribuierenden Possessivpronomen* (PPOSAT) blieben immer
im Nominativ (Singular) in Anlehnung an die Dudengrammatik ([Dudenredaktion (2009)], S. 276): meine → mein, meinem → mein, unsere → unser, ihr → ihr.

- **PREL** (PRELS und PRELAT): Das *attribuierende Relativpronomen* (PRELAT) kommt im vorliegenden Korpus nicht vor. Das *substituierende Personalpronomen* (PRELS) blieb in der lemmatisierten Form immer im Nominativ.


- **PW** (PWS, PWAT, PWAV): Die *substituierenden Interrogativpronomen* (PWS) stehen aufgrund ihres Vorkommens durch die Formen *wer, was* und *welch-* in der lemmatisierten Form ebenfalls als *wer, was* oder *welch-*.

- **PTK** (PTKA, PTKANT, PTKNEG, PTKVZ, PTKZU): Die *Partikeln* wurden aufgrund ihrer unveränderlichen Form als Wort dementsprechend in der lemmatisierten Form aufgeführt.

- **VA** (VAFIN, VAIMP, (VAINF), (VAPP)): Alle *Auxiliarverben* wurden in der lemmatisierten Form immer im Infinitiv aufgeführt.

- **VM** (VMFIN, VMINF, VMPP): Alle *Modalverben* wurden in der lemmatisierten Form immer im Infinitiv aufgeführt.

- **VV** (VVFIN, VVINF, VVIMP, VVPP, VVIZU): Alle *Vollverben* wurden in der lemmatisierten Form immer im Infinitiv aufgeführt.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>JS</td>
<td>m</td>
<td>deutsch</td>
<td>dt.</td>
<td>k.A.</td>
<td>2004-12-31</td>
<td>geb. 1974, Deutschlehrerin</td>
<td>geb. 1977, Psychologe</td>
<td>keine</td>
</tr>
<tr>
<td>JK</td>
<td>m</td>
<td>deutsch</td>
<td>dt.</td>
<td>k.A.</td>
<td>2004-09-20</td>
<td>k.A.</td>
<td>k.A.</td>
<td>k.A.</td>
</tr>
</tbody>
</table>

Tabelle 4.2: Gruppe der Testpersonen
Kapitel 5

Die Arbeit mit dem CorpusAnalyser

In diesem Kapitel möchte ich den CorpusAnalyser vorstellen und insbesondere die Anwendung im Hinblick auf die Zwecke der vorliegenden Arbeit dokumentieren. Der CorpusAnalyser ist ein Open-Source-Programm und wurde von Alexander Mack ([Mack (2014)]) den Anforderungen dieser Arbeit entsprechend programmiert. Es kann kostenfrei verwendet und/oder modifiziert werden nach den Richtlinien der GNU General Public License\(^1\). Die programmierte Datenbank basiert auf der Sprache SQLite3, die dazugehörigen Skripte sind Perl- bzw. Bash-Shell-Skripte. In meiner nun folgenden Beschreibung möchte ich detailliert auf die einzelnen Funktionen eingehen, wobei ich mich an die Struktur halte, die der genannten Webseite zugrunde liegt.\(^2\)

Grundlage für die Arbeit mit dem CorpusAnalyser (folgend CA) bilden XML-Dateien, die zuvor in FOLKER (Abschnitt 4.1.4) oder EXMARaLDA (Abschnitt 4.1.6) bearbeitet wurden. Relevante Teile der jeweiligen XML-Datei wurden anschließend in eine SQLite3-Datenbank (Mack (2014)) überführt. Die Datenbank selbst wurde als Datei auf einer Festplatte gespeichert. Nachdem die Daten erstellt und in die Datenbank importiert worden sind, konnten spezifische Abfragen an die Datenbank gestellt werden. Der Output wurde als CSV-Datei\(^3\) in ein Zielverzeichnis geschrieben. Daneben existieren Skripte, mit denen Audio (.wav)-Dateien modifiziert werden können. Auf diese Weise konnten zum Beispiel nicht verwendete Teile der Audio-Dateien, die im Zuge dieser Arbeit entstanden, stumm geschaltet werden, wobei die Informationen dafür aus einer FOLKER-Datei gezogen wurden.

\(^1\)Weitere Informationen unter http://www.gnu.org/copyleft/gpl.html (zuletzt besucht am 22.10.14)
\(^2\)Alle Abbildungen in diesem Kapitel mit Bezug zum CA stammen von der Webseite: https://github.com/fiedler-mack/CorpusAnalyser.
\(^3\)https://de.wikipedia.org/wiki/CSV_(Dateiformat) (letzter Zugriff 22.10.2014)

5.1 Die Inhalte im Überblick

Project folder structure

Die unten stehende Abbildung (Abbildung 5.1) veranschaulicht die für die Datenbank empfohlene Ordnerrstruktur (folder structure). Diese kann geändert werden, wenn Skripte manuell aufgerufen oder Pfade in den Hilfsskripten (xxx_all.sh) manuell bearbeitet werden.

Wie zu erkennen ist, existiert ein übergeordneter Ordner project root folder mit mehreren Unterordnern. Der erste Unterordner ist zwangsläufig der CA mit dem dazugehörigen Skript 01_wav_convert (siehe: 01_wav_convert). Die einzelnen Skripte und deren Einsatz werden an späterer Stelle noch besprochen (Abschnitt 5.2).
5.1. DIE INHALTE IM ÜBERBLICK

Installation / Preparation

Für die Installation der Datenbank muss die eben beschriebene Ordnerstruktur erstellt werden. Der project folder wird im Ordner root gespeichert (Abbildung 5.2).

Abbildung 5.2: CA: Installation

Copy / save your files to the right place

KAPITEL 5. DIE ARBEIT MIT DEM CORPUSANALYSER

und in den Ordner 01_input_files/exmaralda kopiert und dort gesichert.

Abbildung 5.3: CA: copy and save files

Convert wav-files

In einem vierten Schritt werden die Wave-Dateien konvertiert und alle nicht verwendeten Passagen stumm geschaltet. Dies ist notwendig, weil nicht von allen Probandinnen und Probanden die Einverständniserklärung der Eltern für eine Veröffentlichung der Aufnahme vorliegt. Die Passagen werden auf diese Weise entfernt.

Abbildung 5.4: CA: convert wav-files

Um die nicht verwendeten Passagen zu überschreiben, muss folgendes Skript aufgerufen werden:

01_wav_convert/remove_non_used_area_in_wav_all.sh.

In einer Konsole/einem Terminal sollte nun zum project folder gewechselt werden. Wenn alle Schritte aus Abbildung 5.4 erfolgreich durchgeführt wurden, sollten die neu erstellten Wave-Dateien im folgenden Output-Directory zu finden sein: 02_generated_files/wav.

Create the SQLite3 Database

In diesem Abschnitt geht es um die Frage, wie die SQLite3-Datenbank erstellt wird. Auch dafür wird erneut mit einer Konsole/einem Terminal gearbeitet,
wobei zunächst in den project folder gewechselt wird. Von dort aus erfolgt ein Wechsel zum CA, von wo aus mit dem Befehl \texttt{.\!/create\_database.sh} die Datenbank (Abbildung 5.5) erstellt wird. Nach einem erfolgreichen Ausführen aller Schritte befindet sich die erstellte leere Datenbank im Ordner \texttt{02\_generated\_files/corpus.db}.

Abbildung 5.5: CA: create database

**Import exmaralda files to sqlite db**

Im vorigen Abschnitt wurde gezeigt, wie eine zunächst leere Datenbank erzeugt werden kann. Da zu einem großen Teil mit EXMARaLDA-Dateien gearbeitet wird, soll nun gezeigt werden, wie diese wiederum in die Datenbank importiert werden können.

Abbildung 5.6: CA: import exmaralda files

In einer Konsole/einem Terminal erfolgt dafür zunächst ein Wechsel in den project folder. Durch das Ausführen des Befehls \texttt{./db\_import\_all\_sh} \ldots/\texttt{01\_input\_files/exmaralda/} \ldots/\texttt{02\_generates\_files/corpus.db} befindet sich
94

KAPITEL 5. DIE ARBEIT MIT DEM CORPUSANALYSE


**Generate all results (step 7-12) from db**

Durch das Ausführen des Skriptes Generate all results (step 7-12) from db (Abbildung 5.7) können alle (folgend beschriebenen) Skripte zusammen ausgeführt werden, ohne dass diese einzeln aufgerufen werden müssen. Dies ist dann hilfreich, wenn tatsächlich alle Abfragen (Tag count results, Typetoken results, Typetoken Lemma results, Lemma Tag results, Lemma Bedeutung results und Typetoken Bedeutung results) getätigt werden sollen, da somit viele Arbeitsschritte gespart werden.

[Abbildung 5.7: CA: generate all results]

*This step depends on step 5.*

```
$ cd project
$ cd CorpusAnalyser/db_query_db
$ ./query_all.sh
```

This script will call step 7 to 12 automatically, it’s not necessary to do this manually. All results should be written now - you can stop here.

The results are written to subfolder under 02_db_query_results.

**Generate Tag count results from db**

5.1. DIE INHALTE IM ÜBERBLICK

durch ./db_query_all_tag_count_name.sh in separaten Ordnern unter dem jeweiligen Namen der Probandin/des Probanden gespeichert.

Abbildung 5.8: CA: generate tag count results

Typetoken results from db

Analog zu den Schritten im vorigen Abschnitt können alle Types mit den entsprechenden Tokens ausgezählt werden. Wie gewohnt muss in project folder den und anschließend in den CA gewechselt werden. Durch den Befehl ./db_query_all_typetoken werden alle Tokens (also jedes syntaktische Wort) und alle entsprechenden Types (alle unterschiedlichen Wörter in ihrer lemmatisierten Form) gezählt und im Zielverzeichnis mit dem dazugehörigen Kürzel der Probandin/des Probanden sowie in der jeweiligen Altersgruppe abgelegt. Eine solche Analyse ist hilfreich bei der Ermittlung der Type-Token-Relation. Wenn es um die Verteilung der POS-Tags geht, können interessante Passagen zudem mit konkreten Beispielen belegt werden.

Abbildung 5.9: CA: type-token results

Typetoken Lemma results from db

Durch das Ausführen der Schritte in Abbildung 5.10 werden zusätzlich zu den Tokens und Types je POS-Tag die Lemmata angezeigt. Nach dem Wechseln in den project folder und anschließend in den CA-Ordner muss dazu der Befehl ./db_query_all_typetoken_lemma.sh ausgeführt werden.
KAPITEL 5. DIE ARBEIT MIT DEM CORPUSANALYSER

Abbildung 5.10: CA: type-token-lemma results

Lemma Tags results from db


Abbildung 5.11: CA: lemma tag results

Lemma Bedeutung results from db

Ein ebenso häufig verwendeter Befehl ist ./db_query_all_lemma_bedeutung.sh. Durch das Ausführen werden alle Bedeutungen mit ihren dazugehörigen Lemmata in einzelne Ordner, sortiert nach den jeweiligen Altersgruppen, geschrieben.

Abbildung 5.12: CA: lemma bedeutung results
Typetoken_Bedeutung results from db

Der Befehl ./db_query_all_typetoken_lemma.sh bewirkt, dass alle Lemmata, die pro Bedeutungskategorie vorkommen, sortiert und gezählt werden. Die derartig sortierten Lemmata werden dann ihren Bedeutungskategorien zugeordnet und in CSV-Dateien abgelegt.

Abbildung 5.13: CA: typetoken bedeutung results

5.2 Die Funktionen des CorpusAnalysers

Im nun folgenden Abschnitt möchte ich die für die vorliegende Arbeit besonders wichtigen Funktionen bzw. Skripte des CA näher erläutern und die Zwecke, die damit für diese Arbeit verfolgt werden, darstellen. Der CA wurde, wie bereits erwähnt, für die Zwecke dieser Arbeit entwickelt. Dennoch, oder gerade deswegen, hält er eine Fülle von Funktionen bereit, die jedoch nicht immer zusammen genutzt werden müssen. Jede Unterfunktion kann einzeln oder in Kombination unterschiedliche Zwecke erfüllen. Ich werde auf jedes einzelne Skript eingehen, aber nur die für diese Arbeit relevanten Befehlszeilen ausführlich beschreiben.

5.2.1 01_wave_convert

5.2.2 02_create_database

Hinter dieser Funktion befinden sich die Skripte create_database.sh sowie db_table_structure.sqlite3. Ersteres erstellt eine SQLite3-Datenbank corpus.db und initialisiert die Tabellen, die im zweiten Skript erläutert sind. Wenn bereits eine Datenbank existiert, wird diese gelöscht und mit der neu erstellten Datenbank überschrieben. Das Skript db_table_structure.sqlite3 enthält die Tabellenstruktur der Datenbank. Im Wesentlichen gehören dazu: sprecher, aufnahme, textpostags, wavdatei. Abbildung 5.14 zeigt einen Ausschnitt des Skripts zur Erstellung der Untertabelle sprecher. Enthalten sind id (aufnahme-id), name (des kindes), kuerzel (des kindes), geschlecht, sprache, erstsprache, geburtstdatum. Die ID, die auch bei den dazugehörigen Aufnahmen als Fremdschlüssel (FOREIGN KEY) unter der sprecher_id vorkommt, wird dort eingetragen.

Abbildung 5.14: CA: table structure sprecher

Im Teil aufnahme (Abbildung 5.15) sind die Informationen id, sprecher_id und wavdatei_id enthalten. Ferner sind enthalten der esb_name, datum, kommentar und FOREIGN KEYs, die die IDs der entsprechenden sprecher- und wavdatei-Tabellen auf die Informationen sprecher-id und wavdatei-id verlinken. Die ID, welche auch bei der dazugehörigen textpostags-Tabelle als Fremdschlüssel unter der aufnahme-id vorkommt, wird dort eingetragen. Der esb_name ist die entsprechende EXMARaLDA-Datei. Das datum entspricht dem Aufnahmedatum der Audio-Datei.
5.2. DIE FUNKTIONEN DES CORPUSANALYSERS

Abbildung 5.15: CA: table structure aufnahme

CREATE TABLE aufnahme (id INTEGER PRIMARY KEY, 
sprecher_id INTEGER, 
wavdatei_id INTEGER, 
esb_name VARCHAR(88), 
datum DATE, 
kommentar VARCHAR(308), 
FOREIGN KEY(sprecher_id) REFERENCES sprecher(id), 
FOREIGN KEY(wavdatei_id) REFERENCES wavdatei(id));

**KAPITEL 5. DIE ARBEIT MIT DEM CORPUSANALYSER**

Abbildung 5.16: CA: table structure: textpostags

Abbildung 5.17 zeigt den Tabelleneintrag **wavdatei**. Dieser enthält die Punkte *id* und *name* (der jeweiligen Wave-Datei). Jeder Wave-Datei wird ein Name und eine ID in der Datenbank zugeordnet, welche auch bei den entsprechenden Aufnahmen als Fremdschlüssel (*FOREIGN KEY*) unter der *wavdatei-id* eingetragen wird.

Abbildung 5.17: CA: table structure wav_datei

**5.2.3 03_import_data**

Unter der Funktion **03_import_data** verbergen sich die Skripte `db_import_all.sh` und `db_import_exb_to_db3.pl`. Das Perl-skript liest EXMARaLDA-Dateien im exportierten XML-Format aus und zieht die relevanten Daten aus der jeweiligen Datei. Dann schreibt es die so erhaltenen Werte in die SQLite3-Datenbank. Wenn ein Eintrag in der Datenbank bereits existiert, wird der neue Eintrag ignoriert. Zum Erstellen einer aktualisierten Datenbank, muss das Skript `create_database.sh reinitialisiert werden. Das Skript `db_import_all.sh` wird verwendet, um EXMARaLDA-Dateien in die Datenbank zu importieren. Durch diesen Schritt werden alle EXMARaLDA-Dateien (.exb) aus einem gegebenen Ordner gelesen und der Inhalt der Dateien in die Datenbank geschrieben. Der Inhalt der EXMARaLDA-Dateien wird durch das Unterskript `db_import_exb_to_db3.pl` in SQLite3 importiert und exportiert.
5.2. DIE FUNKTIONEN DES CORPUSANALYSERS

5.2.4 04_query_db


5.2.4.1 01_tag_count

Hinter dieser Funktion verbergen sich vier Skripte, deren Funktionen im Folgenden näher erläutert werden.

db_query_tag_count.sh: Dieses Skript fragt die POS-Tags zwischen den Zeitstempeln in der Datenbank ab und schreibt die Ergebnisse in eine CSV-Datei. Anhand des nun folgenden Ausschnitts aus dem Skript sollen die relevanten Befehle im Hinblick auf die Zwecke dieser Arbeit erläutert werden:

**Skript: db_query_tag_count.sh**

```bash
DB=../../../02_generated_files/corpus.db
if [ -z $1 ] || [ -z $2 ] || [ -z $3 ] || [ -z $4 ] || [ ! -d $4 ]; then
echo "usage $0 daymin daymax postag outputdir"
exit
fi
DAYMIN=$1
DAYMAX=$2
POSTAG="$3"
OUTPUTDIR=$4

echo -n 'SELECT sprecher.name, sprecher.geburtsdatum, ' > query.sql
echo -n 'aufnahme.datum, julianday(aufnahme.datum) ' >> query.sql
echo -n '-julianday(sprecher.geburtsdatum), ' >> query.sql
echo -n 'aufnahme.esb_name, textpostags.wort, ' >> query.sql
echo -n 'textpostags.postag, textpostags.id ' >> query.sql
echo -n 'FROM sprecher,aufnahme,textpostags ' >> query.sql
echo -n 'WHERE julianday(aufnahme.datum) - ' >> query.sql
echo -n 'julianday(sprecher.geburtsdatum) > ' >> query.sql
echo -n $DAYMIN >> query.sql
echo -n ' AND julianday(aufnahme.datum) ' >> query.sql
echo -n '- julianday(sprecher.geburtsdatum) < ' >> query.sql
echo -n $DAYMAX >> query.sql
echo -n ' AND textpostags.postag = "' >> query.sql
echo -n $POSTAG >> query.sql
echo -n '" AND aufnahme.sprecher_id = sprecher.id AND' >> query.sql
```

```bash
de_query_tag_count.sh: Dieses Skript fragt die POS-Tags zwischen den Zeitstempeln in der Datenbank ab und schreibt die Ergebnisse in eine CSV-Datei. Anhand des nun folgenden Ausschnitts aus dem Skript sollen die relevanten Befehle im Hinblick auf die Zwecke dieser Arbeit erläutert werden:

**Skript: db_query_tag_count.sh**

```bash
DB=../../../02_generated_files/corpus.db
if [ -z $1 ] || [ -z $2 ] || [ -z $3 ] || [ -z $4 ] || [ ! -d $4 ]; then
echo "usage $0 daymin daymax postag outputdir"
exit
fi
DAYMIN=$1
DAYMAX=$2
POSTAG="$3"
OUTPUTDIR=$4

echo -n 'SELECT sprecher.name, sprecher.geburtsdatum, ' > query.sql
echo -n 'aufnahme.datum, julianday(aufnahme.datum) ' >> query.sql
echo -n '-julianday(sprecher.geburtsdatum), ' >> query.sql
echo -n 'aufnahme.esb_name, textpostags.wort, ' >> query.sql
echo -n 'textpostags.postag, textpostags.id ' >> query.sql
echo -n 'FROM sprecher,aufnahme,textpostags ' >> query.sql
echo -n 'WHERE julianday(aufnahme.datum) - ' >> query.sql
echo -n 'julianday(sprecher.geburtsdatum) > ' >> query.sql
echo -n $DAYMIN >> query.sql
echo -n ' AND julianday(aufnahme.datum) ' >> query.sql
echo -n '- julianday(sprecher.geburtsdatum) < ' >> query.sql
echo -n $DAYMAX >> query.sql
echo -n ' AND textpostags.postag = "' >> query.sql
echo -n $POSTAG >> query.sql
echo -n '" AND aufnahme.sprecher_id = sprecher.id AND' >> query.sql
```
Durch die Befehle DAYMIN und DAYMAX wird die Zeitspanne eingegrenzt, innerhalb welcher die POS-Tags gezählt und geordnet werden sollen. Ferner wird der Name und das Geburts- sowie Aufnahmedatum jeder Sprecherin und jedes Sprechers zu jedem Zeitpunkt benötigt. Durch die Berechnung in Juliandays\[^4\] können die Zeitangaben in Tagen gemacht werden. Das ist sehr sinnvoll, weil nicht alle Monate die gleiche Anzahl an Tagen aufweisen. Zudem rechnen Programme häufig in Tagen, wodurch weniger Ungenauigkeiten entstehen können. Durch die Befehlszeilen

```
echo -n 'SELECT sprecher.name, sprecher.geburtsdatum, ' > query.sql
echo -n 'aufnahme.datum, julianday(aufnahme.datum) ' >> query.sql
echo -n '-julianday(sprecher.geburtsdatum), ' >> query.sql
```


```
```


Wendet man sich hingegen gegen POSTAGS je Kind untersucht, dann müssen alle POS-Tags aufgeführt werden. Hinter OUTPUT_DIR= wird der Pfad für das Ausgabeverzeichnis angegeben, also der Ort, an dem der Ordnern mit den Ergebnissen liegen soll. Im Falle der vorliegenden Arbeit liegt das OUTPUT_DIR unter ../../../03 db_query_results/tag_count.

Skript: db_query_all_tag_count.sh

```
DAYMINMAX="1281:1495 1496:1708 1709:1983"
POSTAGS="ADJA ADJD ADV APPR APPRART ART CARD FM ITJ KOKOM KON KOUS NE NN PAV"
POSTAGS="$POSTAGS PDAT PDS PIAT PIDAT PIS PPPER PPOSAT PPOSS PRELAT PRELS PRF"
POSTAGS="$POSTAGS PTKA PTKANT PTKNEG PTKVZ PWAT PWAV PWS VAFIN VAIMP"
POSTAGS="$POSTAGS VAINF VAPP VMFIN VMPP VVFIN VVMP VVVIZU VVPP XY"
OUTPUT_DIR=../../../03_db_query_results/tag_count
for d in $DAYMINMAX ; do
  if [[ $d =~ ^(.*)\:(.*)$ ]]; then
    DAYMIN='echo ${BASH_REMATCH[1]}'
    DAYMAX='echo ${BASH_REMATCH[2]}'
    for i in $POSTAGS ; do
      if [ ! -e $OUTPUT_DIR/${DAYMIN}_${DAYMAX} ]; then
        mkdir -p $OUTPUT_DIR/${DAYMIN}_${DAYMAX}
      fi
      echo $DAYMIN $DAYMAX $i $OUTPUT_DIR/${DAYMIN}_${DAYMAX}
      ./db_query_tag_count.sh $DAYMIN $DAYMAX $i
    done
  fi
done
```

db_query_tag_count_name.sh: Das Skript arbeitet im Wesentlichen wie das Skript db_query_tag_count.sh. Der Unterschied ist, dass die POS-Tags nicht nur nach der Altersgruppe sortiert werden, sondern dass auch eine Zuordnung zur jeweiligen Sprecherin/zum jeweiligen Sprecher stattfindet. Die Ergebnisse werden wie gewohnt in eine CSV-Datei geschrieben. Die
KAPITEL 5. DIE ARBEIT MIT DEM CORPUSANALYSER

Befehlszeilen des vorliegenden Skripts entsprechen größtenteils denen des Skripts `db_query_tag_count.sh`. Zusätzlich existieren die Befehlszeilen

```bash
echo -n "$NAME" >> query.sql
echo -n "' AND aufnahme.sprecher_id = sprecher.id AND" >> query.sql
echo -n " textpostags.aufnahme_id = aufnahme.id;" >> query.sql
```

Durch das Einfügen dieser Befehle wird erreicht, dass sich die Abfrage speziell auf eine/n Sprecher/in bezieht (mehr dazu im nächsten Skript).

**Skript: db_query_tag_count_name.sh**

```bash
DB=./././02_generated_files/corpus.db
if [ -z "$1" || -z "$2" || -z "$3" || -z "$4" || -z "$5" ] then
  echo "usage $0 daymin daymax postag outputdir name"
  exit
fi
DAYMIN=$1
DAYMAX=$2
POSTAG="$3"
OUTPUTDIR="$4"
NAME="$5"
```

```bash
echo -n 'SELECT sprecher.name, sprecher.geburtsdatum, ' > query.sql
echo -n 'aufnahme.datum, julianday(aufnahme.datum) ' >> query.sql
echo -n '-julianday(sprecher.geburtsdatum),' >> query.sql
echo -n 'textpostags.wort, ' >> query.sql
echo -n 'textpostags.postag, textpostags.id ' >> query.sql
echo -n ' FROM sprecher,aufnahme,textpostags ' >> query.sql
echo -n 'WHERE julianday(aufnahme.datum) - ' >> query.sql
echo -n 'julianday(sprecher.geburtsdatum) > ' >> query.sql
echo -n $DAYMIN >> query.sql
echo -n ' AND julianday(aufnahme.datum) ' >> query.sql
echo -n '- julianday(sprecher.geburtsdatum) < ' >> query.sql
echo -n $DAYMAX >> query.sql
echo -n ' AND textpostags.postag = "' >> query.sql
echo -n $POSTAG >> query.sql
echo -n '" AND sprecher.kuerzel = "' >> query.sql
echo -n $NAME >> query.sql
echo -n '" AND aufnahme.sprecher_id = sprecher.id AND' >> query.sql
echo -n ' textpostags.aufnahme_id = aufnahme.id;" >> query.sql
COUNT=`cat query.sql | sqlite3 $DB | wc -l`
if [ $COUNT -gt 0 ] ; then
  echo "$COUNT datarows found: "
  echo "create file $OUTPUTDIR/output_${POSTAG}_${DAYMIN}_${DAYMAX}.csv"
cat query.sql | sqlite3 $DB > "$OUTPUTDIR/output_${POSTAG}_${DAYMIN}_${DAYMAX}.csv"
else
  echo "$COUNT datarows found: "
fi
rm query.sql
```

db_query_all_tag_count_name.sh: Dieses Skript fragt die POS-Tags zwischen den Zeitstempeln unter Berücksichtigung des jeweilig gewählten Namens in Kurzform aus der Datenbank ab. Die Resultate werden wie gewohnt in
5.2. DIE FUNKTIONEN DES CORPUSANALYSERS

eine CSV-Datei geschrieben. Die Abfrage aus der Datenbank wird durch
das Unterskript db_query_tag_count.sh initiiert. Die Variablen \textit{DAYMINMAX},
\textit{POSTAGS}, \textit{NAMES} und \textit{OUTPUT_DIR} können und sollten je nach dem Ziel
der jeweiligen Abfrage eigenständig angepasst werden. Die Befehlszeilen
entsprechen denen des Skripts \textit{db_query_all_tag_count.sh} mit dem Zusatz,
dass hier auch die einzelnen Sprecher/innen ausgewählt werden können.
Wenn eine derartige Abfrage mit allen Sprecherkürzeln ausgeführt wird, erhält
man alle POS-Tags in Zuordnung zur jeweiligen Sprecherin/zum jeweiligen
Sprecher, sortiert nach der Altersgruppe als Ergebnis. Der entsprechende
Überordner kann zum Beispiel \texttt{all_tag_count_name} heißen. Als Unterordner
können idealerweise drei Unterordner für je eine Altersgruppe angelegt
werden. Hinter jedem dieser drei Ordner sollten in logischer Konsequenz
jeweils Ordner auftauchen, die einer bestimmten Sprecherin/einem
bestimmten Sprecher zuordenbar sind. In diesen Ordnern werden die
jeweilsigen POS-Tags in einer CSV-Datei abgelegt, die von der jeweiligen
Sprecherin/vom jeweiligen Sprecher in der betreffenden Altersgruppe
geäußert wurden bzw. dem jeweiligen syntaktischen Wort zugeordnet wurden.
Dies stellt nur eine mögliche Ordnerruktur dar, die selbstverständlich je
nach Bedarf geändert werden kann. Ferner ist es vorstellbar, dass nicht
immer alle POS-Tags je Sprecher/in und Altersgruppe von der Datenbank
abgefragt werden. Sollen beispielsweise nur die Verben analysiert werden,
die der Sprecher MK im Zeitraum 1496 bis 1708 verwendet hat, dann müsste
die Zeile \texttt{DAYMINMAX=“1496:1708“} auf diese Weise eingegrenzt werden.
Damit werden lediglich POS-Tags abgefragt, die in diesem Zeitraum geäußert
wurden. Durch Verändern der nächsten Zeile in
\begin{verbatim}
POSTAGS="VAFIN VAIMP VAINF VAPP VMFIN VMPP VVFIN VVIMP VVINF VVIZU VVPP"
\end{verbatim}

werden ausschließlich alle Verben in der genannten Altersgruppe
berücksichtigt. Da jedoch nicht alle Sprecher/innen in Betracht gezogen
werden, sondern nur MK, muss der Befehl \textit{NAMES} geändert werden in
\textit{NAMES=“MK“}: Ein anderes Szenario wäre es beispielsweise herauszufinden,
wie viele Eigennamen (NE) in allen Altersgruppen von jedem einzelnen Kind
geäußert werden. Dann müsste folgende Befehlszeile eingefügt werden:
\begin{verbatim}
DAYMINMAX="1281:1495 1496:1708 1709:1983"
\end{verbatim}

Für die Abfrage der POS-Tags genügt in diesem Falle \textit{POSTAGS=“NE“}; die
Zeile der Sprecher/innen muss wieder alle Namen enthalten, also
KAPITEL 5. DIE ARBEIT MIT DEM CORPUSANALYSER

KAPITEL 5. DIE ARBEIT MIT DEM CORPUSANALYSER

Das vollständige Skript zum Befehl `db_query_all_tag_count_name.sh` sieht folgendermaßen aus:

**Skript: db_query_all_tag_count_name.sh**

```bash
DAYMINMAX="1281:1495 1496:1708 1709:1983"
POSTAGS="ADJA ADJD ADV APPR APPIART ART CARD FM ITJ KOKOM KOUS NE NN PAV"
POSTAGS="$POSTAGS PDAT PDS PIAT PIS PER PPOSAT PRELAT PRELS PRF"
POSTAGS="$POSTAGS PTKA PTKANT PTKNEG PTKVZ PTKZU PWAT PWAV PWS VAFIN VAIMP"
POSTAGS="$POSTAGS VAINF VAPP VMFIN VMINF VMPP VVFIN VVIMP VVINF VVIZU VVPP XY"
NAMES="av mm js rd mk leo lar so lua jk ll ma"
OUTPUT_DIR=../../../03_db_query_results/tag_count
for n in $NAMES ; do
  for d in $DAYMINMAX ; do
    if [[ $d =~ ^(.*)\((.*)\)$ ]] ; then
      DAYMIN='echo ${BASH_REMATCH[1]}'
      DAYMAX='echo ${BASH_REMATCH[2]}'
      for i in $POSTAGS ; do
        if [ ! -e $OUTPUT_DIR/${DAYMIN}_${DAYMAX}/$n ] ; then
          mkdir -p $OUTPUT_DIR/${DAYMIN}_${DAYMAX}/$n
          echo -n "$DAYMIN $DAYMAX $i \\
          ./db_query_tag_count_name.sh $DAYMIN $DAYMAX $i \\
          $OUTPUT_DIR/${DAYMIN}_${DAYMAX}/$n $n
        done
        rmdir –ignore-fail-on-non-empty $OUTPUT_DIR/${DAYMIN}_${DAYMAX}/$n
      done
    fi
  done
for n in $NAMES ; do
```

5.2.4.2 02_typetoken

**db_query_typetoken.sh:** Mit diesem Skript werden die Types und die dazugehörigen Tokens zwischen den Zeitstempeln der EXMARaLDA-Datei abgefragt und in eine CSV-Datei geschrieben. Des Weiteren werden alle vorkommenden Wörter je POS-Tag und Altersgruppe in einer separaten Wortliste aufgeführt. Was sich recht simpel anhört, ist spätestens bei der Betrachtung des dazugehörigen Skripts eine lange Abfolge von Befehlszeilen. Der Befehl, mit dem die eben genannte Wortliste erstellt wird, ist `SELECT textpostags.wort;` man wählt also das jeweilige POS-Tag mit dem dazugehörigen Wort (nicht Lemma). Benötigt werden dafür die Informationen `sprecher, aufnahme, textpostags` aus der Datenbank. Außerdem soll erneut die Zeit eingegrenzt und die Ergebnisse geordnet in den drei Altersgruppen anzeigt werden. Durch die Befehlszeilen:

```bash
echo -n `WHERE julianday(aufnahme.datum)` >> query.sql`
5.2. DIE FUNKTIONEN DES CORPUSANALYSERS

```
echo -n `-julianday(sprecher.geburtsdatum) > ' >> query.sql
echo -n `$DAYMIN` >> query.sql
echo -n `-AND julianday(aufnahme.datum)` `>` query.sql
echo -n `-julianday(sprecher.geburtsdatum) < ' >> query.sql
echo -n `$DAYMAX` >> query.sql
echo -n `-AND textpostags.postag = "` >> query.sql
echo -n `$POSTAG` >> query.sql
echo -n `-AND aufnahme.sprecher_id = sprecher.id` `>` query.sql
echo -n `-AND textpostags.aufnahme_id = aufnahme.id;` `>` query.sql
```

erhält die Datenbank die Information, dass die POS-Tags jeder Sprecherin/jedes Sprechers zu jedem Aufnahmedatum abgefragt werden sollen. Wichtig im vollständigen Skript ist die Variable `WORTLIST` und die dazugehörige vollständige Befehlszeile

```
WORTLIST='cat $OUTPUTDIR/output_wortlist_${POSTAG}_${DAYMIN}_${DAYMAX}.txt'
```

Dadurch werden alle vorkommenden Wörter in einer Liste angezeigt.

Abbildung 5.18: CA: Wortliste 1281 bis 1495 ART

KAPITEL 5. DIE ARBEIT MIT DEM CORPUSANALYSER

Skript: db_query_typetoken.sh

```bash
DB=../../../..02_generated_files/corpus.db
if [-z $1] || [-z $2] || [-z $3] || [-z $4] || [-d $4]; then
    echo "usage $0 daymin daymax postag outputdir"
    exit
fi
DAYMIN=$1
DAYMAX=$2
POSTAG="$3"
OUTPUTDIR="$4"
echo -n 'SELECT textpostags.wort ' > query.sql
echo -n 'FROM sprecher,aufnahme,textpostags ' >> query.sql
echo -n 'WHERE julianday(aufnahme.datum) ' >> query.sql
echo -n '- julianday(sprecher.geburtsdatum) > ' >> query.sql
echo -n $DAYMIN >> query.sql
echo -n ' AND julianday(aufnahme.datum) ' >> query.sql
echo -n '- julianday(sprecher.geburtsdatum) < ' >> query.sql
echo -n $DAYMAX >> query.sql
echo -n ' AND textpostags.postag = "' >> query.sql
echo -n $POSTAG >> query.sql
echo -n '" AND aufnahme.sprecher_id = sprecher.id ' >> query.sql
echo -n 'AND textpostags.aufnahme_id = aufnahme.id;' >> query.sql
```
```
cat query.sql | sqlite3 $DB | sort | uniq > 
$OUTPUTDIR/output_wortlist_${POSTAG}_${DAYMIN}_${DAYMAX}.txt
WORTLIST='cat $OUTPUTDIR/output_wortlist_${POSTAG}_${DAYMIN}_${DAYMAX}.txt'
CNT=0
rm -f $OUTPUTDIR/output_typetoken_count_${POSTAG}_${DAYMIN}_${DAYMAX}.csv
for j in $WORTLIST ; do
    echo -n 'SELECT sprecher.name, sprecher.geburtsdatum, ' > query2.sql
    echo -n 'aufnahme.datum, julianday(aufnahme.datum) ' >> query2.sql
    echo -n '-julianday(sprecher.geburtsdatum), ' >> query2.sql
    echo -n 'aufnahme.esb_name, textpostags.wort, ' >> query2.sql
    echo -n 'textpostags.postag, textpostags.id FROM ' >> query2.sql
    echo -n 'sprecher,aufnahme,textpostags WHERE ' >> query2.sql
    echo -n 'julianday(aufnahme.datum) ' >> query2.sql
    echo -n '- julianday(sprecher.geburtsdatum) > ' >> query2.sql
    echo -n $DAYMIN >> query2.sql
    echo -n ' AND julianday(aufnahme.datum) ' >> query2.sql
    echo -n '- julianday(sprecher.geburtsdatum) < ' >> query2.sql
    echo -n $DAYMAX >> query2.sql
    echo -n ' AND textpostags.postag = "' >> query2.sql
    echo -n $j' >> query2.sql
    echo ' AND aufnahme.sprecher_id = sprecher.id AND ' >> query2.sql
    echo -n 'textpostags.aufnahme_id = aufnahme.id;' >> query2.sql
    echo -n '"' >> query2.sql
    cat query2.sql | sqlite3 $DB | wc -l
    WORDCNT='cat query2.sql | sqlite3 $DB | wc -l'
    WORDCNTSUM=$((WORDCNTSUM + WORDCNT))
    CNT=$((CNT + 1))
    echo $WORDCNT >>
$OUTPUTDIR/output_typetoken_count_${POSTAG}_${DAYMIN}_${DAYMAX}.csv
done
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
5.2. DIE FUNKTIONEN DES CORPUSANALYSERS

rm -f query.sql
rm -f query2.sql

**db_query_all_typetoken.sh**: Dieses Skript fragt ebenso die Types und Tokens zwischen den einzelnen Zeitstempeln in der EXMARaLDA-Datei aus der Datenbank ab und schreibt die Ergebnisse in eine CSV-Datei. Die Abfrage der Datenbank an sich wird durch das oben angeführte Unterkript **db_query_typetoken.sh** ausgeführt. In diesem Skript wird nun einerseits die Altersgruppe, für die die Types und Tokens erfragt werden, eingegrenzt und andererseits die POS-Tags festgelegt, für die die entsprechenden Types und Tokens erfragt werden sollen. Es ist sehr sinnvoll, wenn nicht gar unerlässlich, an dieser Stelle im Skript die Types und die Tokens an Hand der POS-Tags zu identifizieren. Anderenfalls käme es an einigen Stellen der Analyse zu zweideutigen Ergebnissen und die Wörter sowie die Lemmata würden trotz ihrer Formgleichheit unterschiedlichen POS-Tags zugeordnet. In diesem Zusammenhang soll das Wort *schwimmen* mit dem dazugehörigen Lemma *schwimmen* betrachtet werden. In einer Aussage könnte es heißen „Wir schwimmen im See“ und in einer nächsten Aussage „Wir gehen heute mit dem Kindergarten zum Schwimmen“. In beiden Fällen wäre das Lemma *schwimmen*, aber die POS-Tags unterscheiden sich. Demnach ist *schwimmen* in der ersten Aussage ein finites Vollverb (VVFIN), *Schwimmen* in der zweiten Aussage müsste streng genommen als Nomen (NN) getaggt werden (siehe dazu aber den Eintrag unter dem Stichwort *Nomen* in Abschnitt 4.1.7). Das Skript geht bei seiner Suche aus von den POS-Tags und listet am Ende sortiert nach den POS-Tags alle Types mit den dazugehörigen Tokens auf. Eine Möglichkeit wäre es, nur die Types und Tokens abzfragen, die in der Kategorie VVINF und in der Altersgruppe 1709:1983 vorkommen. In diesem Fall stünde in der ersten Zeile **DAYMINMAX=“1709:1983“** und in der folgenden Zeile **POSTAGS=“VVINF“.** Das Programm benötigt also die Information der POS-Tags für die Ausgabe der Types und Tokens, weil dies in den EXMARaLDA-Dateien sowie in der Datenbank so organisiert ist.

**Skript: db_query_all_typetoken.sh**

```bash
POSTAGS=“ADJA ADJD ADV APPR APPRART ART CARD FM ITJ KOKOM KON KOUS NE NN PAV“
POSTAGS=“$POSTAGS PDAT PDS PIAT PID PPER PPOSAT PPOSS PRELAT PRELS PRF“
POSTAGS=“$POSTAGS PTKA PTKANT PTKNEG PTKVZ PWAT PWAV PWS VAFIN VAIMP“
POSTAGS=“$POSTAGS VAINF VAPP VMFIN VMINF VMPP VVFIN VVIMP VVIZU VVPP XY“
OUTPUT_DIR=../../../03_db_query_results/typetoken
for d in $DAYMINMAX ; do
  if [[ $d =~ ^(.*)$ ]]; then
    DAYMIN=`echo $BASH_REMATCH[1]`
    DAYMAX=`echo $BASH_REMATCH[2]`
  fi
  for i in $POSTAGS ; do
    ...
  done
done
```
5.2.4.3 03_typetoken_lemma

db_query_typetoken_lemma.sh: An dieser Stelle werden die Types und die Tokens abgefragt; die Lemmata werden mit den dazugehörigen Häufigkeiten in einer CSV-Datei je POS-Tag ausgegeben. Da sich der erste Teil des Skripts nur unwesentlich vom Skript db_query_typetoken unterscheidet, soll dieser hier nicht näher beschrieben werden. Die Funktion Erstellen einer Wortliste entspricht der des eben erwähnten Skripts db_query_typetoken mit dem Unterschied, dass es sich hier um die Lemmata (SELECT textpostags.lemma) und nicht die tatsächlich geäußerten Wörter (SELECT textpostags.wort) handelt. Im unteren Teil des Skripts ist erkennbar, dass unter anderem die Befehlszeile "AND textpostags.lemma = '1'" an Stelle von "AND textpostags.wort = '1'" enthalten ist. Diese zweite Abfrage (query2.sql) ist insgesamt weniger komplex als die zweite Abfrage (query2.sql) im Skript db_query_typetoken, weil an dieser Stelle nur die Lemmata mit den dazugehörigen POS-Tags relevant sind, nicht aber alle geäußerten Wörter zu jedem Lemma. Zusätzlich ermöglicht das Skript db_query_typetoken_lemma.sh eine Types-zu-Tokens-Analyse jedes einzelnen Tags sowie der zusammengefassten POS-Tags. Als Ergebnis erhält man die prozentualen Anteile, wie sie in Kapitel 6 vorzufinden sind.

Skript: db_query_typetoken_lemma.sh

if [ -e $OUTPUT_DIR/${DAYMIN}_${DAYMAX} ] ; then
  mkdir -p $OUTPUT_DIR/${DAYMIN}_${DAYMAX}
fi

for i in $(echo $POSTAGLIST | tr "+" " ") ; do
  echo $DAYMIN $DAYMAX $i $OUTPUT_DIR/${DAYMIN}_${DAYMAX}
  ./db_query_typetoken.sh $DAYMIN $DAYMAX $i \ $OUTPUT_DIR/${DAYMIN}_${DAYMAX}
done

if [ -z $1 ] || [ -z $2 ] || [ -z $3 ] || [ -z $4 ] || [ ! -d $4 ] ; then
  echo "usage $0 daymin daymax postag outputdir (hint: postag can also be a list like AAA+BBB...)
exit
fi

DAYMIN=$1
DAYMAX=$2
POSTAGLIST="$3"
OUTPUTDIR="$4"
POSTAG_QUERY="AND ( "
for i in $($echo $POSTAGLIST | tr "+" " ") ; do
  POSTAG_QUERY+=$POSTAG_QUERY[POSTAG_QUERY]textpostags.postag = "
  POSTAG_QUERY+=\$i
  POSTAG_QUERY+=" OR "

DB=../../../02_generated_files/corpus.db

if [ -z $1 ] || [ -z $2 ] || [ -z $3 ] || [ -z $4 ] || [ ! -d $4 ] ; then
  echo "usage $0 daymin daymax postag outputdir (hint: postag can also be a list like AAA+BBB...)
  exit
fi

DAYMIN=$1
DAYMAX=$2
POSTAGLIST="$3"
OUTPUTDIR="$4"
POSTAG_QUERY="AND ( "
for i in $($echo $POSTAGLIST | tr "+" " ") ; do
  POSTAG_QUERY+=$POSTAG_QUERY[POSTAG_QUERY]textpostags.postag = "
  POSTAG_QUERY+=\$i
  POSTAG_QUERY+=" OR "

5.2.4.3 03_typetoken_lemma
5.2. DIE FUNKTIONEN DES CORPUSANALYSERS

done
POSTAG_QUERY=${POSTAG_QUERY}'0)' 
 echo -n "SELECT textpostags.lemma > query.sql"
 echo -n "FROM sprecher.aufnahme,textpostags >>> query.sql"
 echo -n "WHERE julianday(aufnahme.datum) >> query.sql"
 echo -n "AND julianday(sprecher.geburtsdatum) > ' >> query.sql"
 echo -n "$DAYMIN >> query.sql"
 echo -n "AND julianday(aufnahme.datum) >> query.sql"
 echo -n "AND julianday(sprecher.geburtsdatum) < ' >> query.sql"
 echo -n "$DAYMAX >> query.sql echo -n "AND textpostags.aufnahme_id = aufnahme.id;" >> query.sql"
 echo -n "$POSTAG_QUERY => query.sql"
 echo -n "AND aufnahme.sprecher_id = sprecher.id >>> query.sql"
 echo -n "AND textpostags.aufnahme_id = aufnahme.id; >> query.sql"
 COUNT='cat query.sql | sqlite3 $DB | wc -l'
 echo ", $COUNT entries found" if [ $COUNT -gt 0 ]; then
cat query.sql | sqlite3 $DB | sort | uniq > \
$OUTPUTDIR/output_wortlist_${POSTAGLIST}_${DAYMIN}_${DAYMAX}.txt
WORTLIST='cat $OUTPUTDIR/output_wortlist_${POSTAGLIST}_${DAYMIN}_${DAYMAX}.txt'
CNT=0
WORDCNTSUM=0
for j in $WORTLIST ; do
 echo -n "SELECT textpostags.postag > query2.sql"
 echo -n "FROM sprecher.aufnahme,textpostags >>> query2.sql"
 echo -n "WHERE julianday(aufnahme.datum) >> query2.sql"
 echo -n "AND julianday(sprecher.geburtsdatum) > ' >> query2.sql"
 echo -n "$DAYMIN >> query2.sql"
 echo -n "AND julianday(aufnahme.datum) >> query2.sql"
 echo -n "AND julianday(sprecher.geburtsdatum) < ' >> query2.sql"
 echo -n "$DAYMAX >> query2.sql echo -n "AND textpostags.lemma = \" >> query2.sql"
 echo -n $j\" >> query2.sql
 echo -n "AND aufnahme.sprecher_id = sprecher.id >>> query2.sql"
 echo -n "AND textpostags.aufnahme_id = aufnahme.id; >> query2.sql"
 echo -n "$j\" >> query2.sql
 WORDCNT='cat query2.sql | sqlite3 $DB | wc -l'
 WORDCNTSUM=$((WORDCNTSUM + WORDCNT))
 CNT=$((CNT + 1))
echo $WORDCNT >>
$OUTPUTDIR/output_typetoken_lemma_${POSTAGLIST}_${DAYMIN}_${DAYMAX}.csv
done
echo "—-|—-" "$OUTPUTDIR/output_typetoken_lemma_${POSTAGLIST}_${DAYMIN}_${DAYMAX}.csv
echo "$CNT|$WORDCNTSUM" "$OUTPUTDIR/output_typetoken_lemma_summary_${DAYMIN}_${DAYMAX}.csv
rm -f query.sql
fi
rm -f query.sql

db_query_all_typetoken_lemma.sh: Dieses Skript fragt die Types und
die Tokens zwischen den einzelnen Zeitstempeln in der jeweiligen
EXMARaLDA-Datei aus der Datenbank ab und schreibt die Ergebnisse in
KAPITEL 5. DIE ARBEIT MIT DEM CORPUSANALYSE

eine CSV-Datei. Die Abfrage der Datenbank an sich wird durch das oben angeführte Unterskript `db_query_typetoken_lemma.sh` ausgeführt. Es ist mit dieser Abfrage möglich, alle oder nur einige POS-Tags abzufragen und dadurch nur die Lemmata ausgeben zu lassen, die den jeweiligen POS-Tags zugeordnet sind. Weiterhin ist es möglich, (wie in allen „...sh“-Skripten) die Altersgruppen einzugrenzen oder aber alle gleichzeitig abzufragen.

Skript: `db_query_all_typetoken_lemma.sh`

```
DAYMINMAX="1281:1495 1496:1708 1709:1983"
POSTAGS="\$POSTAGS PDAT PDS PIAT PIDAT PIS PPER PPOSAT PRF"
POSTAGS="$POSTAGS TPKA PTKANT PTKNEG PTKVZ PTKZU PWAT PWA IVFIN VIFN VIFN VIZU VPP YX"
POSTAGS="$POSTAGS ADJA ADJD ADV APPR ART CARD+CARD"
POSTAGS="$POSTAGS FM+FM ITJ+ITJ KOKOM+KOUS NE+NN"
POSTAGS="$POSTAGS PDAT+PDS+PIAT+PIDAT+PIS PPER+PPER PPOSAT+PPOSS PRELS+PRF"
POSTAGS="$POSTAGS TPKA+PTKANT+PTKNEG+PTKVZ+PTKZU PWAT+PWA IVFIN+VIFN+VIZU+VPP YX"
OUTPUT_DIR=../../../03_db_query_results/typetoken_lemma
for d in $DAYMINMAX ; do
    if [[ $d =~ ^(.*):(.*)$ ]]; then
        DAYMIN='echo ${BASH_REMATCH[1]}'
        DAYMAX='echo ${BASH_REMATCH[2]}'
        rm -f $OUTPUT_DIR/output_typetoken_lemma_summary_${DAYMIN}_${DAYMAX}.csv
        for i in $POSTAGS ; do
            if [ ! -e $OUTPUT_DIR/${DAYMIN}_${DAYMAX} ] ; then
                mkdir -p $OUTPUT_DIR/${DAYMIN}_${DAYMAX}
            fi
            echo -n $DAYMIN $DAYMAX $i $OUTPUT_DIR/${DAYMIN}_${DAYMAX}
        done
    fi
done
CNTSUM=0
WORDCNTSUM=0
for i in $FILE ; do
    if [ ! -f $OUTPUT_DIR/output_typetoken_lemma_summary_${DAYMIN}_${DAYMAX}.csv ] ; then
        FILE='cat $OUTPUT_DIR/output_typetoken_lemma_summary_${DAYMIN}_${DAYMAX}.csv'
    for i in $FILE ; do
        if [[ $i =~ ^(.*)\|.*\|.*\|.*\|(.*)\|(.*)$ ]] ; then
            TAG='echo ${BASH_REMATCH[1]}'
            CNT='echo ${BASH_REMATCH[2]}'
            WORDCNT='echo ${BASH_REMATCH[3]}'
            if [[ $TAG =~ .\+.* ]] ; then
                CNTSUM=$((CNTSUM + CNT))
                WORDCNTSUM=$((WORDCNTSUM + WORDCNT))
            fi
        fi
    done
    if [ $CNTSUM -gt 0 ] && [ $WORDCNTSUM -gt 0 ] ; then
        FILE='cat $OUTPUT_DIR/output_typetoken_lemma_summary_${DAYMIN}_${DAYMAX}.csv'
    for i in $FILE ; do
        if [[ $i =~ ^(.*)\|.*\|.*\|.*\|(.*)\|(.*)$ ]] ; then
            CNT='echo ${BASH_REMATCH[1]}'
            WORDCNT='echo ${BASH_REMATCH[2]}'
            WORDPERCENT='echo "scale=5; \$WORDCNT*100/$WORDCNTSUM" | bc | sed -e "s/\./,/g"'
```

5.2. DIE FUNKTIONEN DES CORPUSANALYSERS


**Skript: db_query_typetoken_lemma_name.sh**

```
DB=./../02_generated_files/corpus.db
  echo "usage $0 daymin daymax postag outputdir (hint: postag can also be a list like AAA+BBB...)
  exit
fi
DAYMIN=$1
DAYMAX=$2
```
KAPITEL 5. DIE ARBEIT MIT DEM CORPUSANALYSE

POSTAGLIST="$3"
OUTPUTDIR=$4
NAME=$5
POSTAG_QUERY="AND ( "
for i in $(echo $POSTAGLIST | tr "+" "\n") ; do
    POSTAG_QUERY=${POSTAG_QUERY}'textpostags.postag = "'
    POSTAG_QUERY=${POSTAG_QUERY}${i}
    POSTAG_QUERY=${POSTAG_QUERY}'" '
    POSTAG_QUERY=${POSTAG_QUERY}'OR "
done
POSTAG_QUERY=${POSTAG_QUERY}'0 ) '

bash

COUNT=$(cat query.sql | sqlite3 $DB | wc -l)
echo , $COUNT entries found

if \[ $COUNT -gt 0 \] ; then
cat query.sql | sqlite3 $DB | sort | uniq 
> $OUTPUTDIR/output_wortlist_${POSTAGLIST}_${DAYMIN}_${DAYMAX}_${NAME}.txt

COUNT=$(cat query2.sql | sqlite3 $DB | wc -l)
COUNT=$(COUNT + 1)

bash

COUNT=$(cat query.sql | sqlite3 $DB | wc -l)
echo , $COUNT entries found

if \[ $COUNT -gt 0 \] ; then
cat query.sql | sqlite3 $DB | sort | uniq 
> $OUTPUTDIR/output_wortlist_${POSTAGLIST}_${DAYMIN}_${DAYMAX}_${NAME}.txt

COUNT=$(cat query2.sql | sqlite3 $DB | wc -l)
COUNT=$(COUNT + 1)

bash

bash

bash

bash

bash

bash

bash

bash

bash

bash

bash

bash

bash
5.2. DIE FUNKTIONEN DES CORPUSANALYSERS

```bash
echo $WORDCNT \n>> $OUTPUTDIR/output_typetoken_lemma_${POSTAGLIST}_${DAYMIN}_${DAYMAX}_${NAME}.csv
done
echo "—-|—-"\n>> $OUTPUTDIR/output_typetoken_lemma_${POSTAGLIST}_${DAYMIN}_${DAYMAX}_${NAME}.csv
echo "$CNT|$WORDCNTSUM"\n>> $OUTPUTDIR/../../output_typetoken_lemma_summary_${DAYMIN}_${DAYMAX}_${NAME}.csv
rm -f query2.sql
fi
rm -f query.sql
db_query_all_typetoken_lemma_name.sh: In Entsprechung zu den anderen "_all...sh"-Skripten wird die Abfrage durch dieses Skript durch das dazugehörige Unterkonzept, in diesem Falle db_query_typetoken_lemma_name.sh, ausgeführt. Unter Verwendung dieses Skriptes bestehen zahlreiche Möglichkeiten, die Abfragen zu kombinieren und auszuführen. Drei der wichtigsten Variablen sind: DAYMINMAX, POSTAGS und NAMES. Durch eine Variation von DAYMINMAX können die Zeiträume eingegrenzt werden. Es gibt die Möglichkeit alle drei Altersgruppen zusammen zu untersuchen oder aber nur eine oder zwei. Ebenso können alle POS-Tags zusammen abgefragt werden (in einer, zwei oder drei Altersgruppen) oder nur bestimmte. In diesen Fällen müssen die Zeilen hinter POSTAGS= angepasst werden. Durch den Befehl NAMES entsteht die Möglichkeit, alle Sprecher/innen in die Analyse einzubeziehen oder nur eine/n oder einige auszuwählen. Um dies an einem Beispiel zu verdeutlichen, soll angenommen werden, dass nur die Lemmata, die sich unter den Adjektiven (ADJA, ADJD) in der Altersgruppe 1709 bis 1983 befinden, von Interesse sind. Zusätzlich beziehen sich diese Informationen nur auf den Sprecher LEO. In diesem Fall sehen die Befehle betreffend DAYMINMAX, POSTAGS und NAMES folgendermaßen aus:

DAYMINMAX="1709:1983" POSTAGS="ADJA ADJD" NAMES="LEO"

Das vollständige Skript besteht aus den folgenden Zeilen:

Skript: db_query_all_typetoken_lemma_name.sh

DAYMINMAX="1281:1495 1496:1708 1709:1983"
POSTAGS="ADJA ADJD ADV APPR APPRART ART CARD FM ITJ KOKOM KON KOUS NE NN PAV"
POSTAGS="$POSTAGS PDAT PID PIAT PIS PPER PPOSAT PPOSS PRELAT PRELS PRF"
POSTAGS="$POSTAGS PTKA PTKANT PTKNEG PTKVZ PTKZU PWAT PWA VAF VAINF VAPP"
POSTAGS="$POSTAGS VAINF VAPP VMFIN VMMP VVFIN VVIMP VVIP VVIZU VVPP XY"
POSTAGS="$POSTAGS ADJA+ADJD ADV+PAV APPR+APPRART ART+CARD"
KAPITEL 5. DIE ARBEIT MIT DEM CORPUSANALYSE

POSTAGS="$POSTAGS VMFIN+VMINF+VMPP VVFIN+VVIMP+VVINF+VVIZU+VVPP XY+XY"
 NAMES="av mm js rd mk leo lar so lua jk ll ma"
OUTPUT_DIR=./..//03_db_query_results/typetoken_lemma
for n in $NAMES ; do
for d in $DAYMINMAX ; do
if [[ $d =~ ^(.*):(.*)$ ]]; then
DUMMYINFO="echo ${BASH_REMATCH[1]}"
DUMMYMAX="echo ${BASH_REMATCH[2]}"
rm -f $OUTPUT_DIR/output_typetoken_lemma_summary_$(DAYMIN)$(DAYMAX)$_{$(n)}.csv
for i in $POSTAGS ; do
if [ ! -e $OUTPUT_DIR/${DAYMIN}_${DAYMAX}/$n ]; then
mkdir -p $OUTPUT_DIR/${DAYMIN}_${DAYMAX}/$n
fi
echo -n $DAYMIN $DAYMAX $i 
$OUTPUT_DIR/${DAYMIN}_${DAYMAX}/$n 
./db_query_typetoken_lemma_name.sh $DAYMIN $DAYMAX $i 
$OUTPUT_DIR/${DAYMIN}_${DAYMAX}/$n $n
fi
rm -rf $OUTPUT_DIR/${DAYMIN}_${DAYMAX}/$n
CNTSUM=0
WORDCNTSUM=0
if [ -f $OUTPUT_DIR/output_typetoken_lemma_summary_$(DAYMIN)$(DAYMAX)$_{$(n)}.csv ]; then
FILE="cat $OUTPUT_DIR/output_typetoken_lemma_summary_$(DAYMIN)$(DAYMAX)$_{$(n)}.csv"
for i in $FILE ; do
if [[ $i =~ ^(.*)\|.*\|.*\|.*\|(.*)\|(.*)$ ]]; then
TAG="echo ${BASH_REMATCH[1]}"
CNT="echo ${BASH_REMATCH[2]}"
WORDCNT="echo ${BASH_REMATCH[3]}"
if [[ $TAG =~ .*\+.* ]]; then
CNTSUM=$((CNTSUM + CNT))
WORDCNTSUM=$((WORDCNTSUM + WORDCNT))
fi
fi
done
if [ -S $CNTSUM -o -S $WORDCNTSUM ]; then
rm -f $OUTPUT_DIR/output_typetoken_lemma_summary_$(DAYMIN)$(DAYMAX)$_{$(n)}.tmp
for i in $FILE ; do
if [[ $i =~ ^(.*)\|.*\|.*\|.*\|(.*)\|(.*)$ ]]; then
CNT="echo ${BASH_REMATCH[1]}"
WORDCNT="echo ${BASH_REMATCH[2]}"
WORDPERCENT="echo "scale=5; $WORDCNT*100/$WORDCNTSUM" | bc | sed -e \"s/\./,/g\"
CNTPERCENT="echo "scale=5; $CNT*100/$CNTSUM" | bc | sed -e \"s/\./,/g\"
echo "$i|$([CNTPERCENT]/$[WORDPERCENT])"
>> $OUTPUT_DIR/output_typetoken_lemma_summary_$(DAYMIN)$(DAYMAX)$_{$(n)}.csv.tmp
fi
done
echo "—-|—-|—-|—-|—-|—-|—-|—-"
>> $OUTPUT_DIR/output_typetoken_lemma_summary_$(DAYMIN)$(DAYMAX)$_{$(n)}.csv.tmp
echo "|
$OUTPUT_DIR/output_typetoken_lemma_summary_$(DAYMIN)$(DAYMAX)$_{$(n)}.csv.tmp
mv $OUTPUT_DIR/output_typetoken_lemma_summary_$(DAYMIN)$(DAYMAX)$_{$(n)}.csv.tmp
$OUTPUT_DIR/output_typetoken_lemma_summary_$(DAYMIN)$(DAYMAX)$_{$(n)}.csv
fi
fi
done
5.2. DIE FUNKTIONEN DES CORPUSANALYSERS

5.2.4.4 04_lemma_tag


Skript: db_query_lemma_tag.sh

```bash
DB=../../../02_generated_files/corpus.db
if [ -z $1 ] || [ -z $2 ] || [ -z $3 ] || [ -z $4 ] || [ ! -d $4 ] ; then
    echo "usage $0 daymin daymax postag outputdir"
    exit
fi
DAYMIN=$1
DAYMAX=$2
```
KAPITEL 5. DIE ARBEIT MIT DEM CORPUSANALYSER

Abbildung 5.19: CA: Beispiel lemma_tag 1496 bis 1708

5.2. DIE FUNKTIONEN DES CORPUSANALYSERS

Abbildung 5.20: CA: Beispiel all_lemma_tag MA 1281 bis 1495

Skript: db_query_all_lemma_tag.sh

```
DASYMINMAX="1281:1495 1496:1708 1709:1983"
POSTAGS="ADJA ADJD ADV APPR APPRART ART CARD"
POSTAGS="$POSTAGS FM ITJ KOKOM KON KOUS NE NN PAV"
POSTAGS="$POSTAGS FM ITJ KOKOM KON KOUS NE NN PAV"
POSTAGS="$POSTAGS FM ITJ KOKOM KON KOUS NE NN PAV"
POSTAGS="$POSTAGS FM ITJ KOKOM KON KOUS NE NN PAV"
POSTAGS="$POSTAGS FM ITJ KOKOM KON KOUS NE NN PAV"
POSTAGS="$POSTAGS FM ITJ KOKOM KON KOUS NE NN PAV"

for d in $DASYMINMAX ; do
  if [[ $d =~ ^(.*):(.*)$ ]]; then
    DAYMIN="echo ${BASH_REMATCH[1]}"
    DAYMAX="echo ${BASH_REMATCH[2]}"
    for i in $POSTAGS ; do
      if ! -e $OUTPUT_DIR/${DAYMIN}_${DAYMAX} ; then
        mkdir -p $OUTPUT_DIR/${DAYMIN}_${DAYMAX}
      fi
      echo $DAYMIN $DAYMAX $i $OUTPUT_DIR/${DAYMIN}_${DAYMAX}
      ./db_query_lemma_tag.sh $DAYMIN $DAYMAX $i $OUTPUT_DIR/${DAYMIN}_${DAYMAX}
    done
  fi
done
```

KAPITEL 5. DIE ARBEIT MIT DEM CORPUSANALYSE

von MA im Zeitraum 1281 bis 1495 geäußert wurden.

Skript: db_query_lemma_tag_name.sh

```
DB=../../../02_generated_files/corpus.db
if [ -z $1 ] || [ -z $2 ] || [ -z $3 ] || [ -z $4 ] || [ -z $5 ]
  then
  echo "usage $0 daymin daymax postag outputdir name"
  exit
fi
DAYMIN=$1
DAYMAX=$2
POSTAG="$3"
OUTPUTDIR=$4
NAME="$5"
echo -n 'SELECT sprecher.name, sprecher.geburtsdatum,' >> query.sql
echo -n 'aufnahme.datum, julianday(aufnahme.datum) '-julianday(sprecher.geburtsdatum),'
  >> query.sql
echo -n 'aufnahme.esb_name, textpostags.wort,' >> query.sql
echo -n 'FROM sprecher,aufnahme,textpostags' >> query.sql
echo -n 'WHERE julianday(aufnahme.datum) '-julianday(sprecher.geburtsdatum)>
  $DAYMIN
  AND julianday(aufnahme.datum) '-julianday(sprecher.geburtsdatum) <
  $DAYMAX
  AND textpostags.postag = "$POSTAG"
  AND sprecher.kuerzel = "$NAME"
  AND aufnahme.sprecher_id = sprecher.id' >> query.sql
COUNT='cat query.sql | sqlite3 $DB | wc -l'
echo "Found $COUNT datarows"
if [ $COUNT -gt 0 ] ; then
  create file $OUTPUTDIR/output_${POSTAG}_${DAYMIN}_${DAYMAX}.csv
  cat query.sql | sqlite3 $DB > $OUTPUTDIR/output_${POSTAG}_${DAYMIN}_${DAYMAX}.csv
fi
rm query.sql
```

db_query_all_lemma_tag_name.sh: Dieses Skript korrespondiert mit
dem vorherigen Skript db_query_lemma_tag_name.sh, von welchem die
eigentliche Abfrage der Lemmata aus der Datenbank ausgeführt wird. Mit
Hilfe dessen können die Altersgruppen eingegrenzt werden - also nur eine,
zwei oder alle drei Gruppen. Ferner können die Lemmata in Bezug auf alle
POS-Tags abgefragt werden.

Abbildung 5.21: CA: Skript db_query_all_lemma_tag_name LL_1709_1983
oder nur einige ausgewählte. Schließlich ist es möglich einzugrenzen, welche Sprecher/innen für die Analyse in Betracht kommen. Abbildung 5.21 zeigt beispielhaft die Lemmata, die LL im Zeitraum 1709 bis 1983 als Nomen (NN) geäußert hat. Im Folgenden ist das entsprechende Skript für diese Abfrage abgebildet:

**Skript: db_query_all_lemma_NN_LL_1709_1983.sh**

```sh
DAYMINMAX="1709:1983"
POSTAGS="NN"
NAMES="ll"
OUTPUT_DIR=../../../03_db_query_results/lemma_tag
for n in $NAMES ; do
  for d in $DAYMINMAX ; do
    if [[ $d =~ ^(.*):(.*)$ ]]; then
      DAYMIN='echo ${BASH_REMATCH[1]}'
      DAYMAX='echo ${BASH_REMATCH[2]}'
      for i in $POSTAGS ; do
        if [ ! -e $OUTPUT_DIR/${DAYMIN}_${DAYMAX}/$n ]; then
          mkdir -p $OUTPUT_DIR/${DAYMIN}_${DAYMAX}/$n
        fi
        echo $DAYMIN $DAYMAX $i $OUTPUT_DIR/${DAYMIN}_${DAYMAX}/$n
        ./db_query_lemma_tag_name.sh $DAYMIN $DAYMAX $i
        rmdir –ignore-fail-on-non-empty $OUTPUT_DIR/${DAYMIN}_${DAYMAX}/$n
      done
    fi
  done
done
```

Im Basisskript können die Variablen `DAYMINMAX`, `POSTAGS` und `NAMES` in Anlehnung an die bisher beschriebenen `__all__...sh`-Skripte modifiziert werden, um so zu spezifischen Ergebnissen zu gelangen.

**Skript: db_query_all_lemma_tag_name.sh**

```sh
DAYMINMAX="1281:1495 1496:1708 1709:1983"
POSTAGS="ADJA ADJD ADV APPR APPRART ART CARD FM ITJ KOKOM KON KOUS NE NN PAV"
POSTAGS="$POSTAGS PDAT PDS PIAT PPER PPOSAT PPRELAT PRELS PRF"
POSTAGS="$POSTAGS PTKA PTKANT PTKNEG PTKVZ PWAT PWAV PWS VAFIN VAIMP"
POSTAGS="$POSTAGS VAINF VAPP VMFIN VMINF VMPP VVFIN VVIMP VVIZU VVPP XY"
NAMES="av mm js rd mk leo lar so lua jk il ma"
OUTPUT_DIR=../../../03_db_query_results/lemma_tag
for n in $NAMES ; do
  for d in $DAYMINMAX ; do
    if [[ $d =~ ^(.*):(.*)$ ]]; then
      DAYMIN='echo ${BASH_REMATCH[1]}'
      DAYMAX='echo ${BASH_REMATCH[2]}'
      for i in $POSTAGS ; do
        if [ ! -e $OUTPUT_DIR/${DAYMIN}_${DAYMAX}/$n ]; then
          mkdir -p $OUTPUT_DIR/${DAYMIN}_${DAYMAX}/$n
        fi
        echo $DAYMIN \ $DAYMAX \ $i \ $OUTPUT_DIR/${DAYMIN}_${DAYMAX}/$n
        ./db_query_lemma_tag_name.sh $DAYMIN $DAYMAX $i
        rmdir –ignore-fail-on-non-empty $OUTPUT_DIR/${DAYMIN}_${DAYMAX}/$n
      done
    fi
  done
done
```
5.2.4.5 05_lemma_bedeutung


Skript: db_query_lemma_bedeutung.sh

```
DB=./..../02_generated_files/corpus.db
if [ -z $1 ] || [ -z $2 ] || [ -z $3 ] || [ -z $4 ] || [ ! -d $4 ] ; then
  echo "usage $0 daymin daymax bedeutung outputdir"
  exit
fi
DAYMIN=$1
DAYMAX=$2
BEDEUTUNG="$3"
OUTPUTDIR="$4"

echo -n 'SELECT sprecher.name, sprecher.geburtsdatum, ' > query.sql
echo -n 'aufnahme.datum, julianday(aufnahme.datum) ' >> query.sql
echo -n '-julianday(sprecher.geburtsdatum), ' >> query.sql
echo -n ' aufnahme.esb_name, textpostags.wort, ' >> query.sql
echo -n 'textpostags.lemma, textpostags.postag, ' >> query.sql
echo -n 'textpostags.bedeutung, textpostags.id ' >> query.sql
echo -n ' FROM sprecher,aufnahme,textpostags ' >> query.sql
echo -n 'WHERE julianday(aufnahme.datum) ' >> query.sql
echo -n '-julianday(sprecher.geburtsdatum) > ' >> query.sql
echo -n $DAYMIN >> query.sql
echo -n ' AND julianday(aufnahme.datum) ' >> query.sql
echo -n '-julianday(sprecher.geburtsdatum) < ' >> query.sql
echo -n $DAYMAX >> query.sql
echo -n ' AND textpostags.bedeutung = "' >> query.sql
echo -n $BEDEUTUNG >> query.sql
echo -n '" AND aufnahme.sprecher_id = sprecher.id ' >> query.sql
echo -n ' AND textpostags.aufnahme_id = aufnahme.id; ' >> query.sql

cat query.sql | sqlite3 $DB
> $OUTPUTDIR/output_${BEDEUTUNG}_${DAYMIN}_${DAYMAX}.csv
rm query.sql
```
5.2. DIE FUNKTIONEN DES CORPUSANALYSERS


Skript: db_query_all_lemma_bedeutung.sh

```bash
DAYMINMAX="1281:1495 1496:1708 1709:1983"
BEDEUTUNG="ad-part ad-qual ad-quant ad-rel"
BEDEUTUNG="$BEDEUTUNG adv-kaus adv-komm adv-lok"
BEDEUTUNG="$BEDEUTUNG adv-mod adv-temp art fm handlung"
BEDEUTUNG="$BEDEUTUNG iij kon-add kon-adv kon-alt kon-klaus kon-spez"
BEDEUTUNG="$BEDEUTUNG kon-temp kon-vgl n-abstr-hdlig n-abstr-maß"
BEDEUTUNG="$BEDEUTUNG n-abstr-vorg n-abstr-vorst n-abstr-wiss n-abstr-zeit"
BEDEUTUNG="$BEDEUTUNG n-abstr-zust n-abstr-eig n-belebt ne"
BEDEUTUNG="$BEDEUTUNG n-unbelebt pav pdat pds"
BEDEUTUNG="$BEDEUTUNG piat pidat pis pper ppos"
BEDEUTUNG="$BEDEUTUNG prels prf pr-kaus pr-lok pr-mod pr-neutr pr-temp"
BEDEUTUNG="$BEDEUTUNG pkw-kaus pkw-komm pkw-mod pkw-neu"
BEDEUTUNG="$BEDEUTUNG pkwz ptk zu pwat pwa vpw sub-fin"
BEDEUTUNG="$BEDEUTUNG sub-klaus sub-kond sub-komm sub-mod-sub neutr sub-temp"
BEDEUTUNG="$BEDEUTUNG v-aux v-kop v-mod"
BEDEUTUNG="$BEDEUTUNG vorgang xy zustand"
OUTPUT_DIR=../../../03_db_query_results/lemma_bedeutung
for d in $DAYMINMAX ; do
  if [[ $d =~ ^(.*)$:.*$ ]]; then
    DAYMIN='echo ${BASH_REMATCH[1]}'
    DAYMAX='echo ${BASH_REMATCH[2]}'
    for i in $BEDEUTUNG ; do
      if ! -e $OUTPUT_DIR/${DAYMIN}_${DAYMAX} ; then
        mkdir -p $OUTPUT_DIR/${DAYMIN}_${DAYMAX}
        ./db_query_lemma_bedeutung.sh $DAYMIN $DAYMAX $i
        done
      fi
    done
  fi
  done
```

5.2.4.6 06_typetoken_bedeutung

db_query_typetoken_bedeutung.sh: Mit dieser Abfrage werden die Types in Bezug auf die Lemmata aus der Datenbank abgefragt und mit den dazugehörigen Häufigkeiten in einer CSV-Datei entsprechend ihrer Bedeutungskategorie ausgegeben. Das Skript db_query_typetoken_bedeutung.sh arbeitet ähnlich wie das Skript db_query_typetoken_lemma.sh
mit dem Unterschied, dass nicht die Lemmata je POS-Tag aufgelistet werden, sondern die Lemmata je Bedeutungskategorie.

db_query_typetoken_bedeutung.sh

```
DB=./../02_generated_files/corpus.db
if [ -z $1 ] || [ -z $2 ] || [ -z $3 ] || [ -z $4 ] || [ ! -d $4 ]; then
    echo "usage $0 daymin daymax bedeutung outputdir"
    exit
fi
DAYMIN=$1
DAYMAX=$2
BEDEUTUNG="$3"
OUTPUTDIR="$4"
echo -n 'SELECT textpostags.lemma ' > query.sql
echo -n 'FROM sprecher,aufnahme,textpostags ' >> query.sql
echo -n 'WHERE julianday(aufnahme.datum) ' >> query.sql
echo -n '- julianday(sprecher.geburtsdatum) > ' >> query.sql
echo -n $DAYMIN >> query.sql
echo -n ' AND julianday(aufnahme.datum) ' >> query.sql
echo -n '- julianday(sprecher.geburtsdatum) < ' >> query.sql
echo -n $DAYMAX >> query.sql
echo -n ' AND textpostags.bedeutung = "' >> query.sql
echo -n $BEDEUTUNG >> query.sql
echo -n '" AND aufnahme.sprecher_id = sprecher.id ' >> query.sql
echo -n 'AND textpostags.aufnahme_id = aufnahme.id;' >> query.sql
cat query.sql | sqlite3 $DB | sort | uniq > $OUTPUTDIR/output_wortlist_${BEDEUTUNG}_${DAYMIN}_${DAYMAX}.txt
WORTLIST='cat $OUTPUTDIR/output_wortlist_${BEDEUTUNG}_${DAYMIN}_${DAYMAX}.txt'
CNT=0
WORDCNTSUM=0
rm -f $OUTPUTDIR/output_typetoken_bedeutung_${BEDEUTUNG}_${DAYMIN}_${DAYMAX}.csv
for i in $WORTLIST ; do
    echo -n 'SELECT textpostags.bedeutung ' > query2.sql
    echo -n 'FROM sprecher,aufnahme,textpostags ' >> query2.sql
    echo -n 'WHERE julianday(aufnahme.datum) ' >> query2.sql
    echo -n '- julianday(sprecher.geburtsdatum) > ' >> query2.sql
    echo -n $DAYMIN >> query2.sql
    echo -n ' AND julianday(aufnahme.datum) ' >> query2.sql
    echo -n '- julianday(sprecher.geburtsdatum) < ' >> query2.sql
    echo -n $DAYMAX >> query2.sql
    echo -n ' AND textpostags.bedeutung = "' >> query2.sql
    echo -n $BEDEUTUNG >> query2.sql
    echo -n '" AND textpostags.lemma = ' >> query2.sql
    echo -n $i" | sqlite3 $DB | wc -1 >> $OUTPUTDIR/output_typetoken_bedeutung_${BEDEUTUNG}_${DAYMIN}_${DAYMAX}.csv
    WORDCNT=`cat query2.sql | sqlite3 $DB | wc -l`
    WORDCNTSUM=`echo $WORDCNTSUM + $WORDCNT`
    CNT=$((CNT + 1))
    echo $WORDCNT >> $OUTPUTDIR/output_typetoken_bedeutung_${BEDEUTUNG}_${DAYMIN}_${DAYMAX}.csv
done
echo "—-|—-" >> $OUTPUTDIR/output_typetoken_bedeutung_${BEDEUTUNG}_${DAYMIN}_${DAYMAX}.csv
```

KAPITEL 5. DIE ARBEIT MIT DEM CORPUSANALYSER
5.2. DIE FUNKTIONEN DES CORPUSANALYSERS

```bash
echo "${BEDEUTUNG}|${DAYMIN}|${DAYMAX}|${CNT}|$WORDCNTSUM"
>> $OUTPUTDIR/../output_typetoken_bedeutung_summary.csv
rm -f query.sql
rm -f query2.sql
db_query_all_typetoken_bedeutung.sh: Mit Hilfe dieses Unterkriptes können die Ergebnisse der Abfrage db_query_typetoken_bedeutung.sh eingegrenzt werden oder aber es können sämtliche Informationen ausgegeben werden. Es ist möglich (wie in dieser Arbeit geschehen), alle Altersgruppen und Bedeutungskategorien in die Analyse einzubeziehen. Dementsprechend werden alle Bedeutungskategorien je Altersgruppe ausgegeben und die dazugehörigen Lemmata alphabetisch sortiert aufgelistet und mit deren Anzahl versehen.

Skript: db_query_all_typetoken_bedeutung.sh

```
DAYMINMAX="1281:1495 1496:1708 1709:1983"
BEDEUTUNG="ad-part ad-qual ad-quant ad-rel"
BEDEUTUNG="$BEDEUTUNG adv-kaus adv-komm adv-lok"
BEDEUTUNG="$BEDEUTUNG adv-mod adv-temp art fm handlung"
BEDEUTUNG="$BEDEUTUNG itj kon-add kon-adv kon-alt kon-kaus kon-spez"
BEDEUTUNG="$BEDEUTUNG kon-temp kon-vgl n-abstr-hdlg n-abstr-maß"
BEDEUTUNG="$BEDEUTUNG n-abstr-vorg n-abstr-vorst n-abstr-wiss n-abstr-zeit"
BEDEUTUNG="$BEDEUTUNG n-abstr-zust n-abstr-eig n-belebt ne"
BEDEUTUNG="$BEDEUTING n-unbelebt pav pdat pds"
BEDEUTUNG="$BEDEUTUNG piat pidat pis pper ppos"
BEDEUTING="$BEDEUTUNG prels prf pr-kaus pr-lok pr-neutr pr-temp"
BEDEUTING="$BEDEUTING ptk-abt ptk-ant ptk-fok ptk-gespr"
BEDEUTING="$BEDEUTING ptk-grad ptk-neg"
BEDEUTING="$BEDEUTING ptkvz ptkzu pwat pwav pws sub-fin"
BEDEUTING="$BEDEUTING sub-kaus sub-kond sub-konz sub-mod-instr sub-neutr sub-temp"
BEDEUTING="$BEDEUTING v-aux v-kop v-mod"
BEDEUTING="$BEDEUTING vorgang xy zustand"
OUTPUT_DIR=../../../03_db_query_results/typetoken_bedeutung
for d in $DAYMINMAX ; do
  if [[ $d =~ ^(.*)(.*)$ ]] ; then
    DAYMIN=\"echo ${BASH_REMATCH[1]}\"
    DAYMAX=\"echo ${BASH_REMATCH[2]}\"
    for i in $BEDEUTING ; do
      mkdir -p $OUTPUT_DIR/${DAYMIN}_${DAYMAX} ; then
        echo $DAYMIN $DAYMAX $i $OUTPUT_DIR/${DAYMIN}_${DAYMAX} $OUTPUT_DIR/${DAYMIN}_${DAYMAX}
        ./db_query_typetoken_bedeutung.sh $DAYMIN $DAYMAX $i $OUTPUT_DIR/${DAYMIN}_${DAYMAX} $OUTPUT_DIR/${DAYMIN}_${DAYMAX}
      done
    done
  fi
done
```

Es wäre ebenso möglich, nur eine oder zwei Altersgruppen oder ausgewählte Bedeutungskategorien zu untersuchen. Dies kann im Skript entsprechend angepasst werden. Soll etwa nur die Altersgruppe 1496_1708 betrachtet
werden und interessieren in dieser Gruppe nur die Adverbien, so würde das Skript zur Ausführung folgendermaßen aussehen:

**Skript: db_query_ADV_1496_1708_typetoken_bedeutung.sh**

```bash
DAYMINMAX="1496:1708"
BEDEUTUNG="adv-kaus adv-komm adv-lok"
BEDEUTUNG="$BEDEUTUNG adv-mod adv-temp"
OUTPUT_DIR=../../../03_db_query_results/typetoken_bedeutung
for d in $DAYMINMAX; do
  if [[ $d =~ ^(.+):(.+)$ ]]; then
    DAYMIN='echo ${BASH_REMATCH[1]}'
    DAYMAX='echo ${BASH_REMATCH[2]}'
    for i in $BEDEUTUNG; do
      if ! -e $OUTPUT_DIR/${DAYMIN}_${DAYMAX}; then
        mkdir -p $OUTPUT_DIR/${DAYMIN}_${DAYMAX}
      fi
      echo $DAYMIN $DAYMAX $i $OUTPUT_DIR/${DAYMIN}_${DAYMAX}
      ./db_query_typetoken_bedeutung.sh $DAYMIN $DAYMAX $i $OUTPUT_DIR/${DAYMIN}_${DAYMAX}
    done
  fi
donedefine
```

5.3 **Möglichkeiten für die Arbeit mit dem CorpusAnalyser**


Es ist weiterhin vorstellbar, dass in einer Analyse nicht nur die Wortarten und Wortbedeutungen von Interesse sind, sondern die Morpheme oder Phoneme einer Sprache in einem gegebenen Kontext. Auch dies ließe sich mit dem CA bewerkstelligen. In der entsprechenden EXMARaLDA-Datei müssten dann nicht die Wörter voneinander separiert aufgeführt und mit den jeweiligen POS-Tags bezeichnet werden. Vielmehr wäre es erforderlich,
die Morpheme bzw. Phoneme voneinander zu trennen. Ob und wie diese Einheiten für eine spätere Analyse gekennzeichnet werden, hängt vom Ziel der jeweiligen Untersuchung ab.
Kapitel 6

Analyse der Wortarten


---

1 Die Datenbank corpus.db wurde im Zuge der Erstellung des CorpusAnalysers ([Mack (2014)]) für die Zwecke dieser Arbeit erarbeitet.
KAPITEL 6. ANALYSE DER WORTARTEN


6.1 Gruppierung der POS-Tags


²Es fehlt die Kategorie FM, weil sie in den Daten dieser Gruppe nicht vorkam. Somit sind hier nur 18 Kategorien enthalten.
6.1. GRUPPIERUNG DER POS-TAGS

Eine Zusammenfassung der Types ist ebenfalls vorhanden, obwohl diese erst weiter unten besprochen werden.

Abbildung 6.1: POS-Tags: tokens_types_gruppiert_1281_1495

<table>
<thead>
<tr>
<th>Kategorie</th>
<th>POS-Tags</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADJ</td>
<td>ADJA + ADJD</td>
</tr>
<tr>
<td>ADV</td>
<td>ADV + PAV</td>
</tr>
<tr>
<td>APPR</td>
<td>APPR + APPRART</td>
</tr>
<tr>
<td>ART</td>
<td>ART</td>
</tr>
<tr>
<td>CARD</td>
<td>CARD</td>
</tr>
<tr>
<td>FM</td>
<td>FM</td>
</tr>
<tr>
<td>ITJ</td>
<td>ITJ</td>
</tr>
<tr>
<td>KON</td>
<td>KOUS + KON + KOKOM</td>
</tr>
<tr>
<td>N</td>
<td>NN + NE</td>
</tr>
<tr>
<td>PIDATS</td>
<td>PDS + PDAT + PI5 + PIAT + PID + PIDAT</td>
</tr>
<tr>
<td>PPER</td>
<td>PPER</td>
</tr>
<tr>
<td>PPOS</td>
<td>PPOSS + PPOSAT</td>
</tr>
<tr>
<td>PREL</td>
<td>(PRELAT)+PRELS + PRF</td>
</tr>
<tr>
<td>PW</td>
<td>PWS + PWAT + PWAV</td>
</tr>
<tr>
<td>PTK</td>
<td>PTKZU + PTKNEG + PTKVZ + PTKANT + PTKA</td>
</tr>
<tr>
<td>VA</td>
<td>VAFIN + (VAINF) + VAIMP + (VAPP)</td>
</tr>
<tr>
<td>VM</td>
<td>VMFIN + (VMINF) + (VMIMP)</td>
</tr>
<tr>
<td>VV</td>
<td>VVFIN + VVINF + VVIMP + VVIP + VVIZU</td>
</tr>
<tr>
<td>XY</td>
<td>XY</td>
</tr>
</tbody>
</table>

Tabelle 6.1: POS-Tags zusammengefasst
**Erläuterungen** zu Tabelle 6.1³:

- ADJ: Adjektive
- ADV: Adverbien und Pronominaladverbien
- APPR: Präpositionen
- ART: Artikel (unbestimmt und bestimmt)
- CARD: Kardinalzahlen
- FM: Fremdsprachliches Material
- ITJ: Interjektionen
- KON: Konjunktionen und Subjunktionen
- N: Nomen und Eigennamen
- PIDATS: Indefinitpronomen; determinierende, attribuierende und substituierende Pronomen
- PPER: Personalpronomen (irreflexiv)
- PPOS: Possessivpronomen
- PREL: Relativpronomen und reflexives Personalpronomen
- PW: substituierende, attribuierende und adverbiale Interrogativ- und Relativpronomen
- PTK: Partikeln
- VA: Auxiliarverben
- VM: Modalverben
- VV: Vollverben
- XY: unverständliche Wörter

³in Klammern stehende POS-Tags kommen in den Daten dieser Arbeit nicht vor, zählen aber grundlegend zu der angegebenen Gruppe.
6.2 Altersgruppen im Vergleich

In diesem Abschnitt und den folgenden Unterabschnitten sollen die Ergebnisse der getätigten Analysen mittels des CA in Bezug auf die Verteilung der POS-Tags dargestellt werden. Zunächst wird die Verteilung der Tokens und die Verteilung der Types aufgezeigt sowie die Relation der Types zu den Tokens. Dazu wurden jeweils alle Tokens sowie alle Types, die in den Zeiträumen 1281 bis 1495, 1496 bis 1708 und 1709 bis 1983 geäußert wurden, addiert. Um das Verhältnis der Types zu den Tokens zu erhalten, wurde die Anzahl der Types durch die Anzahl der Tokens dividiert. Die unten stehende Tabelle fasst kurz das Vorkommen aller Tokens und Types in den jeweiligen Altersgruppen in absoluten Zahlen zusammen, ungeachtet dessen, welches Kind die jeweiligen Tokens und Types äußerte.

<table>
<thead>
<tr>
<th>Altersgruppe</th>
<th>Tokens</th>
<th>Types</th>
</tr>
</thead>
<tbody>
<tr>
<td>1281-1495</td>
<td>2916</td>
<td>462</td>
</tr>
<tr>
<td>1496-1708</td>
<td>12584</td>
<td>1058</td>
</tr>
<tr>
<td>1709-1983</td>
<td>11315</td>
<td>1088</td>
</tr>
</tbody>
</table>

Tabelle 6.2: Types und Tokens in den Altersgruppen

In der ersten Altersgruppe von 1281 bis 1495 Tagen gibt es im Vergleich zu den anderen Gruppen nur wenige Tokens. Rückblickend lässt sich das dadurch erklären, dass die Kinder in diesem Alter in der Regel noch nicht so aufgeschlossen mir gegenüber waren und deshalb nur wenig aus eigener Initiative erzählten. Oft musste ich versuchen, die Kinder zum Sprechen zu animieren. Da dies nicht immer erfolgreich war und die Kinder zudem spontan und aus eigenem Antrieb heraus sprechen sollten, waren weitere Aufnahmen nicht möglich. Dennoch lassen die Daten insgesamt vermuten, dass zumindest ein Trend aufgezeigt werden kann. Die mittlere und die letzte Gruppe weisen jeweils ähnliche Häufigkeiten auf. So äußerten Kinder im Alter von 1496 bis 1708 Tagen im gesamten Aufnahmezeitraum 12584 Wörter (Tokens), darunter 1058 Types. Im Alter von 1709 bis 1983 Tagen wurden 11315 Wörter (Tokens) geäußert, wovon sich 1088 voneinander unterschieden (Types).

6.2.1 Verteilung der Tokens

Unter Zuhilfenahme des CA (Kapitel 5) wurden alle tatsächlich geäußerten Wörter (Tokens) aller Kinder im Hinblick auf ihre Zugehörigkeit zu den POS-Tags analysiert. Des Weiteren wurde die Anzahl der Wörter je POS-Tag und je Altersgruppe vermerkt. Für diesen Analyseschritt wurde die entsprechende Abfrage mit den dazugehörigen Skripten des CA genutzt. Zur Ermittlung der Vorkommenshäufigkeit je POS-Tag und je Altersgruppe
wurde die Abfrage 5.2.4.2 mit den Skripten db_query_typetoken.sh sowie db_query_all_typetoken.sh gestellt. Als Ergebnis erhielt man eine Liste aller geäußerten Wörter je POS-Tag und Altersgruppe. Abbildung 6.5 zeigt beispielsweise alle geäußerten VVIMP (Vollverben im Imperativ) der Gruppe 1709_1983. Die zusammengefassten und gruppierten Ergebnisse sind in der unten stehenden Tabelle (6.3) dargestellt.

<table>
<thead>
<tr>
<th>POS-Tag</th>
<th>1281-1495</th>
<th>POS/ Summe</th>
<th>1496-1708</th>
<th>POS/ Summe</th>
<th>1709-1983</th>
<th>POS/ Summe</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADJ</td>
<td>107</td>
<td>3,7</td>
<td>480</td>
<td>3,8</td>
<td>512</td>
<td>4,5</td>
</tr>
<tr>
<td>ADV+PAV</td>
<td>427</td>
<td>14,6</td>
<td>2175</td>
<td>17,3</td>
<td>1462</td>
<td>12,9</td>
</tr>
<tr>
<td>APPR</td>
<td>81</td>
<td>2,8</td>
<td>474</td>
<td>3,8</td>
<td>362</td>
<td>3,2</td>
</tr>
<tr>
<td>ART</td>
<td>258</td>
<td>8,8</td>
<td>1125</td>
<td>8,9</td>
<td>787</td>
<td>7,0</td>
</tr>
<tr>
<td>CARD</td>
<td>26</td>
<td>0,9</td>
<td>177</td>
<td>1,4</td>
<td>329</td>
<td>2,9</td>
</tr>
<tr>
<td>FM</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0,02</td>
<td>8</td>
<td>0,07</td>
</tr>
<tr>
<td>ITJ</td>
<td>6</td>
<td>0,2</td>
<td>38</td>
<td>0,3</td>
<td>73</td>
<td>0,6</td>
</tr>
<tr>
<td>KON</td>
<td>235</td>
<td>8,1</td>
<td>897</td>
<td>7,1</td>
<td>635</td>
<td>5,6</td>
</tr>
<tr>
<td>N</td>
<td>264</td>
<td>9,1</td>
<td>971</td>
<td>7,7</td>
<td>1245</td>
<td>11,0</td>
</tr>
<tr>
<td>PIDATS</td>
<td>153</td>
<td>5,2</td>
<td>625</td>
<td>4,9</td>
<td>452</td>
<td>4,0</td>
</tr>
<tr>
<td>PPER</td>
<td>283</td>
<td>9,7</td>
<td>1339</td>
<td>10,6</td>
<td>1432</td>
<td>12,7</td>
</tr>
<tr>
<td>PPOS</td>
<td>37</td>
<td>1,3</td>
<td>118</td>
<td>0,9</td>
<td>128</td>
<td>1,1</td>
</tr>
<tr>
<td>PREL</td>
<td>8</td>
<td>0,3</td>
<td>56</td>
<td>0,4</td>
<td>38</td>
<td>0,3</td>
</tr>
<tr>
<td>PTK</td>
<td>309</td>
<td>10,6</td>
<td>1090</td>
<td>8,7</td>
<td>1073</td>
<td>9,5</td>
</tr>
<tr>
<td>PW</td>
<td>44</td>
<td>1,5</td>
<td>175</td>
<td>1,4</td>
<td>137</td>
<td>1,2</td>
</tr>
<tr>
<td>VA</td>
<td>34</td>
<td>1,2</td>
<td>158</td>
<td>1,3</td>
<td>165</td>
<td>1,5</td>
</tr>
<tr>
<td>VM</td>
<td>90</td>
<td>3,1</td>
<td>441</td>
<td>3,5</td>
<td>421</td>
<td>3,7</td>
</tr>
<tr>
<td>VV</td>
<td>493</td>
<td>16,9</td>
<td>2086</td>
<td>16,6</td>
<td>1916</td>
<td>16,9</td>
</tr>
<tr>
<td>XY</td>
<td>61</td>
<td>2,1</td>
<td>157</td>
<td>1,2</td>
<td>140</td>
<td>1,2</td>
</tr>
</tbody>
</table>

Tabelle 6.3: Vorkommen der POS-Tags (Tokens) in allen Altersgruppen im Überblick - gruppiert

Zunächst sind keine signifikanten Unterschiede in der Verwendungshäufigkeit einzelner Wortklassen im Verlauf von 1281 Tagen bis 1983 Tagen erkennbar. Im Gegenteil: Es bietet sich dem Betrachter ein nahezu konstantes Bild der Verteilung hinsichtlich der Verwendung der Wortklassen in den drei Altersgruppen. Leicht auffällig ist die Entwicklung der Verwendung der Konjunktionen (KON). Im Alter von 1281 bis 1495 Tagen nehmen KON einen Anteil von 8,1% hinsichtlich aller verwendeten Wortklassen ein. Im Alter von 1496 bis 1708 Tagen sind es 7,1% und in der Altersgruppe 1709 bis 1983 Tage sind es nur noch 5,6%. An dieser Stelle kann nicht eindeutig gesagt werden, ob es sich um einen weiterführenden Trend handelt. Dazu müssten weitere Altersgruppen getestet werden, die an die letzte Altersgruppe anschließen, also ab einem Alter von 1984 Tagen. Eine ebenso leichte Veränderung zeigt der Gebrauch der Personalpronomen (PPER). Während
Kinder im Alter von 1281 bis 1495 Tagen einen Anteil von 9,7% PPER in ihrer Verwendung aufweisen, sind es in der Gruppe der 1496 bis 1708 Tage alten Kinder 10,6% und bei den Kindern im Alter von 1709 bis 1983 Tagen 12,7%. Auch hier müssten, ähnlich wie bei den Auffälligkeiten bei den KON, weitere Untersuchungen getätigt werden, um herauszufinden, ob es sich um einen fortsetzenden Trend handelt. Die Abbildungen 6.2, 6.3 und 6.4 zeigen die Anteile der gruppierten POS-Tags in den einzelnen Altersgruppen.

Abbildung 6.2: Gruppierte POS-Tags, Verteilung Tokens, 1281 bis 1495 Tage


---

4Dazu werden folgend die drei Verbkategorien VA, VM und VV zusammengefasst. Dies ist insofern sinnvoll, als dass bei den Vergleichsdaten ebenfalls keine Unterteilung in weitere Verbklassen vorgenommen wurde.
KAPITEL 6. ANALYSE DER WORTARTEN

Abbildung 6.3: Gruppierte POS-Tags, Verteilung Tokens, 1496 bis 1708 Tage

9,1%. Alle weiteren Wortarten der STTS, wie sie für die hiesige Untersuchung verwendet wurden, sind meiner Meinung nach nur teilweise mit denen aus Kauschkes Arbeit vergleichbar, da sie einer anderen Klassifikation unterliegen. Zudem wurde die Studie mit einer anderen Methodik durchgeführt, was einen aussagekräftigen Vergleich nicht möglich macht. Die bisherigen Ergebnisse entsprechen jedoch dem, was Kauschke in Anbetracht ihrer Studie prognostizierte:

„Nomen sind von Anfang an vorhanden und wachsen insbesondere während des zweiten Lebensjahres an. […] Mit drei Jahren verwendet kein Kind mehr als 25% Nomen.“ ([Kauschke (1999)], S. 144)

und

„Mit 15 Monaten tauchen erstmals Verben auf, die dann erheblich ansteigen und mit 3 Jahren den stärksten Anteil am Lexikon ausmachen.“ ([Kauschke (1999)], S. 145)

Insgesamt wurden durch alle Altersstufen hinweg Partikeln (PTK) zu einem hohen Anteil verwendet (rund 10% in allen Gruppen). Weiterhin nimmt der Gebrauch der Adjektive (ADJ) mit zunehmendem Alter leicht zu, insbesondere, wenn auch die Kardinalzahlen (CARD) zu den Adjektiven gezählt werden (3,7% Adjektive + 0,9% Kardinalzahlen in Gruppe 1281_1495, 3,8% Adjektive + 1,4% Kardinalzahlen in Gruppe 1496_1708 und 4,5%
Abbildung 6.4: Gruppierte POS-Tags, Verteilung Tokens, 1709 bis 1983 Tage

Adjektive + 2,9% Kardinalzahlen in Gruppe 1709_1983). An späterer Stelle soll in dieser Arbeit geklärt werden, ob es sich inhaltlich dabei um neue Adjektive handelt, die zu den späteren Zeitpunkten auftreten oder aber, ob alle bisher verwendeten Adjektive nur häufiger gebraucht wurden (Abschnitt 7.2).

Abbildung 6.5: Beispiel: tokens_VVIMP_1709_1983

<table>
<thead>
<tr>
<th>Artikel</th>
<th>Wert</th>
</tr>
</thead>
<tbody>
<tr>
<td>deck</td>
<td>3</td>
</tr>
<tr>
<td>lang</td>
<td>1</td>
</tr>
<tr>
<td>geh</td>
<td>1</td>
</tr>
<tr>
<td>gb</td>
<td>4</td>
</tr>
<tr>
<td>guck</td>
<td>21</td>
</tr>
<tr>
<td>nor</td>
<td>2</td>
</tr>
<tr>
<td>komm</td>
<td>7</td>
</tr>
<tr>
<td>ass</td>
<td>3</td>
</tr>
<tr>
<td>leg</td>
<td>5</td>
</tr>
<tr>
<td>mach</td>
<td>5</td>
</tr>
<tr>
<td>nnnn</td>
<td>7</td>
</tr>
<tr>
<td>paas</td>
<td>4</td>
</tr>
<tr>
<td>schau</td>
<td>20</td>
</tr>
<tr>
<td>schaut</td>
<td>1</td>
</tr>
<tr>
<td>zeig</td>
<td>1</td>
</tr>
</tbody>
</table>

15 85
6.2.2 Verteilung der Types

In einem weiteren Schritt wurde die Anzahl und die Verteilung der Types je Altersgruppe ausgewertet. Für diesen Analyseschritt wurde (entsprechend der Analyse der Tokens) die entsprechende Abfrage mit den dazugehörigen Skripten des CA verwendet. Zur Ermittlung der Vorkommenshäufigkeit je POS-Tag und je Altersgruppe wurde die Abfrage 5.2.4.3 mit den Skripten `db_query_typetoken_lemma.sh` sowie `db_query_all_typetoken_lemma.sh` an die Datenbank gestellt. Im Ergebnis erhielt man eine Liste aller Lemmata je POS-Tag und Altersgruppe. Abbildung 6.6 zeigt einen Ausschnitt aus der Datei, die alle Types bzw. Lemmata bezogen auf die finiten Verben (VVFIN) der Kinder der Gruppe 1281_1495 enthält.

Abbildung 6.6: Beispiel: types_VVFIN_1281_1495

Tabelle 6.4 zeigt die Verteilung der Types hinsichtlich der gruppierten POS-Tags je Altersgruppe. Auf den ersten Blick erscheint die Verteilung der einzelnen POS-Tags (ähnlich der Tokens) über die drei Zeiträume hinweg gleichmäßig. Einen leichten Negativtrend zeigt die Verteilung der Präpositionen (APPR). Kinder der ersten Altersgruppe verwendeten einen Anteil von 2,4% verschiedenartiger APPR, Kinder der zweiten Altersgruppe nur noch 1,9% und Kinder der dritten Altersgruppe nur noch 1,7%.
6.2. ALTERSGRUPPEN IM VERGLEICH

<table>
<thead>
<tr>
<th>POS-Tag</th>
<th>1281-1495</th>
<th>POS/Summe</th>
<th>1496-1708</th>
<th>POS/Summe</th>
<th>1709-1983</th>
<th>POS/Summe</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADJ</td>
<td>44</td>
<td>9,5</td>
<td>108</td>
<td>10,2</td>
<td>93</td>
<td>8,5</td>
</tr>
<tr>
<td>ADV+PAV</td>
<td>49</td>
<td>10,6</td>
<td>98</td>
<td>9,3</td>
<td>103</td>
<td>9,5</td>
</tr>
<tr>
<td>APPR</td>
<td>11</td>
<td>2,4</td>
<td>20</td>
<td>1,9</td>
<td>19</td>
<td>1,7</td>
</tr>
<tr>
<td>ART</td>
<td>3</td>
<td>0,6</td>
<td>3</td>
<td>0,3</td>
<td>3</td>
<td>0,3</td>
</tr>
<tr>
<td>CARD</td>
<td>9</td>
<td>1,9</td>
<td>20</td>
<td>1,9</td>
<td>22</td>
<td>2</td>
</tr>
<tr>
<td>FM</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0,2</td>
<td>4</td>
<td>0,4</td>
</tr>
<tr>
<td>ITJ</td>
<td>4</td>
<td>0,9</td>
<td>12</td>
<td>1,1</td>
<td>23</td>
<td>2,1</td>
</tr>
<tr>
<td>KON</td>
<td>11</td>
<td>2,4</td>
<td>19</td>
<td>1,8</td>
<td>17</td>
<td>1,6</td>
</tr>
<tr>
<td>N</td>
<td>141</td>
<td>30,5</td>
<td>428</td>
<td>40,5</td>
<td>423</td>
<td>38,8</td>
</tr>
<tr>
<td>PIDATS</td>
<td>12</td>
<td>2,6</td>
<td>21</td>
<td>2,0</td>
<td>24</td>
<td>2,2</td>
</tr>
<tr>
<td>PPER</td>
<td>7</td>
<td>1,5</td>
<td>7</td>
<td>0,7</td>
<td>7</td>
<td>0,6</td>
</tr>
<tr>
<td>PPOS</td>
<td>4</td>
<td>0,9</td>
<td>5</td>
<td>0,5</td>
<td>5</td>
<td>0,5</td>
</tr>
<tr>
<td>PREL</td>
<td>6</td>
<td>1,3</td>
<td>9</td>
<td>0,8</td>
<td>9</td>
<td>0,8</td>
</tr>
<tr>
<td>PTK</td>
<td>43</td>
<td>9,3</td>
<td>57</td>
<td>5,4</td>
<td>73</td>
<td>6,7</td>
</tr>
<tr>
<td>PW</td>
<td>6</td>
<td>1,3</td>
<td>8</td>
<td>0,8</td>
<td>8</td>
<td>0,7</td>
</tr>
<tr>
<td>VA</td>
<td>3</td>
<td>0,6</td>
<td>3</td>
<td>0,3</td>
<td>3</td>
<td>0,3</td>
</tr>
<tr>
<td>VM</td>
<td>5</td>
<td>1,1</td>
<td>7</td>
<td>0,7</td>
<td>7</td>
<td>0,6</td>
</tr>
<tr>
<td>VV</td>
<td>102</td>
<td>22</td>
<td>229</td>
<td>21,6</td>
<td>240</td>
<td>22</td>
</tr>
<tr>
<td>XY</td>
<td>2</td>
<td>0,4</td>
<td>2</td>
<td>0,2</td>
<td>5</td>
<td>0,5</td>
</tr>
</tbody>
</table>

Tabelle 6.4: Vorkommen der POS-Tags (Types) in allen Altersgruppen im Überblick - gruppiert

Zudem folgen später Einzelanalysen der hier untersuchten Kinder, welche eine gänzlich andere Qualität aufweisen (Abschnitt 6.3).

Abbildung 6.7: Gruppierte POS-Tags, Verteilung Types, 1281 bis 1495 Tage

6.2.3 Das Verhältnis der Types zu den Tokens

Das Verhältnis der Types zu den Tokens soll dargestellt werden, um eventuelle Veränderungen in der Häufigkeit der Verwendung verschiedener Lemmata aufzuzeigen. Nachfolgend ist dieses Verhältnis in Bezug auf die drei Alterszeiträume dargestellt. Die Types- und Tokensanteile wurden hier anhand von Analysen der ungruppierten POS-Tags erfasst (Tabelle 6.5).

<table>
<thead>
<tr>
<th>Zeitraum</th>
<th>1281-1495</th>
<th>1496-1708</th>
<th>1709-1983</th>
</tr>
</thead>
<tbody>
<tr>
<td>Types</td>
<td>528</td>
<td>1207</td>
<td>1236</td>
</tr>
<tr>
<td>Tokens</td>
<td>2916</td>
<td>12584</td>
<td>11315</td>
</tr>
<tr>
<td>Verhältnis Types/Tokens</td>
<td>0,18</td>
<td>0,10</td>
<td>0,11</td>
</tr>
</tbody>
</table>

Tabelle 6.5: Type-Token-Verhältnis in den Altersgruppen

Das Verhältnis des ersten Untersuchungszeitraumes fällt im Gegensatz zur zweiten und dritten Untersuchungsperiode etwas höher aus (0,18 gegenüber 0,10 bzw. 0,11). Es gibt insgesamt aber viel weniger Tokens (2916) als in den beiden anderen Gruppen. Das Verhältnis der Types zu den Tokens beträgt 0,18. Im zweiten bzw. dritten Untersuchungszeitraum beträgt das Verhältnis 0,10 bzw. 0,11. An dieser Stelle kann nicht eindeutig festgelegt werden, ob der Tokens-Wert der ersten Gruppe gewertet werden sollte, weil die untersuchten
6.2. ALTERSGRUPPEN IM VERGLEICH

Abbildung 6.8: Gruppierte POS-Tags, Verteilung Types, 1496 bis 1708 Tage


individuellen Entwicklung bezieht:

„Varianzanalysen zeigen, daß mit zunehmendem Alter keine bedeutsamen Veränderungen stattfinden (F (3,81 = 0,42, p = 0,743). Daraus kann geschlossen werden, daß das Verhältnis von Types zu Tokens gleich bleibt. Die lexikalische Vielfalt unterliegt also keinen entwicklungsbedingten Schwankungen, sondern bleibt relativ konstant (Mittelwerte 0.44, 0.40, 0.38- 0.42).“ ([Kauschke (1999)], S. 143-144)


6.2.4 Zusammenfassung

Durch die Analyse und einen anschließenden Vergleich der drei Altersgruppen sollte ein eventuell auftretender Entwicklungsverlauf aufgezeigt werden. Aufgrund der hier angewandten Methodik der teilnehmenden Beobachtung kann ein Trend der Ergebnisse aufgezeigt werden, nicht aber Signifikanzen. Im Hinblick auf die Verwendungshäufigkeit der Wörter (Tokens) fiel auf, dass Konjunktionen (KON) im Verlauf von 3;5 bis 5;5 Jahren seltener gebraucht wurden, während bei den Personalpronom (PPER) einen leichter Aufwärtstrend erkennbar ist. In Anbetracht der Verschiedenartigkeit der verwendeten Wörter (Types) gab es zwei Auffälligkeiten. Im Verlauf von 3;5 bis 5;5 Jahren wurden weniger unterschiedliche Präpositionen (APPR) gebraucht. Die auffälligste Entwicklung zeigte allerdings die Verwendung verschiedener Nomen (N). Während die Kinder in Gruppe 1 noch 30,5%
KAPITEL 6. ANALYSE DER WORTARTEN


6.3 Individuelle Unterschiede in den Altersgruppen

In einem Folgeschritt wurden die Daten jedes der zwölf untersuchten Kinder einzeln analysiert. Dazu wurden die Daten eines jeden Kindes in jeder der drei Altersgruppen mit Hilfe des CA separiert und anschließend die Art und die Anzahl der jeweils vorkommenden POS-Tags mit den dazugehörigen Wörtern und Lemmata aufgelistet. Um beispielsweise die Anzahl aller Types und Tokens je Kind und Altersgruppe ausfindig zu machen, wurde die Funktion db_query_typedtoken_lemma.sh des CA genutzt (Abschnitt 5.2.4.3).

Durch die Ausführung der Skripte db_query_typedtoken_lemma_name.sh und db_query_all_typedtoken_lemma_name.sh (beide Skripte arbeiten zusammen) erfolgte die Abfrage der Types und Tokens. Dadurch wurden nicht nur - wie in der Abfrage db_query_all_typedtoken_lemma.sh - alle Lemmata je POS-Tag abgefragt und sortiert nach den Altersgruppen in eine CSV-Datei geschrieben. Vielmehr ist bei dieser Abfrage zusätzlich der Name (als Kürzel) übergeordnet, wodurch innerhalb der Altersgruppen noch einmal eine Sortierung nach den Sprechern erfolgte. Wenn nun beispielsweise eine solche ausgegebene Datei geöffnet wird, zeigt sich folgendes Bild (Abbildung 6.10). Diese Grafik zeigt alle von JS geäußerten attributiven (ADJA) im Zeitraum 1709_1983, als JS sich also in einem Altersrahmen von über 4;7 bis
6.3. INDIVIDUELLE UNTERSCHIEDE IN DEN ALTERSGRUPPEN

Abbildung 6.10: Beispiel: JS_lemmata_ADJA_1709_1983

5,5 Jahren befand. Die Ausgabe der CSV-Datei sieht sehr übersichtlich aus, wobei dennoch alle benötigten Informationen enthalten sind. In der untersten Zeile befindet sich die Gesamtanzahl der geäußerten Tokens (46), die Anzahl der Types beläuft sich auf 20 (bis zum Adjektiv weiß). Aufgelistet sind nur die Lemmata, weil die syntaktischen Wörter an dieser Stelle für die Zwecke dieser Arbeit zu wenig Informationen bieten.\(^5\) Zudem sind alle POS-Tags der Form ADJA aufgelistet, d.h., sie können auch inhaltlich beurteilt werden, wenn gewünscht. Ein weiteres Beispiel zeigt Abbildung 6.11, die alle von LEO geäußerten Nomen (NN) im Zeitraum 1709_1983 enthält.


\(^5\)Die Wörter mit den dazugehörigen Lemmata können durch ein Ausführen des Skriptes db_query_all_lemma_tag.sh abgerufen werden.

### 6.3.1 Gruppe 1281 bis 1495

In dieser Altersgruppe gab es lediglich zwei Kinder, die analysierbar waren. Die erste Vermutung war, dass eine Analyse aufgrund der geringen Anzahl der Tokens nicht möglich oder schwierig werden würde. Im weiteren Verlauf wurde aber deutlich, dass selbst mit diesen vergleichsweise wenigen Daten ein Ergebnis zu Tage tritt, welches mit dem anderer Altersgruppen in dieser Arbeit sowie auch mit den Ergebnissen anderer Arbeiten vergleichbar ist.


<table>
<thead>
<tr>
<th>Lemmat</th>
<th>Tokens</th>
</tr>
</thead>
<tbody>
<tr>
<td>auldr</td>
<td>2</td>
</tr>
<tr>
<td>bär</td>
<td>1</td>
</tr>
<tr>
<td>beben</td>
<td>1</td>
</tr>
<tr>
<td>becken</td>
<td>1</td>
</tr>
<tr>
<td>blume</td>
<td>2</td>
</tr>
<tr>
<td>ecke</td>
<td>1</td>
</tr>
<tr>
<td>fall</td>
<td>1</td>
</tr>
<tr>
<td>fuß</td>
<td>1</td>
</tr>
<tr>
<td>gesteht</td>
<td>1</td>
</tr>
<tr>
<td>haus</td>
<td>1</td>
</tr>
<tr>
<td>hühn</td>
<td>1</td>
</tr>
<tr>
<td>karate</td>
<td>1</td>
</tr>
<tr>
<td>ostergras</td>
<td>1</td>
</tr>
<tr>
<td>ostern</td>
<td>1</td>
</tr>
<tr>
<td>pause</td>
<td>1</td>
</tr>
<tr>
<td>schiff</td>
<td>2</td>
</tr>
<tr>
<td>schwarzbär</td>
<td>2</td>
</tr>
<tr>
<td>sommer</td>
<td>1</td>
</tr>
<tr>
<td>spielfigur</td>
<td>1</td>
</tr>
<tr>
<td>taucher</td>
<td>1</td>
</tr>
<tr>
<td>zeit</td>
<td>1</td>
</tr>
</tbody>
</table>

Abbildung 6.11: Beispiel: LEO_lemmata_NN_1709_1983

Tabelle 6.6 zeigt die Verteilung der gruppierten Tags von MA. Nomen (N) und Verben (VA, VM, VV) erscheinen mit 24,7% bzw. 22,4%, Adverbien (ADV) mit einer Häufigkeit von 9,0% bezogen auf die Types. Ein ähnliches Bild zeigt sich bei der Auswertung der Daten von MK (Tabelle 6.7). In Bezug auf die Types erscheinen Nomen (N) mit einer Häufigkeit von 27,1%, Verben (VA, VM, VV) mit einer Häufigkeit von 25,0% und Adverbien (ADV) mit einer Häufigkeit von 12,1%. In den folgenden Abschnitten (6.3.2; 6.3.3) wird deutlich, dass ein ähnliches Verteilungsmuster auch in den höheren Altersgruppen auftritt. Bei einer Betrachtung der Verteilung der Tokens, treten zum Teil Ergebnisse zu Tage, die in der Form nicht erwartet wurden. Der hohe Anteil an Nomen in Bezug auf die Types ist bei einer Betrachtung der Tokens nahezu verschwunden, während die Verteilung der Wortarten insgesamt bei einem Vergleich von Tokens zu Types nicht stark voneinander abweicht. Bei beiden hier untersuchten Kindern in Gruppe 1281 bis 1495 beträgt der Anteil der Types bei den Nomen mehr als doppelt soviel im Vergleich zu den Tokens. Ein höherer Type-Anteil ist zunächst nicht verwunderlich, aber in diesem Ausmaß nicht zu erwarten gewesen. Ähnliches kann in umgekehrter Form bei den Personalpronomen (PPER) beobachtet werden. Ein relativ geringer Type-Anteil steht einem hohen Token-Anteil gegenüber. Dies ist jedoch nicht weiter verwunderlich, weil es insgesamt betrachtet nicht viele verschiedene Personalpronomen im Deutschen gibt, die hätten verwendet werden können. Bei den Verben (VA, VM, VV) sind solch starke Effekte nicht zu verzeichnen. Der Tokens-Anteil ist bei beiden Kindern nur etwas geringer

\(^6\)Ein Vergleich der Daten zu den Adverbien mit denen Kauschkes war nicht möglich, da Kauschke diese Kategorie nicht analysierte.
als der Types-Anteil (20,4% gegenüber 22,4% bei MA und 21,3% gegenüber 25,0% bei MK).

<table>
<thead>
<tr>
<th>Tokens</th>
<th>prozentual %</th>
<th>Types</th>
<th>prozentual %</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADJ</td>
<td>38</td>
<td>6,8</td>
<td>22</td>
</tr>
<tr>
<td>ADV+PAV</td>
<td>58</td>
<td>10,3</td>
<td>16</td>
</tr>
<tr>
<td>APPR</td>
<td>7</td>
<td>1,2</td>
<td>5</td>
</tr>
<tr>
<td>ART</td>
<td>60</td>
<td>10,7</td>
<td>3</td>
</tr>
<tr>
<td>CARD</td>
<td>3</td>
<td>0,5</td>
<td>2</td>
</tr>
<tr>
<td>KON</td>
<td>38</td>
<td>6,8</td>
<td>4</td>
</tr>
<tr>
<td>N</td>
<td>59</td>
<td>10,5</td>
<td>44</td>
</tr>
<tr>
<td>PIDATS</td>
<td>11</td>
<td>2,0</td>
<td>5</td>
</tr>
<tr>
<td>PPER</td>
<td>57</td>
<td>10,2</td>
<td>6</td>
</tr>
<tr>
<td>PPOS</td>
<td>8</td>
<td>1,4</td>
<td>2</td>
</tr>
<tr>
<td>PREL</td>
<td>2</td>
<td>0,4</td>
<td>2</td>
</tr>
<tr>
<td>PTK</td>
<td>74</td>
<td>13,2</td>
<td>21</td>
</tr>
<tr>
<td>PW</td>
<td>19</td>
<td>3,4</td>
<td>4</td>
</tr>
<tr>
<td>VA</td>
<td>12</td>
<td>2,1</td>
<td>2</td>
</tr>
<tr>
<td>VM</td>
<td>12</td>
<td>2,1</td>
<td>4</td>
</tr>
<tr>
<td>VV</td>
<td>91</td>
<td>16,2</td>
<td>34</td>
</tr>
<tr>
<td>XY</td>
<td>12</td>
<td>2,1</td>
<td>2</td>
</tr>
</tbody>
</table>

Tabelle 6.6: MA: Verteilung der Types, gruppiert, 1281 bis 1495

6.3.2 Gruppe 1496 bis 1708

In dieser Altersgruppe bewegen sich die Anteile der Adverbiennomen und Verben bezogen auf die Types analog zu denen der Kinder in der ersten Altersgruppe. Das Vorkommen der Adverbiennomen (ADV) bewegt sich in einer Spanne von 11,3% bis 12,9%, das Vorkommen der Nomen (N) in einer Spanne von 24,7% bis 35,0% und die Verben (VA, VM, VV) in einer Spanne zwischen 22,5% bis 24,3%. Der etwas höhere Anteil hinsichtlich der Nomen bei MK soll zunächst nicht überwertet werden. Bei einer genauen Analyse aller Tokens von MK wird deutlich, dass MK eine besonders große Vielfalt an Eigennamen (NE) im Spiel verwendet, die in dieser Auswertung zusammen mit den Nomen (NN) zur Kategorie N zusammengefasst sind (Abschnitt 6.1). Da MK besonders häufig in Spielsituationen beobachtet wurde, in denen es um die eigene Fantasie geht, finden sich hier besonders viele Wortschöpfungen wie Nemofischchen, Frühlingsfell oder Kriegsstürmer. Diese Nomen finden sich in der Form nicht in konventionellen Wörterbüchern (z. B. [Dudenredaktion (2004)]) wieder, wurden hier aber in ihrer Funktion als Nomen berücksichtigt. Eine andere Erklärung kann in Anlehnung an Kauschkes (1999) Behauptung gemacht werden. Diese konstatiert, dass Kinder im
Tabelle 6.7: MK: Verteilung der Types, gruppiert, 1281 bis 1495

beschrieben und es gibt keine gravierenden individuellen Unterschiede in den Anteilen von Types gegenüber Tokens.

<table>
<thead>
<tr>
<th>POS-Tag</th>
<th>Tokens</th>
<th>prozentual %</th>
<th>Types</th>
<th>prozentual %</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADJ</td>
<td>80</td>
<td>5,9</td>
<td>34</td>
<td>11,6</td>
</tr>
<tr>
<td>ADV+PAV</td>
<td>207</td>
<td>15,3</td>
<td>36</td>
<td>12,3</td>
</tr>
<tr>
<td>APPR</td>
<td>41</td>
<td>3,0</td>
<td>10</td>
<td>3,4</td>
</tr>
<tr>
<td>ART</td>
<td>104</td>
<td>7,7</td>
<td>3</td>
<td>1,0</td>
</tr>
<tr>
<td>CARD</td>
<td>21</td>
<td>1,6</td>
<td>8</td>
<td>2,7</td>
</tr>
<tr>
<td>ITJ</td>
<td>2</td>
<td>0,1</td>
<td>2</td>
<td>0,7</td>
</tr>
<tr>
<td>KON</td>
<td>93</td>
<td>6,9</td>
<td>9</td>
<td>3,1</td>
</tr>
<tr>
<td>N</td>
<td>103</td>
<td>7,6</td>
<td>72</td>
<td>24,7</td>
</tr>
<tr>
<td>PIDATS</td>
<td>60</td>
<td>4,5</td>
<td>12</td>
<td>4,1</td>
</tr>
<tr>
<td>PPER</td>
<td>173</td>
<td>12,8</td>
<td>7</td>
<td>2,4</td>
</tr>
<tr>
<td>PPOS</td>
<td>8</td>
<td>0,6</td>
<td>2</td>
<td>0,7</td>
</tr>
<tr>
<td>PTK</td>
<td>101</td>
<td>7,5</td>
<td>20</td>
<td>6,8</td>
</tr>
<tr>
<td>PW</td>
<td>27</td>
<td>2,0</td>
<td>5</td>
<td>1,7</td>
</tr>
<tr>
<td>VA</td>
<td>26</td>
<td>1,9</td>
<td>2</td>
<td>0,7</td>
</tr>
<tr>
<td>VM</td>
<td>52</td>
<td>3,9</td>
<td>7</td>
<td>2,4</td>
</tr>
<tr>
<td>VV</td>
<td>234</td>
<td>17,4</td>
<td>62</td>
<td>21,2</td>
</tr>
<tr>
<td>XY</td>
<td>16</td>
<td>1,2</td>
<td>1</td>
<td>0,3</td>
</tr>
<tr>
<td></td>
<td>1348</td>
<td></td>
<td>292</td>
<td></td>
</tr>
</tbody>
</table>

Tabelle 6.8: AV: Verteilung der Types, gruppiert, 1496 bis 1708

6.3.3 Gruppe 1709 bis 1983

In der letzten Altersgruppe sind mit einer Anzahl von zehn Kindern die meisten auswertbaren Daten vorhanden. Mengenmäßig unterscheidet sich die Token-Anzahl jedoch nur wenig von der vorhergehenden Gruppe - 12584 gegenüber 11315. Die Spanne des Vorkommens der Adverbien (ADV) bewegt sich zwischen 10,1% und 16,7%, die Spanne der Nomen (N) zwischen 16,7% und 30,7% und die Spanne der Verben (VA, VM, VV) zwischen 21,0% und 28,1% bezogen auf die Types. Auffällig ist hier die Gruppe der Nomen. Mit einem Vorkommen von 16,7% unterscheidet sich dieses Muster von denen der jüngeren Altersgruppen. Dieser geringe Wert kommt in zwei von zehn Fällen vor (LUA, SO) bzw. nur ein wenig höher mit 18,0% bei drei Kindern (AV mit 18,0%, JK mit 18,1% und LAR mit 18,8%). Aus diesem Grund kann meines Erachtens an dieser Stelle nicht mehr von Ausreißern gesprochen werden. Ein möglicher Grund wäre jener, dass die Anzahl der auswertbaren
6.3. INDIVIDUELLE UNTERSCHIEDE IN DEN ALTERSGRUPPEN

<table>
<thead>
<tr>
<th>POS-Tag</th>
<th>Tokens</th>
<th>prozentual %</th>
<th>Types</th>
<th>prozentual %</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADJ</td>
<td>41</td>
<td>3,9</td>
<td>27</td>
<td>9,4</td>
</tr>
<tr>
<td>ADV+PAV</td>
<td>152</td>
<td>14,6</td>
<td>33</td>
<td>11,5</td>
</tr>
<tr>
<td>APPR</td>
<td>58</td>
<td>5,6</td>
<td>14</td>
<td>4,9</td>
</tr>
<tr>
<td>ART</td>
<td>121</td>
<td>11,6</td>
<td>3</td>
<td>1,0</td>
</tr>
<tr>
<td>CARD</td>
<td>6</td>
<td>0,6</td>
<td>3</td>
<td>1,0</td>
</tr>
<tr>
<td>ITJ</td>
<td>2</td>
<td>0,2</td>
<td>2</td>
<td>0,7</td>
</tr>
<tr>
<td>KON</td>
<td>79</td>
<td>7,6</td>
<td>11</td>
<td>3,8</td>
</tr>
<tr>
<td>N</td>
<td>123</td>
<td>11,8</td>
<td>79</td>
<td>27,5</td>
</tr>
<tr>
<td>PIDATS</td>
<td>53</td>
<td>5,1</td>
<td>12</td>
<td>4,2</td>
</tr>
<tr>
<td>PPER</td>
<td>93</td>
<td>8,9</td>
<td>5</td>
<td>1,7</td>
</tr>
<tr>
<td>PPPOS</td>
<td>12</td>
<td>1,1</td>
<td>2</td>
<td>0,7</td>
</tr>
<tr>
<td>PREL</td>
<td>7</td>
<td>0,7</td>
<td>4</td>
<td>1,4</td>
</tr>
<tr>
<td>PTK</td>
<td>56</td>
<td>5,4</td>
<td>18</td>
<td>6,3</td>
</tr>
<tr>
<td>PW</td>
<td>13</td>
<td>1,2</td>
<td>4</td>
<td>1,4</td>
</tr>
<tr>
<td>VA</td>
<td>9</td>
<td>0,9</td>
<td>2</td>
<td>0,7</td>
</tr>
<tr>
<td>VM</td>
<td>33</td>
<td>3,2</td>
<td>6</td>
<td>2,1</td>
</tr>
<tr>
<td>VV</td>
<td>174</td>
<td>16,7</td>
<td>60</td>
<td>20,9</td>
</tr>
<tr>
<td>XY</td>
<td>12</td>
<td>1,1</td>
<td>2</td>
<td>0,7</td>
</tr>
<tr>
<td>1044</td>
<td>287</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Tabelle 6.9: LEO: Verteilung der Types, gruppiert, 1496 bis 1708

Tokens dieser Kinder zu gering war. Ähnlich geringe Werte weisen jedoch auch andere Kinder mit einem höheren prozentualen Nomenanteil auf, so dass eine leichte Veränderung der Verteilung der Wortarten hinsichtlich der Nomen ab einem Alter von 1709 Tagen angenommen werden kann.

Bei einer Betrachtung der Tokens fällt auf, dass auch hier die Gruppe der Nomen besonders hervorsticht. Während bei den Types noch immer relativ hohe Anteile von bis zu 30,7% zu verzeichnen sind, liegt er bei den Tokens weit darunter und ist vergleichbar mit jenem der vorhergehenden Altersgruppen. Dennoch sind Types-Anteile bezüglich der Nomen zu verzeichnen, die mit einem geringen Prozentsatz von ca. 16% bis 18% auf eine allmähliche allgemeine Veränderung in der Komposition des Lexikons hinweisen. Bei den Verben sowie allen weiteren Wortarten sind keine Besonderheiten erkennbar.⁷

6.3.4 Vergleich der individuellen Verläufe mit den Altersgruppen

Nach den bisherigen Analysen der vorliegenden Daten wird ersichtlich, dass die Daten aus Abschnitt 6.2 nicht ohne Weiteres vergleichbar sind mit den Ergebnissen aus Abschnitt 6.3. Sieht man sich die Verteilung der Wortarten

⁷Ausgenommen sind an dieser Stelle die Personalpronomen (PPER), die, wie oben erläutert, nie einen bestimmten Type-Anteil überschreiten können.

Eine Betrachtung der Verteilung der Wortarten und insbesondere das
6.3. **INDIVIDUELLE UNTERSCHIEDE IN DEN ALTERSGRUPPEN**

<table>
<thead>
<tr>
<th>POS-Tag</th>
<th>Tokens</th>
<th>prozentual %</th>
<th>Types</th>
<th>prozentual %</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADJ</td>
<td>325</td>
<td>3,7</td>
<td>90</td>
<td>11,1</td>
</tr>
<tr>
<td>ADV+PAV</td>
<td>1603</td>
<td>18,4</td>
<td>91</td>
<td>11,3</td>
</tr>
<tr>
<td>APPR</td>
<td>309</td>
<td>3,5</td>
<td>20</td>
<td>2,5</td>
</tr>
<tr>
<td>ART</td>
<td>763</td>
<td>8,8</td>
<td>3</td>
<td>0,4</td>
</tr>
<tr>
<td>CARD</td>
<td>125</td>
<td>1,4</td>
<td>13</td>
<td>1,6</td>
</tr>
<tr>
<td>FM</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0,2</td>
</tr>
<tr>
<td>ITJ</td>
<td>29</td>
<td>0,3</td>
<td>8</td>
<td>1,0</td>
</tr>
<tr>
<td>KON</td>
<td>648</td>
<td>7,4</td>
<td>17</td>
<td>2,1</td>
</tr>
<tr>
<td>N</td>
<td>586</td>
<td>6,7</td>
<td>283</td>
<td>35,0</td>
</tr>
<tr>
<td>PIDATS</td>
<td>428</td>
<td>4,9</td>
<td>21</td>
<td>2,6</td>
</tr>
<tr>
<td>PPER</td>
<td>912</td>
<td>10,5</td>
<td>7</td>
<td>0,9</td>
</tr>
<tr>
<td>PPOS</td>
<td>72</td>
<td>0,8</td>
<td>5</td>
<td>0,6</td>
</tr>
<tr>
<td>PREL</td>
<td>42</td>
<td>0,5</td>
<td>7</td>
<td>0,9</td>
</tr>
<tr>
<td>PTK</td>
<td>810</td>
<td>9,3</td>
<td>50</td>
<td>6,2</td>
</tr>
<tr>
<td>PW</td>
<td>120</td>
<td>1,4</td>
<td>6</td>
<td>0,7</td>
</tr>
<tr>
<td>VA</td>
<td>93</td>
<td>1,1</td>
<td>3</td>
<td>0,4</td>
</tr>
<tr>
<td>VM</td>
<td>305</td>
<td>3,5</td>
<td>7</td>
<td>0,9</td>
</tr>
<tr>
<td>VV</td>
<td>1422</td>
<td>16,3</td>
<td>174</td>
<td>21,5</td>
</tr>
<tr>
<td>XY</td>
<td>112</td>
<td>1,3</td>
<td>1</td>
<td>0,1</td>
</tr>
<tr>
<td></td>
<td>8706</td>
<td>808</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Tabelle 6.11: MK: Verteilung der Types, gruppiert, 1496 bis 1708

Vorkommen von Nomen und Verben (bezogen auf die Types) bei jedem Kind separat liefert folgende Ergebnisse:

- **AV**: Nomen 24,7%, Verben 24,3%
- **LEO**: Nomen 27,5%, Verben 23,7%
- **LUA**: Nomen 26,0%, Verben 22,5%
- **MK**: Nomen 35,0%, Verben 22,8%

KAPITEL 6. ANALYSE DER WORTARTEN

<table>
<thead>
<tr>
<th>POS-Tag</th>
<th>Tokens</th>
<th>prozentual %</th>
<th>Types</th>
<th>prozentual %</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADJ</td>
<td>4</td>
<td>1,9</td>
<td>3</td>
<td>3,4</td>
</tr>
<tr>
<td>ADV+PAV</td>
<td>22</td>
<td>10,6</td>
<td>9</td>
<td>10,1</td>
</tr>
<tr>
<td>APPR</td>
<td>13</td>
<td>6,3</td>
<td>7</td>
<td>7,9</td>
</tr>
<tr>
<td>ART</td>
<td>15</td>
<td>7,2</td>
<td>3</td>
<td>3,4</td>
</tr>
<tr>
<td>CARD</td>
<td>3</td>
<td>1,4</td>
<td>2</td>
<td>2,2</td>
</tr>
<tr>
<td>KON</td>
<td>15</td>
<td>7,2</td>
<td>4</td>
<td>4,5</td>
</tr>
<tr>
<td>N</td>
<td>22</td>
<td>10,6</td>
<td>16</td>
<td>18,0</td>
</tr>
<tr>
<td>PIDATS</td>
<td>4</td>
<td>1,9</td>
<td>3</td>
<td>3,4</td>
</tr>
<tr>
<td>PPER</td>
<td>36</td>
<td>17,3</td>
<td>4</td>
<td>4,5</td>
</tr>
<tr>
<td>PPOS</td>
<td>2</td>
<td>1,0</td>
<td>2</td>
<td>2,2</td>
</tr>
<tr>
<td>PTK</td>
<td>14</td>
<td>6,7</td>
<td>6</td>
<td>6,7</td>
</tr>
<tr>
<td>PW</td>
<td>6</td>
<td>2,9</td>
<td>4</td>
<td>4,5</td>
</tr>
<tr>
<td>VA</td>
<td>6</td>
<td>2,9</td>
<td>2</td>
<td>2,2</td>
</tr>
<tr>
<td>VM</td>
<td>10</td>
<td>4,8</td>
<td>3</td>
<td>3,4</td>
</tr>
<tr>
<td>VV</td>
<td>34</td>
<td>16,3</td>
<td>20</td>
<td>22,5</td>
</tr>
<tr>
<td>XY</td>
<td>2</td>
<td>1,0</td>
<td>1</td>
<td>1,1</td>
</tr>
</tbody>
</table>

208 89

Tabelle 6.12: AV: Verteilung der Types, gruppiert, 1709 bis 1983

von 35,0% personenabhängig ist (Daten aus Gruppe 1709_1983 zu MK liegen nicht vor). Ein erster Blick lässt vermuten, dass es sich nicht um die Eigenheiten einer Person handelt, denn im Zeitraum 1281_1495 machten Nomen (N) einen Anteil von 8,7% aller von MK geäußerten Tokens aus, wovon 27,1% Types waren. In der unten stehenden Übersicht sind alle von MK geäußerten Nomen in zwei Zeiträumen aufgeführt, geordnet nach Alphabet und versehen mit der Anzahl ihres Vorkommen.

6.3. INDIVIDUELLE UNTERSCHIEDE IN DEN ALTERSGRUPPEN

<table>
<thead>
<tr>
<th>POS-Tag</th>
<th>Tokens</th>
<th>prozentual %</th>
<th>Types</th>
<th>prozentual %</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADJ</td>
<td>35</td>
<td>6,2</td>
<td>13</td>
<td>7,8</td>
</tr>
<tr>
<td>ADV+PAV</td>
<td>69</td>
<td>12,2</td>
<td>24</td>
<td>14,5</td>
</tr>
<tr>
<td>APPR</td>
<td>9</td>
<td>1,6</td>
<td>5</td>
<td>3,0</td>
</tr>
<tr>
<td>ART</td>
<td>47</td>
<td>8,3</td>
<td>2</td>
<td>1,2</td>
</tr>
<tr>
<td>CARD</td>
<td>20</td>
<td>3,5</td>
<td>9</td>
<td>5,4</td>
</tr>
<tr>
<td>ITJ</td>
<td>11</td>
<td>1,9</td>
<td>4</td>
<td>2,4</td>
</tr>
<tr>
<td>KON</td>
<td>21</td>
<td>3,7</td>
<td>5</td>
<td>3,0</td>
</tr>
<tr>
<td>N</td>
<td>46</td>
<td>8,1</td>
<td>30</td>
<td>18,1</td>
</tr>
<tr>
<td>PIDATS</td>
<td>31</td>
<td>5,5</td>
<td>11</td>
<td>6,6</td>
</tr>
<tr>
<td>PPER</td>
<td>73</td>
<td>12,9</td>
<td>4</td>
<td>2,4</td>
</tr>
<tr>
<td>PPOS</td>
<td>1</td>
<td>0,2</td>
<td>1</td>
<td>0,6</td>
</tr>
<tr>
<td>PREL</td>
<td>1</td>
<td>0,2</td>
<td>1</td>
<td>0,6</td>
</tr>
<tr>
<td>PTK</td>
<td>57</td>
<td>10,1</td>
<td>14</td>
<td>8,4</td>
</tr>
<tr>
<td>PW</td>
<td>8</td>
<td>1,4</td>
<td>5</td>
<td>3,0</td>
</tr>
<tr>
<td>VA</td>
<td>7</td>
<td>1,2</td>
<td>1</td>
<td>0,6</td>
</tr>
<tr>
<td>VM</td>
<td>28</td>
<td>4,9</td>
<td>4</td>
<td>2,4</td>
</tr>
<tr>
<td>VV</td>
<td>89</td>
<td>15,7</td>
<td>32</td>
<td>19,3</td>
</tr>
<tr>
<td>XY</td>
<td>14</td>
<td>2,5</td>
<td>1</td>
<td>0,6</td>
</tr>
<tr>
<td></td>
<td>567</td>
<td>166</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| soldat (1), spielzeugtag (1), stapel (1), stein (1), strafkarte (2), sturm (2), tiger (1), tobeland (3), toilette (1), tor (1), trampolin (1), turm (8), uno (14), vater (1), verwandte (1), vorschüler (1), wasser (1), wolf (3) zahl (2), zaun (1), zigarette (1) |

## Kapitel 6. Analyse der Wortarten

<table>
<thead>
<tr>
<th>POS-Tag</th>
<th>Tokens</th>
<th>prozentual %</th>
<th>Types</th>
<th>prozentual %</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADJ</td>
<td>94</td>
<td>4,0</td>
<td>44</td>
<td>8,5</td>
</tr>
<tr>
<td>ADV+PAV</td>
<td>343</td>
<td>14,4</td>
<td>72</td>
<td>13,9</td>
</tr>
<tr>
<td>APPR</td>
<td>98</td>
<td>4,1</td>
<td>14</td>
<td>2,7</td>
</tr>
<tr>
<td>ART</td>
<td>183</td>
<td>7,7</td>
<td>3</td>
<td>0,6</td>
</tr>
<tr>
<td>CARD</td>
<td>83</td>
<td>3,5</td>
<td>13</td>
<td>2,5</td>
</tr>
<tr>
<td>ITJ</td>
<td>20</td>
<td>0,8</td>
<td>13</td>
<td>2,5</td>
</tr>
<tr>
<td>KON</td>
<td>171</td>
<td>7,2</td>
<td>13</td>
<td>2,5</td>
</tr>
<tr>
<td>N</td>
<td>237</td>
<td>10,0</td>
<td>146</td>
<td>28,2</td>
</tr>
<tr>
<td>PIDATS</td>
<td>88</td>
<td>3,7</td>
<td>15</td>
<td>2,9</td>
</tr>
<tr>
<td>PPER</td>
<td>294</td>
<td>12,4</td>
<td>7</td>
<td>1,4</td>
</tr>
<tr>
<td>PPOS</td>
<td>23</td>
<td>1,0</td>
<td>5</td>
<td>1,0</td>
</tr>
<tr>
<td>PREL</td>
<td>13</td>
<td>0,5</td>
<td>5</td>
<td>1,0</td>
</tr>
<tr>
<td>PTK</td>
<td>175</td>
<td>7,4</td>
<td>39</td>
<td>7,5</td>
</tr>
<tr>
<td>PW</td>
<td>25</td>
<td>1,1</td>
<td>5</td>
<td>1,0</td>
</tr>
<tr>
<td>VA</td>
<td>46</td>
<td>1,9</td>
<td>3</td>
<td>0,6</td>
</tr>
<tr>
<td>VM</td>
<td>86</td>
<td>3,6</td>
<td>7</td>
<td>1,4</td>
</tr>
<tr>
<td>VV</td>
<td>371</td>
<td>15,6</td>
<td>110</td>
<td>21,2</td>
</tr>
<tr>
<td>XY</td>
<td>28</td>
<td>1,2</td>
<td>3</td>
<td>0,6</td>
</tr>
<tr>
<td></td>
<td>2378</td>
<td></td>
<td>517</td>
<td></td>
</tr>
</tbody>
</table>

Tabelle 6.14: JS: Verteilung der Types, gruppiert, 1709 bis 1983

Tabelle 6.15: LAR: Verteilung der Types, gruppiert, 1709 bis 1983

(4), stunde (1), tank (4), tankwasser (1), tante (1), teil (1), tier (1), tisch (2), tischspiel (1), tor (1), totenkopf (2), training (3), tuch (1), tür (1), turm (1), uboot (1), uhr (1), uno (18), verstecke (1), vogel (8), vogelhäuschen (1), vogelnest (1), vorschau (1), waage (3), wachtmeister (1), waffe (2), wärme (4), wasser (1), weg (1), welt (1), winter (10), winterfell (1), wochenende (1), wohnung (2), wolke (2), wünscher (3), würfel (5), wüste (1), zahn (1), zauber (2), zauberfeuer (1), zauberlaser (1), zauberstab (4), zeichen (1), zeit (1), zoo (1), zoowärter (2)

Anhand der vorliegenden Inhalte kann zunächst nur schwer ein Grund für die hohe Type-Anzahl im Zeitraum 1496_1708 ausgemacht werden. Einige Lemmata treten mit einer leicht erhöhten Häufigkeit auf, wie zum Beispiel Feuer (10), Fußball (10), Hause (19), Kreuz (11), Krieg (8), Spiel (8), Uno (18), Winter (10). Diese Nomen kommen zwar zum Teil auch im Zeitraum 1281_1495 vor, jedoch selten derart häufig. Es fällt zudem auf, dass viele der sehr häufig auftretenden Nomen (aber auch seltener geäußerte) in einer Spielsituation geäußert wurden. Dies lässt ein Blick in die entsprechenden EXMARaLDA-Dateien sowie in die entsprechenden Kontextbeschreibungen vermuten. Man könnte nun annehmen, dass sich der Type-Anteil mit einem steigenden Token-Anteil aller Wörter ebenso erhöht und bei den meisten hier untersuchten Kindern ist dies auch der Fall. Es gibt jedoch Ausnahmen, die
<table>
<thead>
<tr>
<th>POS-Tag</th>
<th>Tokens</th>
<th>prozentual %</th>
<th>Types</th>
<th>prozentual %</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADJ</td>
<td>10</td>
<td>5,3</td>
<td>9</td>
<td>8,3</td>
</tr>
<tr>
<td>ADV+PAV</td>
<td>22</td>
<td>11,6</td>
<td>14</td>
<td>12,8</td>
</tr>
<tr>
<td>APPR</td>
<td>12</td>
<td>6,3</td>
<td>5</td>
<td>4,6</td>
</tr>
<tr>
<td>ART</td>
<td>15</td>
<td>7,9</td>
<td>3</td>
<td>2,8</td>
</tr>
<tr>
<td>CARD</td>
<td>5</td>
<td>2,6</td>
<td>3</td>
<td>2,8</td>
</tr>
<tr>
<td>KON</td>
<td>15</td>
<td>7,9</td>
<td>4</td>
<td>3,7</td>
</tr>
<tr>
<td>N</td>
<td>30</td>
<td>15,9</td>
<td>25</td>
<td>22,9</td>
</tr>
<tr>
<td>PIDATS</td>
<td>5</td>
<td>2,6</td>
<td>4</td>
<td>3,7</td>
</tr>
<tr>
<td>PPER</td>
<td>20</td>
<td>10,6</td>
<td>5</td>
<td>4,6</td>
</tr>
<tr>
<td>PPOS</td>
<td>3</td>
<td>1,6</td>
<td>3</td>
<td>2,8</td>
</tr>
<tr>
<td>PREL</td>
<td>1</td>
<td>0,5</td>
<td>1</td>
<td>0,9</td>
</tr>
<tr>
<td>PTK</td>
<td>5</td>
<td>2,6</td>
<td>5</td>
<td>4,6</td>
</tr>
<tr>
<td>PW</td>
<td>1</td>
<td>0,5</td>
<td>1</td>
<td>0,9</td>
</tr>
<tr>
<td>VA</td>
<td>6</td>
<td>3,2</td>
<td>2</td>
<td>1,8</td>
</tr>
<tr>
<td>VM</td>
<td>3</td>
<td>1,6</td>
<td>3</td>
<td>2,8</td>
</tr>
<tr>
<td>VV</td>
<td>33</td>
<td>17,5</td>
<td>21</td>
<td>19,3</td>
</tr>
<tr>
<td>XY</td>
<td>3</td>
<td>1,6</td>
<td>1</td>
<td>0,9</td>
</tr>
<tr>
<td></td>
<td>189</td>
<td></td>
<td>109</td>
<td></td>
</tr>
</tbody>
</table>

Tabelle 6.16: LEO: Verteilung der Types, gruppiert, 1709 bis 1983

diese These wieder ins Wanken bringen. So weist LL in Gruppe 1709_1983 einen recht hohen Nomenanteil (N) von 30,7% bezüglich der Types auf bei nur 1175 geäußerten Nomen. In Relation zur Gesamtanzahl der geäußerten Wörter weist kein anderes Kind einen derart hohen Type-Anteil auf. Dennoch können anhand der vorliegenden Daten keine Zusammenhänge zwischen Tokenanzahl, Alter der Kinder, Geschlecht und dem Type-Anteil bezüglich der Nomen hergestellt werden. Bestenfalls ist eine leichte Tendenz zu erkennen, nach der der Type-Anteil mit zunehmendem Alter abnimmt. Vielmehr macht es jedoch den Anschein, dass die unterschiedlichen Nomenanteile an verschiedene Situationen gebunden sind, was in Abschnitt 6.4 näher betrachtet werden soll.

unverständliche Ausdrücke etc.\textsuperscript{8} Templin kam zu folgenden Ergebnissen in Bezug auf das Vorkommen der Nomen und Verben (hinsichtlich der Types):

- 3 Jahre: 25,5\% Nomen, 23,4\% Verben
- 3;5 Jahre: 23,8\% Nomen, 24,1\% Verben
- 4 Jahre: 24,2\% Nomen, 24,0\% Verben
- 4;5 Jahre: 25\% Nomen, 23,3\% Verben


\textsuperscript{8}In dieser Arbeit gibt es für Eigennamen sowie für unverständliche Ausdrücke eigene Kategorien, NE und XY. Da es sich angesichts der Datenlage nur um recht wenige Tokens innerhalb der Kategorie NE handelt, soll trotzdem ein kurzer Vergleich mit Templins Ergebnissen vorgenommen werden.

„Nomen sind von Anfang an vorhanden und wachsen insbesondere während des zweiten Lebensjahres an. [...] Mit drei Jahren verwendet kein Kind mehr als 25% Nomen.“ ([Kauschke (1999)], S. 144)

und

„Mit 15 Monaten tauchen erstmals Verben auf, die dann erheblich ansteigen und mit 3 Jahren den stärksten Anteil am Lexikon ausmachen. ([Kauschke (1999)], S. 145)
6.3. **INDIVIDUELLE UNTERSCHIEDE IN DEN ALTERSGRUPPEN**

<table>
<thead>
<tr>
<th>POS-Tag</th>
<th>Tokens</th>
<th>prozentual %</th>
<th>Types</th>
<th>prozentual %</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADJ</td>
<td>193</td>
<td>4,8</td>
<td>42</td>
<td>8,0</td>
</tr>
<tr>
<td>ADV+PAV</td>
<td>513</td>
<td>12,7</td>
<td>64</td>
<td>12,2</td>
</tr>
<tr>
<td>APRR</td>
<td>114</td>
<td>2,8</td>
<td>14</td>
<td>2,7</td>
</tr>
<tr>
<td>ART</td>
<td>278</td>
<td>6,9</td>
<td>3</td>
<td>0,6</td>
</tr>
<tr>
<td>CARD</td>
<td>100</td>
<td>2,5</td>
<td>13</td>
<td>2,5</td>
</tr>
<tr>
<td>ITJ</td>
<td>16</td>
<td>0,4</td>
<td>6</td>
<td>1,1</td>
</tr>
<tr>
<td>KON</td>
<td>229</td>
<td>5,7</td>
<td>12</td>
<td>2,3</td>
</tr>
<tr>
<td>N</td>
<td>504</td>
<td>12,5</td>
<td>160</td>
<td>30,6</td>
</tr>
<tr>
<td>PIDATS</td>
<td>167</td>
<td>4,2</td>
<td>20</td>
<td>3,8</td>
</tr>
<tr>
<td>PPER</td>
<td>492</td>
<td>12,2</td>
<td>7</td>
<td>1,3</td>
</tr>
<tr>
<td>PPOS</td>
<td>44</td>
<td>1,1</td>
<td>4</td>
<td>0,8</td>
</tr>
<tr>
<td>PREL</td>
<td>14</td>
<td>0,3</td>
<td>6</td>
<td>1,1</td>
</tr>
<tr>
<td>PTK</td>
<td>397</td>
<td>9,9</td>
<td>46</td>
<td>8,8</td>
</tr>
<tr>
<td>PW</td>
<td>45</td>
<td>1,1</td>
<td>6</td>
<td>1,1</td>
</tr>
<tr>
<td>VA</td>
<td>42</td>
<td>1,0</td>
<td>3</td>
<td>0,6</td>
</tr>
<tr>
<td>VM</td>
<td>159</td>
<td>3,9</td>
<td>6</td>
<td>1,1</td>
</tr>
<tr>
<td>VV</td>
<td>677</td>
<td>16,8</td>
<td>110</td>
<td>21,0</td>
</tr>
<tr>
<td>XY</td>
<td>44</td>
<td>1,1</td>
<td>1</td>
<td>0,2</td>
</tr>
<tr>
<td></td>
<td>4028</td>
<td></td>
<td>523</td>
<td></td>
</tr>
</tbody>
</table>

Tabelle 6.19: MM: Verteilung der Types, gruppiert, 1709 bis 1983

Die erste Aussage trifft nicht zu, was die Ergebnisse dieser Arbeit zumindest hinsichtlich des Vorkommens der Types anbelangt. Allerdings können die Ergebnisse nicht exakt verglichen werden, da jeweils andere Methoden verwendet wurden. Unzutreffend ist an dieser Stelle auch die zweite Aussage, nämlich, dass Verben mit 3 Jahren den stärksten Anteil am Lexikon ausmachen. Vielmehr nehmen die Verben neben den Nomen sowohl hier als auch in Templins Ergebnissen einen großen Anteil am Lexikon bei Kindern im Vorschulalter ein (ca. 25% Nomen und ca. 22% bis 25% Verben) ([Templin (1957)], S. 101).

Tabelle 6.20: RD: Verteilung der Types, gruppiert, 1709 bis 1983

<table>
<thead>
<tr>
<th>POS-Tag</th>
<th>Tokens</th>
<th>prozentual %</th>
<th>Types</th>
<th>prozentual %</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADJ</td>
<td>33</td>
<td>4,4</td>
<td>22</td>
<td>8,5</td>
</tr>
<tr>
<td>ADV+PAV</td>
<td>76</td>
<td>10,1</td>
<td>33</td>
<td>12,7</td>
</tr>
<tr>
<td>APPR</td>
<td>33</td>
<td>4,4</td>
<td>10</td>
<td>3,8</td>
</tr>
<tr>
<td>ART</td>
<td>73</td>
<td>9,7</td>
<td>3</td>
<td>1,2</td>
</tr>
<tr>
<td>CARD</td>
<td>18</td>
<td>2,4</td>
<td>11</td>
<td>4,2</td>
</tr>
<tr>
<td>FM</td>
<td>2</td>
<td>0,3</td>
<td>2</td>
<td>0,8</td>
</tr>
<tr>
<td>ITJ</td>
<td>2</td>
<td>0,3</td>
<td>2</td>
<td>0,8</td>
</tr>
<tr>
<td>KON</td>
<td>30</td>
<td>4,0</td>
<td>6</td>
<td>2,3</td>
</tr>
<tr>
<td>N</td>
<td>95</td>
<td>12,6</td>
<td>61</td>
<td>23,6</td>
</tr>
<tr>
<td>PIDATS</td>
<td>32</td>
<td>4,2</td>
<td>8</td>
<td>3,1</td>
</tr>
<tr>
<td>PPER</td>
<td>88</td>
<td>11,7</td>
<td>5</td>
<td>1,9</td>
</tr>
<tr>
<td>PPOS</td>
<td>12</td>
<td>1,6</td>
<td>2</td>
<td>0,8</td>
</tr>
<tr>
<td>PREL</td>
<td>1</td>
<td>0,1</td>
<td>1</td>
<td>0,4</td>
</tr>
<tr>
<td>PTK</td>
<td>73</td>
<td>9,7</td>
<td>23</td>
<td>8,8</td>
</tr>
<tr>
<td>PW</td>
<td>12</td>
<td>1,6</td>
<td>5</td>
<td>1,9</td>
</tr>
<tr>
<td>VA</td>
<td>21</td>
<td>2,8</td>
<td>3</td>
<td>1,2</td>
</tr>
<tr>
<td>VM</td>
<td>19</td>
<td>2,5</td>
<td>5</td>
<td>1,9</td>
</tr>
<tr>
<td>VV</td>
<td>128</td>
<td>17,0</td>
<td>56</td>
<td>21,6</td>
</tr>
<tr>
<td>XY</td>
<td>6</td>
<td>0,8</td>
<td>1</td>
<td>0,4</td>
</tr>
<tr>
<td></td>
<td>754</td>
<td>259</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

6.3. **INDIVIDUELLE UNTERSCHIEDE IN DEN ALTERSGRUPPEN**

<table>
<thead>
<tr>
<th>POS-Tag</th>
<th>Tokens</th>
<th>prozentual %</th>
<th>Types</th>
<th>prozentual %</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADJ</td>
<td>33</td>
<td>6,6</td>
<td>16</td>
<td>9,3</td>
</tr>
<tr>
<td>ADV+PAV</td>
<td>68</td>
<td>13,5</td>
<td>23</td>
<td>13,3</td>
</tr>
<tr>
<td>APPR</td>
<td>9</td>
<td>1,8</td>
<td>4</td>
<td>2,3</td>
</tr>
<tr>
<td>ART</td>
<td>20</td>
<td>4,0</td>
<td>3</td>
<td>1,7</td>
</tr>
<tr>
<td>CARD</td>
<td>8</td>
<td>1,6</td>
<td>5</td>
<td>2,9</td>
</tr>
<tr>
<td>ITJ</td>
<td>2</td>
<td>0,4</td>
<td>2</td>
<td>1,2</td>
</tr>
<tr>
<td>KON</td>
<td>20</td>
<td>4,0</td>
<td>7</td>
<td>4,0</td>
</tr>
<tr>
<td>N</td>
<td>45</td>
<td>8,9</td>
<td>29</td>
<td>16,8</td>
</tr>
<tr>
<td>PIDATS</td>
<td>28</td>
<td>5,6</td>
<td>10</td>
<td>5,8</td>
</tr>
<tr>
<td>PPER</td>
<td>69</td>
<td>13,7</td>
<td>6</td>
<td>3,5</td>
</tr>
<tr>
<td>PPOS</td>
<td>6</td>
<td>1,2</td>
<td>2</td>
<td>1,2</td>
</tr>
<tr>
<td>PTK</td>
<td>71</td>
<td>14,1</td>
<td>23</td>
<td>13,3</td>
</tr>
<tr>
<td>PW</td>
<td>7</td>
<td>1,4</td>
<td>4</td>
<td>2,3</td>
</tr>
<tr>
<td>VA</td>
<td>6</td>
<td>1,2</td>
<td>2</td>
<td>1,2</td>
</tr>
<tr>
<td>VM</td>
<td>22</td>
<td>4,4</td>
<td>5</td>
<td>2,9</td>
</tr>
<tr>
<td>VV</td>
<td>84</td>
<td>16,7</td>
<td>31</td>
<td>17,9</td>
</tr>
<tr>
<td>XY</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>0,6</td>
</tr>
<tr>
<td></td>
<td>503</td>
<td>173</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Tabelle 6.21: SO: Verteilung der Types, gruppiert, 1709 bis 1983

ist, so Kegel ([Kegel (1987)], S. 49). An dieser Stelle wird deutlich, dass vor allem das Untersuchen des passiven Wortschatzes eine Herausforderung darstellt, da nur sehr schwer bewertet werden kann, wann und ob ein Kind ein Wort tatsächlich verstanden hat. Bisher ist mir keine Arbeit bekannt, die die gesamte Struktur des Wortschatzes, sei es bei Kindern oder bei Erwachsenen, abschließend in einer experimentellen Studie untersucht hat. Wie bereits festgestellt wurde, ist dies nur schwer bis gar nicht möglich. Eine Untersuchung, die den gesamten aktiven und passiven Wortschatz eines Individuums erfassen will, setzt unterschiedliche Schritte voraus, die zum jetzigen Zeitpunkt nicht erfüllbar sind. Es müsste zunächst der gesamte passive Wortschatz erfasst werden, was weder experimentell noch durch Beobachtungen, Befragungen, Wortlisten etc. möglich wäre. Selbst wenn es gelingen würde, zumindest den aktuellen Wortschatz zu einem gegebenen Zeitpunkt eines Individuums zu erfassen, würde doch der passive Wortschatz niemals erfasst werden können. Doch die Annahme durch Fragebögen zum Wortverständnis und durch andere Tests sei es möglich den passiven Wortschatz zu ermitteln, wirft einige Fragen auf. Rechnerisch gesehen könnte so vielleicht ein Großteil der verstandenen Lexeme erreicht werden, aber was ist mit Wörtern wie *Feinlöwe* oder *wurstig*? Diese Wörter gibt es - laut Duden - nicht im Wortschatz der deutschen Sprache ([Dudenredaktion (2004)])

Dennoch ist die Bedeutung je nach Kontext verständlich. Und genau hier liegt

6.3.5 Zusammenfassung

Durch eine Betrachtung der Wortarten gemäß der STTS (hier vorliegend in gruppiert Form) wurde zunächst ein Überblick über die Verwendung und Verteilung der Wortarten bei jedem Kind in jeder Altersgruppe geschaffen. Auffällig war, dass die Verteilung in den Gruppen sowie bei einem Vergleich aller Kinder (egal welchen Alters) sehr ähnliche prozentuale Anteile aufweist. Dabei ist es egal, ob ein Kind im gesamten Aufnahmezeitraum sehr viele oder nur sehr wenige Tokens zu den Aufnahmen beisteuerte. Die wenigsten Wörter finden sich bei LEO in Gruppe 3 mit 189 Tokens. Die meisten Wörter
weiteren Verlauf der Sprache genutzt werden kann ([Kauschke (1999)], S. 153), kann insbesondere erweitert werden um folgende These:

*Der Spracherwerb im 3. Lebensjahr kann als ein prognostisches Mittel für den weiteren Verlauf der Sprachentwicklung verwendet werden.*

Vor allem die Verteilung der Verben, sowohl in Bezug auf die Types als auch auf die Tokens ist vergleichbar mit 36 Monate alten Kindern und bleibt über den Zeitraum von 3;5 bis 5;5 Jahren nahezu unverändert erhalten.

### 6.4 Analyse der ersten einhundert Wörter je Kind

In diesem Abschnitt soll untersucht werden, ob die geäußerten Wörter jedes Kindes, unabhängig vom Alter, den vorgehenden Ergebnissen entsprechen. Diese Analyse kann deshalb vorgenommen werden, weil die Ergebnisse in Abschnitt 6.3 andeuten, dass es ab einem Alter von 3 Jahren keine großen Unterschiede in der Verteilung der Wortarten gibt bzw., dass geringfügige Unterschiede (vor allem im Hinblick auf die Types) von der individuellen Situation, nicht aber vom Alter abhängen. Es wurde lediglich festgestellt, dass es mit zunehmendem Alter vermutlich zu einer Reduktion des Nomenanteils bezogen auf die Types kommt, da die Anteile in der dritten Altersgruppe bei fünf Kindern (AV: 18,0%, JK: 18,1%, LAR: 18,8%, LUA: 16,7%, SO: 16,8%) unter 19% lagen im Gegensatz zu den Vergleichsgruppen. Dieser leichte Abwärtstrend müsste weiter untersucht werden, um bestätigt werden zu können. Denn insbesondere bei diesen fünf Kindern waren auch die Tokens-Anteile niedriger als bei Kindern, die besonders hohe prozentuale Nomen-Anteile aufweisen, so dass nicht ausgeschlossen werden kann, dass die Tokens-Anzahl einen entscheidenden Einfluss auf die Types-Anzahl hat. Im Folgenden sind die ersten einhundert gesprochenen Wörter (Tokens) jedes Kindes aufgeführt. Da diese Tokens aus ihrem ursprünglichen Kontext extrahiert wurden, sind diese fernab der cGAT-Transkriptionsregeln, wie sie in den FOLKER-Dateien angewendet wurden, aufgeschrieben. Es wird lediglich das Ende einer Äußerung durch // gekennzeichnet. Alle Wörter (sowie die dazugehörigen Lemmata, POS-Tags und Bedeutungskategorien) befinden sich in der SQLite3-Datenbank des CA sowie in gesonderten Dateien. Des Weiteren sind alle Situationsbeschreibungen der einzelnen Aufnahmedaten in einer Tabelle in Abschnitt 4.1.1 zusammengefasst.

---

9Hierzu zählen sowohl die erstellten EXMARaLDA-Dateien sowie alle Ausgabedateien im CSV-Format, die im Zuge der Analyse unter der Anwendung des CA entstanden sind.
Detailliertere Beschreibungen befinden sich zudem in den Protokollen, die während bzw. kurz nach den Aufnahmesituationen angefertigt wurden.


- AV: Bei den ersten einhundert geäußerten Wörtern handelt es sich eindeutig um eine Erzählsls situation mit der Interviewerin. Lediglich ab dem 74. Wort ist eine angedeutete Spielsituation erkennbar. AV spielt jedoch nicht selber, sondern fordert die Interviewerin auf, Karten eines Kartenspiels aufzudecken.
• LEO: Bis zu Wort 33 singt LEO ein Lied ("Ich gehe mit meiner Laterne"). Darauffolgend handelt es sich eindeutig um eine Erzählsituation mit der Interviewerin. LEO erzählt, dass er schon gut malen und mit einer Schere schneiden kann. Außerdem erläutert er, was ein Geheimnis ist.


• LL: Die ersten einhundert geäußerten Wörter von LL können sowohl einer Erzähl- als auch einer Spielsituation zugeordnet werden. Die gesamte Situation findet während eines Spiels statt, wobei LL in der ersten Hälfte mit der Interviewerin spricht. Später (ab Wort 57) beteiligt sich LL auch am Spiel, was durch die Verwendung von Nomen, die als Elemente zum Spiel gehören, deutlich wird (mehrfache Äußerung der Nomen Birne, Traktor, Schwein, Schmetterling, die auf den Karten eines Memory-Spiels abgebildet sind).

• LUA: LUA befindet sich mit der Interviewerin in einer Erzählsituation. Auffällig sind gehäufte Wiederholungen in den hier geäußerten Passagen. Ob diese einen Einfluss auf die Anteile und schließlich auf die Verteilung der Wortarten haben, kann nicht abschließend beurteilt werden. Da die Wiederholungen jedoch nicht standardmäßig bei LUA auftreten (im Vergleich zu allen von LUA geäußerten Tokens), ist davon auszugehen, dass diese keinen Einfluss auf die Verteilung der Wortarten haben.

• MA: Die ersten einhundert geäußerten Wörter gehören zu einer Erzählsituation zwischen MA und der Interviewerin. MA erzählt, was sie gleich malen möchte und beschreibt eine Situation, in der ein Vogel draußen am Fenster sitzt.

• MK: MK erzählt von den Erlebnissen seines Geburtstages. Es handelt sich demnach um eine Erzählsituation.

• MM: MM befindet sich zwar gerade in einem (Karten-)Spiel. Es geht jedoch vorwiegend darum, wer schneller war und wer die Karten verteilt.


• SO: Dem Inhalt der ersten einhundert Tokens nach zu urteilen, befindet sich SO in einem Spiel zusammen mit anderen Kindern. Da es jedoch zu einem Streit darüber kommt, welches Kind wie viele Karten bekommt und welches Kind wann an der Reihe ist, entsteht gar kein Spiel. Demnach kann die Situation eher einer Erzählsituation zugeordnet werden.


\textsuperscript{11}Die hier aufgeführten prozentualen Anteile beziehen sich auf die ersten einhundert syntaktischen Wörter je Kind. Die Werte sind nicht gesondert in einer Tabelle aufgeführt.
6.4. ANALYSE DER ERSTEN EINHUNDERT WÖRTER JE KIND

Befanden sich Kinder in einer sogenannten Erzählsituation, verwendeten sie (zumindest in der vorliegenden Arbeit) tendenziell weniger Nomen, da diese scheinbar zur Beschreibung von Situationen weniger notwendig sind.


6.5 Die Wortartenverteilung bei Erwachsenen


Das Vorkommen von Nomen und Verben bezogen auf die Tokens ergab

Im Folgenden sollen nun die Daten eines jeden einzelnen Kindes betrachtet werden. In Abschnitt 6.3 wurden die Anteile aller Wortarten je Kind und Altersgruppe in Bezug auf die Types und die Tokens vermerkt. Zunächst zu den Types: In Gruppe 1 kamen Nomen mit einem Anteil von 24,7% (MA) bzw. 27,1% (MK) und Verben mit 22,4% (MA) bzw. 25,0% (MK) vor. In Gruppe 2
KAPITEL 6. ANALYSE DER WORTARTEN

belief sich das Vorkommen der Nomen auf 24,7% bis 35,0% bei vier Kindern, wobei 35,0% nur bei MK auftraten. Verben traten mit einer Häufigkeit von 22,5% bis 24,3% auf. Gruppe 3 lieferte Anteile von 16,7% bis 30,7% Nomen bei insgesamt 10 Kindern, wobei rund 17% zweimal (LUA und SO) und rund 18% dreimal (AV, LAR und JK) auftraten. Die restlichen Nomenanteile bewegten sich zwischen 22,9% und 30,7%. Verben kamen bei diesen 10 Kindern mit 21,0% bis 28,1% vor. An dieser Stelle ist es nicht einfach, einen einheitlichen Bezug herzustellen zu den Daten der Erwachsenen. Auffällig ist aber, dass unter Ausschluss des 35,0%-Vorkommens der unterschiedlichen Nomen bei MK in Gruppe 2, der Anteil der Nomen kontinuierlich niedriger ist. In Gruppe 3 gab es gar zwei Kinder, die einen Anteil von rund 17% aufwiesen und drei Kinder mit einem Anteil von unter 19% Nomen. Im Vergleich zum Vorkommen von 19% bei den Erwachsenen ist das erstaunlich wenig. Man kann also annehmen, dass sich die Verwendung verschiedener Nomen mit zunehmenden Alter langsam reduziert. Ähnlich verhält es sich bei den Verben. In Gruppe 3 verwendete kein Kind mehr als 28,1% Verben bezogen auf die Types. Das ist im Vergleich zu 18,5% bei den Erwachsenen zwar erheblich mehr. Die meisten Kinder wiesen jedoch einen Anteil von 22% bis 24% auf, während der Wert von 28,1% bei nur einem Kind vorkam. In Gruppe 1 war der geringste Anteil 22,4%, in Gruppe 2 lag er bei 22,5%. Man könnte also auch hier annehmen (unter Ausschluss der 28,1% in Gruppe 3), dass die Verwendung von Verben bezogen auf die Types in einem Alterszeitraum von 3;5 bis 5;5 Jahren kontinuierlich niedrig ist und durchaus mit den von Kauschke (2007) ermittelten Daten von 16,3% vergleichbar ist. Um zu prüfen, ob sich die Types-Anteile bezüglich der Verben mit zunehmendem Alter weiter reduzieren, müssten weiterführende Analysen gemacht werden, die über das sechste Lebensjahr von Kindern hinausgehen.

In Bezug auf die Tokens sieht ein Vergleich der Daten der Kinder zu den Erwachsenen etwas anders aus. Wie weiter oben dargestellt, kamen Verben bei allen Kindern verteilt über alle Altersgruppen mit einem Anteil von 21,0% bis 28,1% bezogen auf die Types vor. Hinsichtlich der Tokens sind es 20,4% bis 24,0%. Im Vergleich zu den Daten der Erwachsenen ist der Anteil zwar höher, erstaunlicherweise erscheint er aber stabil und es gibt keine großen Unterschiede zwischen den Types- und Tokens-Anteilen. Bei den Nomen reichte die Spanne bei den Types von 16,7% bis 35,0%, bei den Tokens von 5,9% bis 15,9%. Dieser Unterschied ist im Vergleich zu den Daten der Erwachsenen zum Teil enorm. Hier steht ein Typeanteil von 19%, einem Tokenanteil von 9% jeweils bezüglich der Nomen gegenüber. Erstaunlich ist hingegen, dass der Anteil der Tokens bei den hier untersuchten Kindern mit 5,9% bis 15,9% ähnlich niedrig ist wie jener der Erwachsenen.
Kapitel 7

Der Grundwortschatz semantisch betrachtet


7.1. DIE BEDEUTUNGSKATEGORIEN


7.1 Die Bedeutungskategorien

Die folgend aufgeführten Kriterien, nach denen ein Lemma einer bestimmten Bedeutungskategorie zugeordnet ist, sind angelehnt an die Regeln der Dudengrammatik ([Dudenredaktion (2009)]). Die Kategorien sind

¹Aufgrund des sehr großen Umfangs des erstellten Lexikons und der dazugehörigen Excel-Dateien ist dieses im Anhang dieser Arbeit zu finden.
KAPITEL 7. DER GRUNDWORTSCHATZ SEMANTISCH BETRACHTET

nicht alphabetisch geordnet, sondern entsprechen in ihrer Reihenfolge weitestgehend den Kategorien der STTS und ihrer dortigen Anordnung (Tabelle 3.2).

7.1.1 Adjektive

Unter grammatischen Gesichtspunkten sind Adjektive Wörter, die flektiert werden können. Dementsprechend sind sie veränderbar nach Kasus, Numerus und Genus. Im Hinblick auf semantische Kriterien unterscheidet der Duden folgende Klassen (Regel 459):

- qualifizierende Adjektive: Farbe (rot, grün), Form (eckig, rund), Geschmack (süß, bitter), Oberfläche (rau, glatt), Temperatur (warm, kalt), Ästhetik (schön, hässlich), Moral (gut, böse), Intellekt (klug, witzig), Räumliche Dimension (hoch, breit), Zeitliche Dimension (früh, spät), Wahrheitsgehalt (wahrscheinlich, angeblich)

  Beispiel (1): Komm, wir malen ein so großes Pferd, ok? (024_MK.exb)

  Beispiel (2): Sieht lustig aus, oder? (039_MA.exb)

- relationale Adjektive: Geografie (afrikanisch, asiatisch), Staat/Volk/Sprache (englisch, französisch), Religion (katholisch, islamisch), Epoche (römisch, mittelalterlich), Beruf (ärztlich), Bereich (wirtschaftlich, technisch), Stoff (golden, hölzern), Zeitpunkt (heutig, gestrig), Räumliche Lage (vordere, linke)

  Beispiel (3): Hier, chinesisches Essen. 'Ne kleine Portion. (041_LEO.exb)

  Beispiel (4): Hier ist der kleine Käsepfannkuchen mit einem schweizer Käse. (041_LEO.exb)


  Beispiel (5): das dritte Holzstück (013_JS.exb)

  Beispiel (6): Ich habe null, neun, sechs gemacht. (033_JK.exb)

- adjektivisch gebrauchte Partizipien (Regel 458): z. B. fliegend in fliegender Fisch

  Beispiel (7): ein fliegender Fisch. (048_LEO.exb)
7.1. DIE BEDEUTUNGSKATEGORIEN

In der hier getätigten Analyse kommen qualifizierende Adjektive (ad-qual), quantifizierende Adjektive (ad-quant), relative Adjektive (ad-rel) und adjektivisch gebrauchte Partizipien (ad-part) vor. Bei den quantitativen Adjektiven handelt es sich um alle zuvor mit CARD getaggen Zahlwörter. Diese werden im Folgenden mit ad-quant bezeichnet, da auch die Dudengrammatik dies als eine Vorgehensweise zulässt (Regel 509).

7.1.2 Adverbien

Die Dudengrammatik unterscheidet vier große Gruppen von Adverbien (ab Regel 837):


  Beispiel (8): Ich habe **hier** einen Ritter. (023_LUA.exb)

  Beispiel (9): **Hier unten** läuft das ganz groß raus. (033_MK.exb)

- **Temporaladverb** (Regel 850): situiert ein Geschehen, ein Objekt o. ä. in der Zeit oder bezeichnet eine temporale Beziehung hinsichtlich des Zeitpunkts (z. B. jetzt, nun, heute, gestern, morgen, eben, neulich, morgens, mittags, dann), der Dauer (z. B. immer, stets, lange, bisher), der Wiederkehr bzw. Wiederholung (z. B. manchmal, nochmal, montags, dienstags, nachmittags) und der Zeitbewertung bzw. -erwartung (z. B. bereits, schon, noch).

  Beispiel (10): Ich schlafe **jetzt** bei den Eltern. (011_MM-exb)

  Beispiel (11): Ich habe **gerade** nicht geklingelt. (020_LAR.exb)

- **Modaladverb** (Regel 852): gibt sowohl die Qualität (Art und Weise) als auch die Quantität (Menge, Ausmaß) an. Beispiele für qualitative Modaladverbien sind: so, genauso, ebenfalls, anders, nebenbei, gern, unversehens. Beispiele für quantitative Modaladverbien sind: sehr, größtenteils, einigermaßen, halbwegs, teilweise, allein.

  Beispiel (12): Da ist **nur** ein bisschen von mir drin. (040_MK.exb)

  Beispiel (13): Kannst auch **so rum drehen**. (046_LUA.exb)

Beispiel (14): Sonst können wir das nicht spielen. (050_JS.exb)

Beispiel (15): Also ich habe nicht über die Linie gemalt. (054_MK.exb)

Eine weitere und 5. Gruppe betrifft die sogenannten Kommentaradverbien (Regel 868), die sich funktional teilweise wie Abtönungspartikeln verhalten. Kommentaradverbien beziehen sich immer auf den ganzen Satz, weshalb sie auch Satzadverbien genannt werden (leider, allerdings, natürlich, selbstverständlich, vielleicht).

Beispiel (16): Das kannst du leider nicht sein. (049_MK.exb)

Beispiel (17): Aber vielleicht sind sie ja jetzt da. (050_LUA.exb)


Beispiel (18): Außerdem spielen wir, wenn uns langweilig wird. (033_MK.exb)

Beispiel (19): Was steht da, da, da_rin. (048_AV.exb)

Weitere Adverbien mit besonderen Funktionen sind interrogative Adverbien (Regel 856) und relative Adverbien (Regel 857), bei denen es sich um W-Fragewörter handelt. Da diese bereits in den STTS als Relativ- und Interrogativpronomen bezeichnet wurden, werden sie auch im Folgenden als Pronomen beschrieben. In der folgenden Übersicht aller Adverbien wird demnach unterschieden zwischen Lokaladverbien (adv-lok), Temporaladverbien (adv-temp), Modaladverbien (adv-mod), Kausaladverbien (adv-kaus), Kommentaradverbien (adv-komm) sowie den zu den Pronomen zugeordneten Pronominaladverbien (pav).
7.1. Präpositionen

Präpositionen stehen vor ihrem Bezugswort (Regel 895), z. B. nach Hause. Daneben existieren Postpositionen (meiner Meinung nach), die dem Bezugswort nachgestellt sind sowie Zirkumpositionen (um Himmels willen), die das Bezugswort umrahmen. Sowohl Post- als auch Zirkumpositionen kommen in den Daten der vorliegenden Arbeit nicht vor und wurden deshalb nicht weiter beachtet. Laut Dudengrammatik gibt es etwa 20 Präpositionen, die im Deutschen besonders häufig vorkommen. Dazu zählen:

in, mit, von, an, auf, zu, bei, nach, um, für, aus, vor, über, durch, unter, gegen, hinter, bis, neben, zwischen

Im Hinblick auf semantische Aspekte können fünf Gruppen von Präpositionen unterschieden werden:

- **Lokale Präposition** zur Bezeichnung des Raumes, der Lage, der Richtung (Regel 905): ab, an, auf, aus, außer, außerhalb, bei, bis, durch, nach, neben, von, vor, zu

**Beispiel** (20): Eine Babykuh hat bei Mama getrunken. (019_RD.exb)

**Beispiel** (21): auf dem Spielplatz, glaube ich (020_SO.exb)

- **Temporale Präposition** zur Bezeichnung des Zeitpunkts und der Dauer (Regel 906): ab, an, auf, bei, binnen, für, gegen, in, innerhalb, mit, nach, seit, über, um, unter, vor, während, zu, zwischen

**Beispiel** (22): Im Moment brauchen wir die nicht. (026_LUA.exb)

**Beispiel** (23): Das habe ich auch schon mal vor langer Zeit gemacht. (053_AV.exb)

- **Modale Präposition** zur Bezeichnung der Art und Weise (Regel 907): auf, aus, außer, bei, bis, an, für, gegen, mit, ohne, unter, von, wider, zu

**Beispiel** (24): Ich gehe mit meiner Laterne. (036_LEO.exb)

**Beispiel** (25): In der Zeit wird Papier für euch geliefert. (033_RD.exb)

- **Kausale Präposition** zur Bezeichnung des Grundes, des Anlasses, der Einräumung, der Einschränkung und des Zwecks (Regel 908): auf, aus, bei, durch, für, kraft, laut, mangels, mit, mittels, nach, trotz, über, um, unter, von, wegen

**Beispiel** (26): Vielleicht ist es wegen dem? (038_LL.exb)

**Beispiel** (27): Wegen dir spiele ich nicht mehr mit. (048_JS.exb)
KAPITEL 7. DER GRUNDWORTSCHATZ SEMANTISCH BETRACHTET

- **Neutrale (leere) Präposition** als Verbindungsglied ohne eigene Bedeutung (Regel 908): hoffen auf, leiden unter, bitten um, Achtung vor, Garantie für, stolz auf, froh über, fähig zu

Beispiel (28): Hier werden Äpfel zu Mehl gepresst. (033_JS.exb)

Beispiel (29): Ich bin am Verlieren. (047_RD.exb)

Die folgenden Bezeichnungen wurden für die semantische Analyse der Präpositionen verwendet: Lokale Präpositionen (pr-lok), Temporale Präpositionen (pr-temp), Modale Präpositionen (pr-mod), Kausale Präpositionen (pr-kaus) und Neutrale Präpositionen (pr-neutr).

7.1.4 Artikel

In der vorangegangenen Wortartenanalyse wurde der Artikel nicht nach bestimmt (der, die, das) und unbestimmt (ein/eine) unterschieden, sondern stets mit ART bezeichnet. Aufgrund des zahlreichen Vorkommens von Artikeln aber auch der Schwierigkeit einer semantischen Einordnung sollen Artikel an dieser Stelle nicht weiter gruppiert werden als bisher. **Bestimmte** und **unbestimmte Artikel** werden dementsprechend als Artikel mit dem Kürzel **art** bezeichnet.

Beispiel (30): Das ist ja auch nicht eine Schwimmhalle, das ist ein Schwimmbad. (019_MK.exb)

Beispiel (31): Dann kommst du immer mit dem Baby hier her zu uns. (057_MK.exb)

7.1.5 Fremdsprachliches Material

**Fremdsprachliches Material** kam nur sehr selten in den Daten dieser Arbeit vor. Da es sich zudem nur um die Nennung einzelner Wörter (oder Zahlen) handelt, sollen auch diese Lemmata nicht weiter gruppiert werden. Sie werden in der vorliegenden semantischen Betrachtung mit **fm** bezeichnet.

Beispiel (32): ... habe ich jetzt **Ice Age** drei ausgeliehen. (046_MK.exb)

7.1.6 Interjektionen

Die **Interjektionen** werden von der Dudengrammatik unter dem Eintrag **Partikeln** behandelt (Regel 887). Dies ist unter semantischen Aspekten sicher sinnvoll, da die Interjektion als Ausdruckspartikel eine gesonderte Gruppe ausmacht. Da in der hier durchgeführten Analyse mittels der STTS (Kapitel 6) die Interjektionen von den Partikeln gesondert betrachtet und analysiert
wurden, werden sie auch an dieser Stelle gesondert aufgeführt. Unter semantischen Gesichtspunkten zählen sie aber zu den Partikeln und sind somit, wie schon erwähnt, eine Ausdruckspartikel zum Ausdruck spontaner, reaktiver Emotionen und Bewertungen. Beispiele: pfui, hurra, huch, igitt, juhu, oje, pst!

Die Interjektionen werden in ihrer semantischen Kategorie mit itj gekennzeichnet.

Beispiel (33): pfui deibel (039_LEO.exb)

7.1.7 Konjunktionen und Subjunktionen

Bei den Junktionen wurde bisher unterschieden zwischen Konjunktionen (KON), Vergleichskonjunktionen (KOKOM) und unterordnenden Konjunktionen bzw. Subjunktionen (KOUS). Die Dudengrammatik unterscheidet, wenn es um die Funktion und Bedeutung geht, zwei große Gruppen, die wiederum weiter untergliedert sind: Konjunktionen und Subjunktionen.

Die Bezeichnung in Klammern steht für das Kürzel, das der jeweiligen Junktion in der semantischen Analyse entspricht.

Konjunktionen (Regeln 934 bis 940)

Die Konjunktionen werden unterschieden in:

- additive Konjunktionen (kon-add) (Regel 935): und, plus
  
Beispiel (34): Ich habe einen riesigen Schnabel und damit töte ich alle Fische. (042_MK.exb)

- alternative Konjunktionen (kon-alt) (Regel 936): oder
  
Beispiel (35): Im Sommer kann ich beim Halbfinale oder beim richtigen Finale zugucken. (058(JS).exb)

- adversative und konzessive Konjunktionen (kon-adv) (Regel 937): aber, doch, jedoch, sondern, bloß
  
Beispiel (36): (ich habe) drei (NN), aber dafür hast du mehr (036_LAR.exb)

- spezifizierende Konjunktionen (kon-spez) (Regel 938): außer
  
Beispiel (37): hier darf keiner durch, außer den ... (024_MK.exb)

- kausale Konjunktionen (kon-kaus) (Regel 939): denn, weil, da
  
Beispiel (38): aber die beiden Zahlen stimmen nicht, denn die Zwei ist zu klein (030_MK.exb)

- vergleichende Konjunktionen (kon-vgl) (Regel 940): wie, als
  
Beispiel (39): noch größer wie der meine (039_MA.exb)

Grammatikalisch korrekt müsste es an dieser Stelle heißen „noch größer als meiner“.
Subjunktionen (Regeln 941 bis 952)

Die Subjunktionen werden unterschieden in:

- **neutrale Subjunktionen** (*sub-neutr*) (Regel 942): *dass*, *ob*
  
  **Beispiel** (40): *Der will immer bestimmen, *dass* der Punkt so lang geht.* (044_MK.exb)

- **temporale Subjunktionen** (*sub-temp*) (Regel 943): *als*, *wenn*, *indem*, *nachdem*, *bis*, *bevor*
  
  **Beispiel** (41): *Haben wir gesehen, *als* wir mit Mama wieder rausgegangen sind.* (054_AV.exb)

- **konditionale Subjunktionen** (*sub-kond*) (Regel 944): *wenn*, *falls*, *ob*
  
  **Beispiel** (42): *Wenn es raus geht, dann pack’ es da rein.* (023_LUA.exb)

- **adversative Subjunktionen** zur Bezeichnung des Gegensatzes (*sub-adv*) (Regel 945): *anstatt*
  
  Adversative Subjunktionen kommen in den Daten dieser Arbeit nicht vor.

- **restriktive Subjunktionen** zur Bezeichnung der Einschränkung (*sub-restr*) (Regel 946): *außer dass*
  
  Restriktive Subjunktionen kommen in den Daten dieser Arbeit nicht vor.

- **modal-instrumentale Subjunktionen** zur Bezeichnung des Mittels, um das im Hauptsatz genannte Ziel zu erreichen (*sub-mod-instr*) (Regel 947): *indem, ohne dass*
  
  **Beispiel** (43): *Wenn man fertig getankt hat, kann man gleich wieder losfliegen *ohne* zu bezahlen.* (046_MK.exb)

- **Subjunktionen zum Ausdruck eines Vergleichs** (*sub-vgl*) (Regel 948): *als*, *wie*
  
  Subjunktionen zum Ausdruck eines Vergleich kommen in den Daten dieser Arbeit nicht vor.

- **kausale Subjunktionen** (*sub-kaus*) (Regel 949): *weil*, *wo*, *umso mehr/weniger als*
7.1. DIE BEDEUTUNGSKATEGORIEN

Beispiel (44): ja, weil ich krank bin (033_MK.exb)

- **konsekutive Subjunktionen** (*sub-kons*) (Regel 950): sodass, (so) - dass
  
  Konsekutive Subjunktionen kommen in den Daten dieser Arbeit nicht vor.

- **finale Subjunktion** (*sub-fin*) (Regel 951): damit, dass, um zu

  Beispiel (45): Ja! Und weit, damit ich reinpasse. (019_JS.exb)

- **konzessive Subjunktionen** (*sub-konz*) (Regel 952): obwohl, auch wenn

  Beispiel (46): Wir beide sind keine Anfänger, **obwohl** das nicht stimmt. (058_JS.exb)

7.1.8 **Substantive**

Die Substantive als Inhaltskategorie ließen bereits bei einem ersten Blick auf die geäußerten Tokens vermuten, dass es nicht leicht sein wird, Gruppierungen zu schaffen, die alle Substantive in ihrer Fülle zusammenfassen. So schreibt auch die Dudengrammatik, dass eine Einteilung der Substantive aus unterschiedlichen Perspektiven hinsichtlich der Bedeutung möglich ist (Regel 220). Hinzu kommt, dass zwischen den einzelnen Gruppen Zusammenhänge bestehen und Substantive nicht immer eindeutig einer Gruppe zugeordnet werden können. Folgende Bedeutungsgruppen werden unterschieden:

**Konkreta und Abstrakta (Regel 221)**

Konkreta nennt man Substantive, mit denen etwas Gegenständliches bezeichnet wird (*Mensch, Mann, Frau, Kind, Blume, Tisch, Wald*). Abstrakta sind Substantive, mit denen etwas Nichtgegenständliches bezeichnet wird (*Geist, Seele, Leben, Schlag, Angst, Wurf, Frieden, Liebe*).

Die Abstrakta sind im Duden folgendermaßen gegliedert:

- **Menschliche Vorstellungen** (*n-abstr-vorst*): Geist, Seele

  Beispiel (47): Ein Schlitzohr hat *Geister* in der Flasche. (011_MM.exb)

- **Handlungen** (*n-abstr-hdlg*): Schlag, Wurf, Schnitt, Boykott

  Beispiel (48): Ich hab' das (mit), nicht mit *Absicht* umgekippt. (026_LUA.exb)

- **Vorgänge** (*n-abstr-vorg*): Leben, Sterben, Schwimmen, Schlaf, Reise
KAPITEL 7. DER GRUNDWORTSCHATZ SEMANTISCH BETRACHTET

Beispiel (49): Dieses Mal wird es von der Herstellung ein bisschen (anders). (033_JS.exb)

- Zustände (n-abstr-zust): Friede, Ruhe, Angst, Liebe, Alter

Beispiel (50): Soll ich dir sagen, warum ich solche Angst hatte? (058_JS.exb)

- Eigenschaften (n-abstr-eig): Würde, Verstand, Ehrlichkeit, Krankheit, Dummheit

Beispiel (51): Das ist nicht die Länge. (049_MK.exb)

- Verhältnisse oder Beziehungen (n-abstr-verh): Ehe, Freundschaft, Nähe, Unterschied

Abstrakte Nomen in der Kategorie Verhältnis oder Beziehung kamen in den Daten dieser Arbeit nicht vor.

- Wissenschaften, Künste (n-abstr-wiss): Biologie, Mathematik, Musik

Beispiel (52): ... der kann nicht so viel Deutsch. (024_LUA.exb)

- Maß- und Zeitbegriffe (n-abstr-maß bzw. n-abstr-zeit): Meter, Watt, Gramm, Jahr, Stunde, Mai

Beispiel (53): Nächste Woche ist schon Weihnachten. (040_JS.exb)


Belebtheit (Regel 222)

Konkreta werden im Duden näher nach Belebtheit bestimmt und auch in der vorliegenden Arbeit wurden die konkreten Substantive nach Belebtheit (n-belebt) und Unbelebtheit (n-unbelebt) unterschieden.
7.1. DIE BEDEUTUNGSKATEGORIEN

• belebte Substantive: Mensch, Katze, Käfer, Baum, Alge

Beispiel (54): Eine Babykuh hat bei Mama getrunken. (019_RD.exb)

Beispiel (55): Kimba ist der kleine weiße Löwe. (019_JS.exb)

• unbelebte Substantive: Fahrzeug, Weg, Stein

Beispiel (56): Hier kommen die ganzen Holzteile und Porzellan (...) (013_JS.exb)

Beispiel (57): Hier kommen die Murmeln rein. (024 LUA.exb)


Eigennamen und Appellativa (Regel 223)

Eine Unterscheidung zwischen Eigennamen und Appellativa wird in der Literatur meistens derart vorgenommen, dass Eigennamen ein bestimmtes Individuum bezeichnen; Appellativa hingegen bezeichnen eine Gattung oder Klasse. Eigennamen bezeichnen demnach keine besondere Klasse mit bestimmten Eigenschaften, sondern jeweils ein bestimmtes Individuum ([Dudenredaktion (2009)]).

Beispiele für Appellativa (nach [Dudenredaktion (2009)]):

• Katzen können in der Dunkelheit gut sehen.

• Autos verschmutzen die Umwelt.

• Bären sind Raubtiere.

Appellativa werden im Duden in folgende Gruppen eingeteilt:

• Volksbezeichnungen (Regel 226): Italiener, Russin

• Produktbezeichnungen (Regel 227): Apple Power Macintosh, VW Golf

Variant
• Nomenklaturen (Regel 228): das Insekt, der Tausendfüßer, der Käfer

Beispiele für Eigennamen (nach [Dudenredaktion (2009)]):

• Paul ist sieben Jahre alt.

• die Vereinigten Arabischen Emirate

• Frankreich grenzt an Deutschland.

Eigennamen können laut Dudengrammatik in weitere Gruppen unterteilt werden, die hier zu Zwecken der Einheitlichkeit zwar aufgeführt, in der späteren semantischen Betrachtung jedoch nicht derart spezifisch verwendet wurden (Regel 225).

• Personennamen: Moritz, Gertrud, Elisabeth

• Geographische Namen: Finnland, das Allgäu, die Zugspitze

• Astronomische Eigennamen: Saturn (Planet), der Halleysche Komet

• Institutionen, Organisationen, Firmen: der Stadtrat, das Statistische Bundesamt

• Zeitungen, Zeitschriften, Bücher: die Morgenpost, die Zeitschrift für Sprachwissenschaft, die Bibel

• Einzelne Ereignisse: der Schwarze Freitag, der Zweite Weltkrieg

• Andere Objekte und Erscheinungen: das Weiße Haus, der Schiefe Turm von Pisa

• Übernamen: der Alte Fritz (= Friedrich der Große), der Rote Planet (=Mars) (Beispiele aus [Dudenredaktion (2009)])


Beispiel (58): MK, ich bin auf dem Hochbett! (024 LUA.exb)

Beispiel (59): Kannst du mit mir Memory spielen? Memory, Julia? (045 MM.exb)
7.1.9 Pronomen

In der Dudengrammatik erfolgt in Regel 349 (Seite 252) eine Beschreibung der Semantik von Pronomen (und Artikelwörtern). Demzufolge haben Pronomen verweisende, zeigende, fragende oder quantifizierende Funktion. Pronomen können zudem die Funktion einer Nominalphrase übernehmen. Im weiteren Verlauf werden alle im Deutschen auftretenden Pronomenarten beschrieben und in einer Tabelle gegenübergestellt (Regel 350). Die dort aufgeführten Pronomenarten entsprechen im Wesentlichen den in den STTS aufgelisteten POS-Tags, weshalb an dieser Stelle keine Abweichungen zu den semantischen Kategorien auftreten. Folgend werden deshalb die Pronomen mit ihren Entsprechungen in den STTS aufgeführt. Dies entspricht ferner der Beschreibung in der semantischen Analyse, wobei die Bezeichnung für die Bedeutung in Kleinbuchstaben erfolgt. In Klammern steht zuerst die Bezeichnung für die POS-Tags, dahinter die Bezeichnung für die Bedeutungsanalyse. Angegeben sind nur jene Kategorien, die auch tatsächlich in den Daten vorkommen:

- **Personalpronomen** (PPER, pper): ich, mich, mir, meiner, wir, uns, unser, du, dich, dir, deiner, ihr, euch, euer, er, sie, es, ihn, seiner, sie, ihnen, ihrer

  Beispiel (60): Ich habe kein rot mehr. (053_AV.exb)

  Beispiel (61): Wenn du mal wieder den Roboter brauchst, musst du nur anrufen. (033_JS.exb)

- **Reflexivpronomen** (PRF, prf): mich, mir, uns, dich, dir, euch, sich, einander

  Beispiel (62): Nein, der muss sich selbst was holen. (033_JK.exb)

  Beispiel (63): Ich wünsche mir rot. (036_MM.exb)

- **Possessivpronomen** (PPOSAT, pposat, PPOSS, pposs): mein - unser, dein - euer, sein - ihr

  Beispiel (64): Unsere/pposat Marina kommt heute. Das ist Mamas Freundin. (038_MM.exb)

  Beispiel (65): das ist dein(e)s/pposs (038_MM.exb)

- **attribuierende und substituierende Demonstrativpronomen** (PDAT, pdat, PDS, pds): der, die, das, dieser - jener, derjenige, derselbe

---

3 Die bestimmten und unbestimmten Artikel werden, wie schon erwähnt, unter dem Abschnitt 7.1.4 behandelt. Die attribuierenden Demonstrativpronomen (PDAT) sind hinsichtlich ihrer semantischen Zuordnung eigentlich Artikel. Da sie bei der Bezeichnung mittels der POS-Tags jedoch das Tag PDAT erhalten haben, sollen sie der Übersichtlichkeit halber auch hier mit dem Kürzel pdat bezeichnet werden, wohlwissend, dass dies strenggenommen in semantischer Hinsicht kein Demonstrativpronomen darstellt.
Beispiel (66): Wenn du zum Beispiel dieses\(^4\) Körbchen hier hast. (031_MK.exb)

Beispiel (67): Das\(\) ist ein ganz schön weiter Weg. (044_MK.exb)

- Relativpronomen (PRELS, prels): der, die, das, welcher, wer, was

Beispiel (68): Das ist die Nummer eins, die du einstellst. (048_MK.exb)

Beispiel (69): alle Steine, die bei mir draußen liegen (039_LEO.exb)

- Interrogativ- und Relativpronomen (PWAT, pwat, PWAV, pwav, PWS, pws): wer, was, welcher, was für (einer)

Beispiel (70): In welche\(\) Schachtel sollen die Murmeln? (024_LUA.exb)

Beispiel (71): Warum\(\) macht ihr sowas? (046_LUA.exb)

Beispiel (72): Was\(\) ist, was\(\) wollen wir jetzt machen? (033_MK.exb)

- Indefinitum (PIAT, piat, PIDAT, pidat, PIS, pis): man, jederman, jemand, irgendjemand, niemand, nichts, etwas, irgendswas, was, irgendeiner, wer, alle, jeder, beide, einige, manche, welche, solche, irgendeine, kein, ein bisschen, ein wenig, ein paar

Beispiel (73): wenn man keine\(\) Karte mehr hat (026_MK.exb)

Beispiel (74): die beiden\(\) Zahlen stimmen nicht (030_MK.exb)

Beispiel (75): aber man\(\) muss gut aufpassen (013_JS.exb)

### 7.1.10 Partikeln


Generell soll erwähnt sein, dass es - ebenso wie bei der Analyse anderer Wortarten - bei der Zuordnung der Partikeln zu ihren semantischen

\(^4\)Dieses ist wie schon erwähnt eigentlich Determiner. Aufgrund der Kennzeichnung mit dem POS-Tag PDAT durch die STTS bleibt die Bezeichnung pdat hier erhalten.
Kategorien vorkommen kann, dass eine andere Person eine andere Kategorie gewählt hätte. Aufgrund zum Teil subjektiver Entscheidungen ist das unvermeidbar. In den STTS werden beispielsweise viele Partikelnarten zu den Adverbien gezählt, was in diesem Abschnitt aufgrund der Referenzierung auf die Dudengrammatik nicht geschehen soll (z.B. [Schiller, Teufel, Stöckert und Thielen (2009)]). Wie alle anderen Zuordnungen, die im Verlauf getätigt wurden und werden, wurde auch diese Analyse mit besten Wissen und Gewissen vorgenommen.

**Gradpartikeln (Regel 871)**


Beispiele ([Dudenredaktion (2009)]: wenig, etwas, einigermaßen, fast, ziemlich, so, sehr, besonders, ganz⁵, zu, gar, überhaupt, viel (+ADJ), total, echt, unheimlich, schön

**Beispiel** (76): Dann muss man überhaupt nichts bezahlen. (033_JS.exb)

**Beispiel** (77): Mir tut es gar nicht weh. (043_RD.exb)

**Fokuspartikeln (Regel 873)**

Fokuspartikeln kennzeichnen jenen Teil des Satzes, der den größten Mitteilungswert hat. Sie werden im Folgenden mit ptk-fok bezeichnet.

Beispiele ([Dudenredaktion (2009)]: nur, allein, bloß, sogar, selbst, besonders, auch

**Beispiel** (78): Und ich spiele das sogar. (046_MK.exb)

**Beispiel** (79): Die brauchen selber keine Wärme. (046_MK.exb)

**Negationspartikeln (Regel 874)**

Die meisten Partikeln sind dadurch gekennzeichnet, dass man sie weglassen kann, ohne die Grammatik eines Satzes zu verletzen. Negationspartikeln hingegen können nicht weggelassen werden, weil sich der Wahrheitswert der jeweiligen Aussagen verändern würde. Die häufigste Negationspartikel

⁵Ganz kann je nach Kontext sowohl Partikel als auch Adjektiv sein.
Im Deutschen, um eine Aussage oder einen Satz zu negieren ist *nicht*. In der semantischen Analyse wurden alle einen Satz, eine Aussage oder einen Sachverhalt negierenden Lexeme wie *nein, nie* und *niemals* neben *nicht* als Negationspartikel mit dem Kürzel *ptk-neg* bezeichnet.

**Beispiel (80):** habe noch **nie** damit gespielt (031_SO.exb)

**Beispiel (81):** Ich weiß es **nicht** mehr, wie der heißt. (042_MK.exb)

### Abtönungspartikeln (Regel 875)

Die Abtönungspartikel, auch Modalpartikel genannt, drückt Einstellungen, Annahmen, Bewertungen und Erwartungen der Sprecherin/des Sprechers bezüglich eines Sachverhaltes aus. Sie bezieht sich auf den gesamten Satz. Das Kürzel für die Abtönungspartikel ist im Folgenden *ptk-abt*.

Beispiele ([Dudenredaktion (2009)]: *nicht, ja, bloß, doch, schon, denn, wohl, aber, nur, halt, eben, mal, auch, eigentlich, etwa, vielleicht, ruhig*

**Beispiel (82):** Findus, hack **doch** nicht auf mich drauf! (013_JS.exb)

**Beispiel (83):** Ist **denn** das ein Briefumschlag? (033_RD.exb)

### Gesprächspartikeln (Regel 880)


Beispiele ([Dudenredaktion (2009)]: *ja, ähm, äh, so, also, dann, nun, gut, naja, klar, sicher, genau, aber, übrigens, entschuldigung, natürlich, hoffentlich, leider, kaum, wie bitte?, was?, hm*

**Beispiel (84):** Äh, weiß ich nicht. (036_LAR.exb)

**Beispiel (85):** Hm, hab’ ich vergessen. (038_LL.exb)
Antwortpartikeln, Partikeln mit Verbzusatz und Partikeln mit „zu“ neben dem Infinitv

Diese Partikeln haben in der Dudengrammatik keine eigenen Einträge. Sie wurden in der semantischen Analyse jedoch weiterhin gemäß ihrer Funktion als POS-Tag in den STTS verwendet. Dabei entsprechen Antwortpartikeln (ptk-ant) der Kategorie PTKANT (ja, doch, bitte, bitteschön, danke, ok bzw. okay), die Partikeln mit Verbzusatz (ptkvz) entsprechen der Kategorie PTKVZ gemäß den STTS. Hinter dem Kürzel ptkzu verbirgt sich die Partikel zu, die zusammen mit einem Verb den Infinitiv bildet.

Beispiel (86): Dankeschön/ptk-ant (038_LL.exb)

Beispiel (87): Gut, dann geb’ ich jedem ein paar ab/ptkvz. (020_MM.exb)

Beispiel (88): ohne eine Pause zu/ptkzu machen (053_LEO.exb)

7.1.11 Verben


Beispiel (89): Wir spielen doch nur Räuber. (023_LUA.exb)

Beispiel (90): Ich habe dort schon gebastelt. (040_MK.exb)

- **Vorgangsverben** sind nicht *agentiv*. Sie beschreiben dynamische Sachverhalte, die nicht unter der Kontrolle eines Agens stehen und können *telisch* oder *atelisch* sein. Vorgangsverben werden mit dem Kürzel *(v-)*vorgang bezeichnet. Beispiele: erfrieren, wachsen, schlafen

Beispiel (91): In der Zeit wird Papier für euch geliefert. (033_RD.exb)

Beispiel (92): Hier schlafen ganz viele Krippenkinder. (038_LL.exb)


Beispiel (93): Ich glaube mein Papa... (020_LAR.exb)

Beispiel (94): Wir haben ’ne Meisterin. (020_MM.exb)

**Auxiliarverben, Modalverben, Kopulaverben**

• (Anna) ist (gesund). (Anna) wird (gesund). (Anna) bleibt (gesund).

• (Anna) ist (Chefin). (Anna) wird (Chefin). (Anna) bleibt (Chefin). (Regel 1202)

Des Weiteren können bei Kopulaverben auch Prädikative mit identifizierender Bedeutung stehen:

• (Der Gärtner) war (der Mörder). (Regel 1203)

Manchmal sind insbesondere auch Adverbialien (z. B. Lokal- oder Modaladverbialien) als Ergänzung bei Kopulaverben möglich:

• (Anna) ist (in Paris). (Der Ring) ist (aus Silber). (Regel 1203)


Beispiel (95): du hast aufgedeckt (031_MM.exb) (Auxiliarverb)

Beispiel (96): Oder man springt runter mit einem Seil, das hier angefesselt ist. (041_MK.exb) (Auxiliarverb)

Beispiel (97): Ich will mir dir Ritterburg spielen. (023_MK.exb) (Modalverb)

Beispiel (98): Soll ich dir sagen, was KD immer macht? (039_LEO.exb) (Modalverb)

Beispiel (99): Das sind keine Indiander. (039_LEO.exb) (Kopulaverb)
Beispiel (100): Das ist im Aquarium. (053_RD.exb) (Kopulaverb)

Wie in diesem Abschnitt dargestellt, ergeben sich für die Bezeichnung der Verben sechs Kategorien: (v-)handlung, (v-)vorgang, (v-)zustand, v-mod, v-aux und v-kop. Im hier erstellten Lexikon (Anhang) werden Verben der Handlung, des Vorgangs und des Zustandes der Übersichtlichkeit halber mit v-handlung, v-vorgang und v-zustand gekennzeichnet, wobei v- die Kategorie Verb einleitet.

7.2 Tendenzen in der Verwendung von Inhaltswörtern

In den folgenden Unterabschnitten werden ausgesuchte Lemmata im Hinblick auf mögliche sprachliche und kognitive Entwicklungsschritte der hier untersuchten Kinder - an dieser Stelle zunächst die Inhaltskategorien - vorgestellt. Die Auflistung aller Lemmata in den drei Altersgruppen, geordnet nach ihren jeweiligen Bedeutungskategorien, befindet sich als gesondertes Dokument als Teil des erstellten Lexikons im Anhang dieser Arbeit (Zusammenfassung_Lemmata_in_den_Gruppen).

7.2.1 Nomen (n-abstr-eig, n-abstr-hdlg, n-abstr-maß, n-abstr-vorg, n-abstr-vorst, n-abstr-zeit, n-abstr-zust, n-belebt, n-unbelebt, ne)


<table>
<thead>
<tr>
<th>Nomen/Altersgruppe</th>
<th>Gruppe 1281_1495</th>
<th>Gruppe 1496_1708</th>
<th>Gruppe 1709_1983</th>
</tr>
</thead>
<tbody>
<tr>
<td>abstrakte Begriffe (Auswahl)</td>
<td>Geheimnis (Gespannt), Idee, Kontrolle, Krach, Schuld, Welt</td>
<td>Angst, Ärger, Durst, Hilfe, Idee, Krieg, Schuld, Zauber</td>
<td>Angst, Chance, Geist, Glück, Idee, Lüge, Möglichkeit, Pech</td>
</tr>
</tbody>
</table>

Tabelle 7.1: Nomen in den Altersgruppen

Tabelle 7.1 zeigt sehr deutlich, dass die hier untersuchten Kinder mit zunehmendem Alter eine Vielzahl von Komposita produzierten, die ein hohes Maß an Kreativität aufweißen. Insbesondere in den Gruppen 2 und 3 zeugen Nomen wie Helflöwe, Feinlöwe, Giftgurke, Apfelpresse oder Papierfisch von
KAPITEL 7. DER GRUNDWORTSCHATZ SEMANTISCH BETRACHTET


7.2.2 Verben (v-aux, v-kop, v-mod, v-hdlg, v-vorgang, v-zustand)


---

\(^6\)Hinsichtlich der semantischen Kategorien sind unter den Vollverben alle Handlungs-, Vorgangs- und Zustandsverben zusammengefasst.
### Tabelle 7.2: Verben in den Altersgruppen

<table>
<thead>
<tr>
<th>Verben/ Altersgruppe</th>
<th>1281_1495</th>
<th>1496_1708</th>
<th>1709_1983</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Modalverben</strong></td>
<td>können (33), müssen (25), wollen (23), dürfen (8), sollen (1)</td>
<td>können (159), müssen (139), wollen (71), dürfen (35), sollen (27), mögen (6), möchten (4)</td>
<td>müssen (129), können (121), dürfen (65), wollen (54), sollen (26), möchten (19), mögen (7)</td>
</tr>
<tr>
<td><strong>Vollverben</strong> (häufigste)</td>
<td>gucken (20), machen (19), wissen (19), sehen (18), gehen (17), spielen (12)</td>
<td>gucken (150), machen (138), wissen (90), kommen (76), spielen (70), gehen (66), glauben (10)</td>
<td>machen (77), wissen (70), gehen (58), kommen (55), spielen (55), glauben (25)</td>
</tr>
</tbody>
</table>

7.2.3 Adjektive (ad-part, ad-qual, ad-quant, ad-rel)


<table>
<thead>
<tr>
<th>Adjektive/</th>
<th>1281_1495</th>
<th>1496_1708</th>
<th>1709_1983</th>
</tr>
</thead>
<tbody>
<tr>
<td>Altersgruppe</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>groß (13), cool (6), ganz (5), richtig (5), klein (4), schnell (4)</td>
<td>klein (28), gut (20), groß (19), lang (15), richtig (14)</td>
<td>gut (33), klein (26), schnell (21), groß (20), richtig (15)</td>
<td></td>
</tr>
<tr>
<td>Farbadjektiv</td>
<td>gold (8)</td>
<td>grün (26), rot (27), gelb (22)</td>
<td>rot (44), grün (40), gelb (39), blau (24), orange (19)</td>
</tr>
</tbody>
</table>

Tabelle 7.3: Adjektive in den Altersgruppen


In Gruppe 1 äußerten die Probandinnen und Probanden:

- groß - klein, gut - schlecht, früh - spät

In Gruppe 2 wurden folgende Antonyme geäußert:

- groß - klein, gut - schlecht, kurz - lang, neu - alt, richtig - falsch, tief - flach, viel - wenig, weit - nah, schwer - leicht, früh - spät, voll - leer, warm - kühl sowie viel - wenig

In den Daten von Gruppe 3 befinden sich die folgenden Antonyme:

- alt - neu, breit - schmal, richtig - falsch, groß - klein, kurz - lang, leicht - schwer, teuer - billig, viel - wenig, dick - dünn, gut - schlecht, hoch - tief, schmal - weit sowie traurig - lustig

Diese Entwicklung geht laut Kauschke (2012) einher mit dem Erwerb des hierarchisch gegliederten Nomenlexikons, wobei Kinder Wörter in weiteren semantischen Zusammenhängen erwerben. Neben dem Erwerb der Bedeutungsgleichheit bzw. der Synonymie steht der Erwerb von Oppositionsbeziehungen wie Antonymie (heiß - kalt), Kontradiktion (tot - lebendig), Konversion (ziehen - schieben) sowie die Teil-Ganzes-Beziehung
7.2. TENDENZEN IN DER VERWENDUNG VON INHALTSWÖRTERN


Abbildung 7.6: Adjektive in den drei Altersgruppen (Ausschnitt)

In einer früheren Studie untersuchte Kuczaj (1982a) ein Kind und fand heraus, dass es die Gegensatzpaare hot - cold zur Bezeichnung von Temperaturunterschieden zuerst erwarb; später erst kam cool - warm dazu ([Kuczaj (1982)]). Selbst nachdem das Kind gelernt hatte, dass hot...
KAPITEL 7. DER GRUNDWORTSCHATZ SEMANTISCH BETRACHTET


Hinsichtlich der Verwendung von Adjektiven in Bezug auf die Bedeutungskategorien (ad-part, ad-qual, ad-quant, ad-rel) konnten folgende Tendenzen beobachtet werden: Insgesamt weisen die hier untersuchten Kinder mit zunehmendem Alter ein immer differenzierteres Adjektivlexikon (vor allem in Bezug auf die Kategorie ad-qual) auf, das insbesondere ab der zweiten Altersgruppe zum Teil sehr spezifische Adjektive enthält (u.a. magnetisch, kugelrund, unendlich, ungerecht, vernünftig, verrückt). Adjektivische Partizipien wurden in nur einem Fall in Gruppe 2 von LEO geäußert (fliegender).

7.2.4 Adverbien (adv-kaus, adv-komm, adv-lok, adv-mod, adv-temp)


Besonders auffällig bei der Analyse der kausalen Adverbien (adv-kaus) war, dass das Adverb also erst ab der zweiten Altersgruppe (ab 4;1) Jahren geäußert wurde (50 mal in Gruppe 2, 12 mal in Gruppe 3). Zu diesem Phänomen liegen in der aktuellen und auch älteren Forschungslandschaft keine bzw. nur sehr wenige bekannte Studien und Ergebnisse vor. Zum einen besteht die Möglichkeit, dass also zwar im Lexikon der Kinder der ersten Gruppe enthalten ist, aber nicht verwendet wurde. Zum anderen besteht die Annahme, dass also noch nicht im (produktiven) Lexikon der hier untersuchten Kinder enthalten ist.

In Anbetracht der lokalen (adv-lok), modalen (adv-mod) und temporalen Adverbien (adv-temp) kann mit zunehmendem Alter eine immer differenziertere Verwendung dieser Adverbien verzeichnet werden.
Lokale Adverbien wie *dazwischen*, *nebeneinander* oder *seitwärts* kommen in dieser spezifischen Form erst ab 4;1 Jahren vor. Ebenso weisen die temporalen Adverbien *gleichzeitig* oder *inzwischen* auf ein sehr spezifisches Adverblexikon ab einem Alter von 4;1 Jahren hin. Derartig konkrete Angaben zum Ort oder zur Zeit konnten bei jüngeren Kindern der Gruppe 1 nicht beobachtet werden.

### 7.3 Tendenzen in der Verwendung von Funktionswörtern

An dieser Stelle sollen mögliche Entwicklungstendenzen der hier verwendeten Funktionswörter, und damit verbunden den konkret verwendeten Lemmata, über die drei Alterszeiträume hinweg dokumentiert werden. Zu den Funktionswörtern zählen dabei *Artikel, Pronomen, Präpositionen* und *Konjunktionen* ([Bussmann (1983)]) sowie in dieser Arbeit außerdem die *Interjektionen* und *Partikeln*. Hinsichtlich der Funktionswörter konnten an einigen Stellen Entwicklungstendenzen beobachtet werden, die sich - ebenso wie die Verteilung der Wortarten - vor allem ab Altersgruppe 1406_1708, also ab ca. 4;1 Jahren, bemerkbar machen.


#### 7.3.1 Interjektionen (*itlj*)

Interjektionen wurden in allen drei Altersgruppen geäußert, wobei ab der zweiten Gruppe ein differenzierteres Vokabular hinsichtlich der Interjektionen erkennbar ist. In Gruppe 1 (1281_1495) wurden nur vier verschiedene Interjektionen geäußert (*aua, cool, hey, pff*). Ab einem Alter von 4;1 Jahren bzw. in den Gruppen 2 und 3 befinden sich sehr spezifische und - so scheint es - zielgerichtete Interjektionen (*pfui, wow, juhu, tschüs, mann, prima, yippie*).
Dies lässt vermuten, dass Interjektionen, zumindest was die hiesigen Daten anbelangt, ab einem Alter von 4;1 Jahren sehr viel bewusster eingesetzt werden und dementsprechend eine andere Funktion erfüllen als bei jüngeren Kindern.

### 7.3.2 Konjunktionen (kon-add, kon-adv, kon-alt, kon-kaus, kon-spez, kon-temp, kon-vgl) und Subjunktionen (sub-fin, sub-kaus, sub-kond, sub-konz, sub-mod-instr, sub-neutr, sub-temp)

Hinsichtlich der Konjunktionen können einige wenige Entwicklungstendenzen festgestellt werden. Die adversative Konjunktion *sondern* kommt in Gruppe 1 nicht vor, während sie in Gruppe 2 und 3 mit drei bzw. zwei Fällen vertreten ist. Die spezifizierende Konjunktion *außer* hingegen kommt mit einem Fall nur in Gruppe 1 vor. Die temporale Konjunktion *als* tritt mit ebenfalls einem Fall nur in Gruppe 2 auf. An dieser Stelle sollte aufgrund der geringen Fallzahlen nicht von Entwicklungstendenzen gesprochen werden. Dazu wären weitere Daten erforderlich.


Grundsätzlich betrachtet kann aufgrund der vermehrten Verwendung von
Subjunktionen ab einem Alter von 4;1 Jahren ein vermutlich gesteigertes syntaktisches Wissen um die Bildung bzw. Einleitung von Nebensätzen angenommen werden.

7.3.3 Pronomen \( (pav, pdat, pds, piat, pidat, pis, pper, ppos, prels, prf, pwat, pwav, pws) \)

Bei der Analyse der Wortarten nach den STTS (Kapitel 6) sowie der Zuordnung der Lemmata zu ihren Bedeutungskategorien wurden die Pronominaladverbien \( (pav) \) zu den Adverbien gezählt. Doch die Pronominaladverbien stellen einen Grenzfall zwischen Adverb und Pronomen dar (siehe zum Beispiel [Dudenredaktion (2009)]). Deshalb werden sie an dieser Stelle ausnahmsweise den Pronomen zugeordnet, weil sie sich im gesamten Korpus eher bei den Funktionswörtern, denn bei den Inhaltswörtern, einreihen. Es fällt auf, dass erst in den Gruppen 2 und 3, also ab einem Alter von 4;1 Jahren die Pronominaladverbien \( \text{außerdem} \) und \( \text{trotzdem} \) auftreten. Des Weiteren finden sich in diesen Altersgruppen die Pronominaladverbien \( \text{wofür} \), wenn auch nur mit je einem Fall. In Gruppe 1 sind lediglich Pronominaladverbien vertreten, die mit \( \text{da(r)-} \) oder \( \text{hier-} \) gebildet werden (siehe auch [Dudenredaktion (2009)], Regel 860).

Bei den attribuierenden Indefinitpronomen ohne \( (piat) \) und mit Determiner \( (pidat) \) sowie bei den substituierenden Indefinitpronomen \( (pis) \) ist ab Gruppe 2 eine größere Vielfalt im Vokabular zu verzeichnen. In der Kategorie \( pdat \) treten ab einem Alter von 4;1 Jahren Pronomen wie \( \text{(ein) bisschen, solch-, andere, beide, jeder oder auch derselbe} \) auf, die in der ersten Altersgruppe noch nicht zu finden sind. Ähnlich sieht es in der Kategorie \( pidat \) aus. In Gruppe 1 tritt lediglich das Pronomen \( \text{beide} \) auf, während in den Gruppen 2 und 3 \( \text{andere, ein, (ein) paar und bisschen} \) hinzukommen. Ein ähnliches Muster ist bei den substituierenden Indefinitpronomen erkennbar, wenn auch die verwendeten Pronomen bereits in Gruppe 1 recht vielfältig verwendet wurden. In Gruppe 2 und 3 treten neben den sehr häufig verwendeten Pronomen \( \text{all-, ein, etwas} \) und \( \text{man} \) auch seltener verwendete Pronomen wie \( \text{irgendwas, irgendswelch-}, \text{jemand, jeder, manch-, niemand, wenig} \) oder \( \text{welch-} \) auf. Letztere sind in Gruppe 1 noch nicht vertreten.

Bei den Personalpronomen \( (pper) \) und Relativpronomen \( (prels) \) sind keine Besonderheiten erkennbar. Im Hinblick auf die reflexiven Personalpronomen \( (prf) \) ist, wie auch bei den anderen Pronomen, ein etwas differenzierteres Vokabular ab der zweiten Altersgruppe erkennbar, das ebenfalls bei den Possessivpronomen \( (ppos) \) verzeichnet werden kann.
7.3. TENDENZEN IN DER VERWENDUNG VON FUNKTIONSWÖRTERN

7.3.4 Präpositionen \( (pr-kaus, pr-lok, pr-mod, pr-neutr, pr-temp) \)


Bezüglich der modalen Präpositionen \( (pr-mod) \) befinden sich in Gruppe 1 lediglich \textit{für}, \textit{mit} und \textit{zu}. Danach tauchen auch Präpositionen auf wie \textit{gegen}, \textit{ohne} oder \textit{um}. Ein ähnliches Muster ist bei den temporalen Präpositionen \( (pr-temp) \) erkennbar. Während in Gruppe 1 nur die Präpositionen \textit{in} und \textit{seit} verwendet wurden, finden sich bei älteren Kindern auch nähere Bestimmungen der Zeit durch Präpositionen wie \textit{an}, \textit{vor}, \textit{zu} und \textit{nach}.

7.3.5 Partikeln \( (ptk-abt, ptk-ant, ptk-fok, ptk-gespr, ptk-grad, ptk-neg, ptkvz, ptkzu) \)


Die Verwendung von Abtönungspartikeln \( (ptk-abt) \), Fokuspartikeln \( (ptk-fok) \) und Gesprächspartikeln \( (ptk-gespr) \) erscheint ab Gruppe 2 (1496_1708) vielfältiger. Aufgrund der recht bedeutungsarmen Inhalte einzelner Partikeln (z. B. \textit{äh}, \textit{mh}, \textit{ne}, \textit{oh}, \textit{ah}) lassen sich jedoch nur schwer Tendenzen ausmachen. Lediglich ein Blick auf die Produktion der Gradpartikeln \( (ptk-grad) \) lässt einen gezielteren Einsatz mittels zum Teil spezifischer Partikeln vermuten. Besonders auffällig ist hier die Verwendung der Partikeln \textit{überhaupt},
besonders, ziemlich und sehr, die im Vokabular der jüngeren Kinder in Gruppe 1 nicht vorkommen und erst in Gruppe 2 und 3 auftreten.

7.4 Hinweise zur Verwendung des Lexikons

werden. Eine Veröffentlichung dieser Daten in digitaler Form ist, wie eingangs schon erwähnt, leider nicht möglich.

7.4.1 Beispielhafte Ausgabedateien

An dieser Stelle sollen der Vollständigkeit halber einige der Ausgabedateien, die mittels der Ausführungen durch die Skripte des CA erstellt wurden, gezeigt werden. Insbesondere vor dem Hintergrund, dass durch die hier angewandte Methodik im gesamten Verlauf dieser Arbeit zahlreiche und sehr vielfältige Möglichkeiten mit den in Kapitel 5 verwendeten Skripten entstanden, rechtfertigen diesen Unterabschnitt.

- Ein Ausschnitt aus der Ausgabedatei, die durch das Ausführen des Skriptes `db_query_all_tag_count.sh` entstanden ist, ist in Abbildung 7.7 zu sehen, die einige der von LEO und LUA geäußerten Nomen im Zeitraum 1496 bis 1708 zeigt.

```
<table>
<thead>
<tr>
<th>Datum</th>
<th>Tag</th>
<th>Zeitraum</th>
<th>Tag-Count</th>
<th>Lemma</th>
<th>EXB</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005-07-02</td>
<td>2009-09-29</td>
<td>1550.0</td>
<td>023_LUA_lem.exb</td>
<td>bauernhof</td>
<td>NN</td>
</tr>
<tr>
<td>2005-07-02</td>
<td>2009-09-29</td>
<td>1550.0</td>
<td>023_LUA_lem.exb</td>
<td>ritter</td>
<td>NN</td>
</tr>
<tr>
<td>2005-07-02</td>
<td>2009-09-29</td>
<td>1550.0</td>
<td>023_LUA_lem.exb</td>
<td>ritterteil</td>
<td>NN</td>
</tr>
<tr>
<td>2005-07-02</td>
<td>2009-09-29</td>
<td>1550.0</td>
<td>023_LUA_lem.exb</td>
<td>pferden</td>
<td>NN</td>
</tr>
<tr>
<td>2005-07-02</td>
<td>2009-09-29</td>
<td>1550.0</td>
<td>023_LUA_lem.exb</td>
<td>pferden</td>
<td>NN</td>
</tr>
<tr>
<td>2005-07-02</td>
<td>2009-09-29</td>
<td>1550.0</td>
<td>023_LUA_lem.exb</td>
<td>stäcke</td>
<td>NN</td>
</tr>
<tr>
<td>2005-07-02</td>
<td>2009-09-29</td>
<td>1550.0</td>
<td>023_LUA_lem.exb</td>
<td>sættel</td>
<td>NN</td>
</tr>
<tr>
<td>2005-07-02</td>
<td>2009-09-29</td>
<td>1550.0</td>
<td>023_LUA_lem.exb</td>
<td>cousine</td>
<td>NN</td>
</tr>
<tr>
<td>2005-07-02</td>
<td>2009-09-29</td>
<td>1550.0</td>
<td>023_LUA_lem.exb</td>
<td>kindergarten</td>
<td>NN</td>
</tr>
<tr>
<td>2005-07-02</td>
<td>2009-09-29</td>
<td>1550.0</td>
<td>023_LUA_lem.exb</td>
<td>cousine</td>
<td>NN</td>
</tr>
<tr>
<td>2005-07-02</td>
<td>2009-09-29</td>
<td>1550.0</td>
<td>023_LUA_lem.exb</td>
<td>cousine</td>
<td>NN</td>
</tr>
<tr>
<td>2005-07-02</td>
<td>2009-09-29</td>
<td>1550.0</td>
<td>023_LUA_lem.exb</td>
<td>pferden</td>
<td>NN</td>
</tr>
<tr>
<td>2005-07-02</td>
<td>2009-09-29</td>
<td>1550.0</td>
<td>023_LUA_lem.exb</td>
<td>stück</td>
<td>NN</td>
</tr>
<tr>
<td>2005-07-02</td>
<td>2009-09-29</td>
<td>1550.0</td>
<td>023_LUA_lem.exb</td>
<td>ritterteil</td>
<td>NN</td>
</tr>
<tr>
<td>2005-07-02</td>
<td>2009-09-29</td>
<td>1550.0</td>
<td>023_LUA_lem.exb</td>
<td>reich</td>
<td>NN</td>
</tr>
<tr>
<td>2005-07-02</td>
<td>2009-09-29</td>
<td>1550.0</td>
<td>023_LUA_lem.exb</td>
<td>pfand</td>
<td>NN</td>
</tr>
<tr>
<td>2005-07-02</td>
<td>2009-09-29</td>
<td>1550.0</td>
<td>023_LUA_lem.exb</td>
<td>berg</td>
<td>NN</td>
</tr>
<tr>
<td>2005-07-02</td>
<td>2009-09-29</td>
<td>1550.0</td>
<td>023_LUA_lem.exb</td>
<td>borg</td>
<td>NN</td>
</tr>
<tr>
<td>2005-07-02</td>
<td>2009-09-29</td>
<td>1550.0</td>
<td>023_LUA_lem.exb</td>
<td>gold</td>
<td>NN</td>
</tr>
<tr>
<td>2005-07-02</td>
<td>2009-09-29</td>
<td>1550.0</td>
<td>023_LUA_lem.exb</td>
<td>gold</td>
<td>NN</td>
</tr>
<tr>
<td>2005-07-02</td>
<td>2009-09-29</td>
<td>1550.0</td>
<td>023_LUA_lem.exb</td>
<td>gold</td>
<td>NN</td>
</tr>
<tr>
<td>2005-07-02</td>
<td>2009-09-29</td>
<td>1550.0</td>
<td>023_LUA_lem.exb</td>
<td>sand</td>
<td>NN</td>
</tr>
<tr>
<td>2005-07-02</td>
<td>2009-09-29</td>
<td>1550.0</td>
<td>023_LUA_lem.exb</td>
<td>murmeln</td>
<td>NN</td>
</tr>
<tr>
<td>2005-07-02</td>
<td>2009-09-29</td>
<td>1550.0</td>
<td>023_LUA_lem.exb</td>
<td>golden</td>
<td>NN</td>
</tr>
<tr>
<td>2005-07-02</td>
<td>2009-09-29</td>
<td>1550.0</td>
<td>023_LUA_lem.exb</td>
<td>schnauchel</td>
<td>NN</td>
</tr>
<tr>
<td>2005-07-02</td>
<td>2009-09-29</td>
<td>1550.0</td>
<td>023_LUA_lem.exb</td>
<td>hause</td>
<td>NN</td>
</tr>
<tr>
<td>2005-07-02</td>
<td>2009-09-29</td>
<td>1550.0</td>
<td>023_LUA_lem.exb</td>
<td>rauber</td>
<td>NN</td>
</tr>
</tbody>
</table>
```

Abbildung 7.7: tag count_NN_1496_1708

- Abbildung 7.8 zeigt einen Ausschnitt einer Ausgabedatei, die durch das Ausführen des Skriptes `db_query_all_tag_count_name.sh` entstanden ist und einige der von JS geäußerten Adverbien (ADV) im Zeitraum 1709 bis 1983 enthält.
KAPITEL 7. DER GRUNDWORTSCHATZ SEMANTISCH BETRACHTET

Abbildung 7.8: tag count_JS_ADV_1709_1983

- Abbildung 7.9 zeigt einen Ausschnitt der im Zeitraum 1281 bis 1495 geäußerten Vollverben im Infinitiv (VVINF) bezogen auf die Tokens. Grundlegend hierfür war das Ausführen des Skriptes db_query_all_typetoken.sh.

- In Abbildung 7.10 ist ein Ausschnitt der von allen Sprecherinnen und Sprechern im Zeitraum 1709 bis 1983 geäußerten Adjektive (ADJ) zu sehen. Erforderlich war das Ausführen des Befehls db_query_all_typetoken_lemma.sh.

- In Abbildung 7.11 sind auschnitthaft die Modalverben (VMFIN) der Sprecher/innen, die diese im Zeitraum 1496 bis 1708 verwendet haben, zu sehen. Durch Ausführen des Skriptes db_query_all_lemma_tag.sh werden neben den Wörtern auch die Lemmata und die dazugehörigen POS-Tags mit angezeigt.

7.5 Zusammenfassung

Abbildung 7.9: type token_VVINF_1281_1495

- Durch das Ausführen des Befehls `db_query_all_typetoken_bedeutung.sh` entstehen Dateien, wie sie in Abbildung 7.13 zu sehen sind. Die Abbildung zeigt die Ausgabedatei für die Altersgruppe 1709 bis 1983 mit den entsprechenden Lokaladverbien (`adv-lok`), die in diesem Zeitraum geäußert wurden.

7.5 Zusammenfassung

In diesem Kapitel wurde, unter der Voraussetzung aller bisher getätigten Schritte, die eigentliche Erstellung des produktiven Lexikons der hier untersuchten Kinder vorgestellt und mit allen Ergebnissen präsentiert. Der dabei wichtigste Schritt war die Zuordnung aller Lemmata zu sogenannten Bedeutungskategorien, die sich auf semantische Aspekte beziehen. Im Gegensatz dazu bezog sich die Zuordnung der Lemmata zu den POS-Tags vornehmlich auf lexikalisch-synaktische Aspekte. Mithilfe des CA (Kapitel 5) wurden alle neuen Daten unmittelbar in die Datenbank importiert, wodurch mit Hilfe von zwei neu erstellten Skripten (5.2.4.5 und 5.2.4.6) das eben beschriebene Lexikon mit seinen zusätzlichen Dateien erstellt werden konnte. Durch die nachfolgende Betrachtung aller Bedeutungskategorien in den Altersgruppen (Abschnitte 7.2 und 7.3) konnten Entwicklungstendenzen ausgemacht werden, die es wert sind, in Zukunft näher untersucht zu werden. Es bleibt vor allem die Frage, ob es sich um tatsächliche Tendenzen handelt, die ab einem Alter von 4;1 Jahren auftauchen oder aber, ob die erhobenen Daten in der Altersgruppe 1 zu gering waren im Gegensatz zu
den Gruppen 2 und 3. Erstaunlich ist, dass sich die Tendenz einer Änderung der semantischen Aspekte im Lexikon der hier untersuchten Kinder ab einem Alter von 4;1 Jahren einfügt in die ermittelten Ergebnisse bezüglich der Wortartenverteilung (Kapitel 6), die ebenfalls eine Veränderung der Verteilung der Wortarten ab einem Alter von 4;1 Jahren vermuten lassen. Diese Ergebnisse geben Anlass dazu anzunehmen, dass sich das Lexikon ab einem Alter von vier Jahren grundlegend neu zu strukturieren beginnt.
7.5. ZUSAMMENFASSUNG

Abbildung 7.11: lemma tag_VMFIN_1496_1708
Abbildung 7.12: lemma bedeutung_v-vorgang_1281_1495
<table>
<thead>
<tr>
<th>Wort</th>
<th>Häufigkeit</th>
</tr>
</thead>
<tbody>
<tr>
<td>außen</td>
<td>13</td>
</tr>
<tr>
<td>dia</td>
<td>144</td>
</tr>
<tr>
<td>dazwischen</td>
<td>1</td>
</tr>
<tr>
<td>dort</td>
<td>1</td>
</tr>
<tr>
<td>draußen</td>
<td>2</td>
</tr>
<tr>
<td>drüber</td>
<td>5</td>
</tr>
<tr>
<td>durch</td>
<td>1</td>
</tr>
<tr>
<td>entlang</td>
<td>1</td>
</tr>
<tr>
<td>hier</td>
<td>4</td>
</tr>
<tr>
<td>herum</td>
<td>1</td>
</tr>
<tr>
<td>hier</td>
<td>197</td>
</tr>
<tr>
<td>hin</td>
<td>6</td>
</tr>
<tr>
<td>hinten</td>
<td>1</td>
</tr>
<tr>
<td>hinüber</td>
<td>1</td>
</tr>
<tr>
<td>hoch</td>
<td>4</td>
</tr>
<tr>
<td>innen</td>
<td>1</td>
</tr>
<tr>
<td>irgendwo</td>
<td>2</td>
</tr>
<tr>
<td>nebeneinander</td>
<td>1</td>
</tr>
<tr>
<td>oben</td>
<td>9</td>
</tr>
<tr>
<td>raus</td>
<td>4</td>
</tr>
<tr>
<td>rein</td>
<td>8</td>
</tr>
<tr>
<td>runter</td>
<td>6</td>
</tr>
<tr>
<td>unten</td>
<td>3</td>
</tr>
<tr>
<td>zurück</td>
<td>3</td>
</tr>
</tbody>
</table>

Abbildung 7.13: type token bedeutung_adv-lok_1709_1983
Kapitel 8

Diskussion


---

\(^1\) Alle Äußerungen, auch jene mit Aussprachefehlern bzw. -schwierigkeiten, können bei Bedarf durch ein Abspielen der Audiodateien in FOLKER verfolgt werden.

KAPITEL 8. DISKUSSION

Sprache von Erwachsenen erscheint die Entscheidung für die STTS sinnvoll.

Im anschließenden Methodikteil wurden die Daten aller hier aufgeführten Probandinnen und Probanden dargelegt, das Diktiergerät sowie die Arbeit mit der hier verwendeten Software beschrieben (Kapitel 4). Hervorzuheben ist dabei insbesondere die Arbeit mit dem EXMARaLDA Partitur Editor, welcher für zahlreiche korpuslinguistische Arbeiten empfehlenswert ist. Es können nicht nur die Lemmata getaggt und mit Beschreibungen jeglicher Art versehen werden, auch die Analyse von Morphemen und Phonemen wäre darstellbar. Dafür ist in jedem Falle eine gute Datenaufbereitung durch eine qualifizierte Person notwendig, da dies nicht von der Software übernommen werden kann. Weiterhin habe ich versucht, alle Sonderfälle, die beim Taggen und bei der Lemmatisierung der gesammelten Daten auftreten können, eingrenzen und so gut wie möglich zu beschreiben.

Kinder auf gleiche Weise zu analysieren, um herauszufinden, ab welchem Alter keine gravierenden Unterschiede mehr in der Wortartenverteilung auftreten.

In der folgenden Analyse (Kapitel 6) dieser Arbeit wurden mehrere Auswertungen mit den hier erhaltenen Daten vorgenommen. Zunächst wurden alle drei betrachteten Altersgruppen (1281 bis 1495, 1496 bis 1708, 1709 bis 1983) im Hinblick auf die Verteilung der Wortarten miteinander verglichen. Dabei wurde zwischen allen verwendeten (syntaktischen) Wörtern (Tokens) sowie zwischen dem Anteil verschiedener Wörter (Types) bezüglich der Lemmata unter diesen Tokens unterschieden. Ein erster, sehr grober Vergleich ließ vermuten, dass die Verteilung der Wortarten sowohl bezüglich der Types als auch der Tokens über die Altersgruppen hinweg annähernd gleich bleibt. Bei einem genaueren Blick auf die Wortartenverteilung hinsichtlich der Tokens fiel Folgendes auf: Adverbien und Pronominaladverbien (ADV + PAV) werden prozentual recht häufig verwendet (je 14,6% - 17,3% - 12,9%). Der Anteil der Konjunktionen (KON) nimmt im Verlauf leicht ab (je 8,1% - 7,1% - 5,6%) ebenso wie der Anteil der attribuierenden, substituierenden und Indefinitpronomen (PIDATS) (je 5,2% - 4,9% - 4,0%). Die Verwendung der Personalpronomen (PPER) steigt mit zunehmendem Alter leicht an (je 9,7% - 10,6% - 12,7%) wie auch die Verwendung der Adjektive (ADJ) (je 3,7% - 3,8% - 4,5%). Erstaunlicherweise bleibt die Verwendungshäufigkeit der Verben (VA, VM und VV) nahezu gleich über die Altersgruppen hinweg (je 21,2% - 21,4% - 22,1%). Der Anteil der Nomen (N) steigt nur leicht mit zunehmendem Alter (je 9,1% - 7,7% - 11,0%) an, zeigt jedoch keinen kontinuierlichen Anstieg aufgrund des vorübergehenden Abfalls in Gruppe 2. In Bezug auf diese Werte lassen sich folgende hypothetische Aussagen treffen:

Annahme 1: Im Verlauf von 3;5 bis 5;5 Jahren verwenden Kinder zu einem großen Anteil Adverbien, die bis zu einem Fünftel des gesamten Wortschatzes ausmachen können. Nomen nehmen keinen überaus hohen Anteil am Vokabular ein, wodurch davon ausgegangen werden kann, dass nach dem dritten Lebensjahr der hier untersuchten Kinder der Anteil drastisch abnimmt und auf einem Niveau von ca. 10% zunächst stabil bleibt (eventuell bis ins Erwachsenenalter).

Laut Tomasello würden Kinder im Alter von 4 bis 5 Jahren dabei noch Schwierigkeiten haben. Er bemängelt jedoch, dass es bisher keine Studien gäbe, die die kommunikativen Bedingungen untersucht haben, in denen Kinder ein Pronomen gegenüber einem Nomen gewählt haben. Laut ihm sind zur Annäherung an diese Frage fünf Hypothesen möglich:

1. Junge Kinder verwenden denselben referierenden Ausdruck wie der Erwachsene, mit dem sie sich gerade unterhalten. Wenn der Erwachsene auf ein Objekt mit einem Nomen referiert, so wird auch das Kind dies tun.

2. Junge Kinder folgen dem Prinzip der 'mutual exclusivity' und verwenden den Objektnamen, wenn sie ihn kennen und ein Pronomen, wenn sie den Namen nicht kennen.


Campbell et al. (2000) versuchten jede dieser fünf Hypothesen zu testen, zum einen mit Kindern im Alter von 2;6 Jahren und zum anderen mit Kindern im Alter von 3;6 Jahren. Sie wollten herausfinden, ob Kinder Pronomen unterschiedlich von der von Erwachsenen gebrauchten Form verwenden, wenn

1. der Erwachsene zuvor ein Pronomen/Nomen verwendete
2. der Erwachsene das Zielereignis miterlebt
3. die Kinder unbekannte und schwierige Nomen verwenden mussten.

Das Ergebnis war, dass die Verwendung unterschiedlicher Pronomen abhängig war vom unmittelbar vorausgehenden Diskurs und von der jeweils gestellten Frage. Wurde zum Beispiel gefragt „What did X do?“, so gebrauchten die Kinder ein Pronomen oder eine Nullreferenz. Wurde hingegen gefragt „What happened?“, so antworteten Kinder eher mit einem Nomen „The boy...“ ([Campbell, Brooks und Tomasello (2000)]). Die Datenlage ist wie bereits erwähnt recht klein. Doch geben die eben erwähnte Studie sowie die


---

3Damit beziehe ich mich lediglich auf die syntaktischen Muster, nicht aber auf den Erwerb der Morphosyntax, die insbesondere im Deutschen zu diesem Zeitpunkt in der Regel noch nicht abgeschlossen ist. (Höhle (2012)), S. 135-137)

Einen weiteren Anhaltspunkt dafür, dass Kinder im Alter von 3;5 bis 5;5 Jahren bereits einen großen Teil der Struktur der jeweiligen Sprache verinnerlicht haben, deutet Pinker (1996) an und verdeutlicht dies an einem


Auch Höhle (2012) postuliert, dass der Erwerb der grundlegenden grammatischen Fähigkeiten im Laufe der ersten vier Lebensjahre abgeschlossen ist, was die Erkenntnisse dieser Arbeit in Bezug auf die unwesentlichen Veränderungen (vor allem ab Gruppe 2 bzw. ab ca. 4;1 Jahren) in der Verteilung der Wortarten unterstreicht. Die Entwicklung konversationeller und pragmatischer Fähigkeiten hingegen dauere bis weit in das Schulalter hinein. Demzufolge seien zu Beginn des Schulalters kindliche Erzählungen oft geprägt durch eine geringe lexikalische Diversität, was sich in einer geringen Anzahl verschiedener Inhaltswörter bei einem hohen...

In Kapitel 6 wurde ferner die Wortartenverteilung hinsichtlich der Types ermittelt, also, wie viele verschiedene Lemmata die Kinder pro Wortart verwendeten. Auf den ersten Blick nehmen erneut die Adverbien und Pronominaladverbien (ADV + PAV) einen erwähnenswerten Anteil am Vokabular ein, unabhängig von der Altersgruppe (je 10,6% - 9,3% - 9,5%). Diese Anteile sind im Vergleich zu den Tokens etwas niedriger. Nichtsdestotrotz gehört diese Kategorie neben den Nomen und Verben zu jener mit dem höchsten Anteil am Gesamtvokabular. Hinsichtlich der Nomen wurde festgestellt, dass diese im Vergleich zu ihrer Verwendungshäufigkeit einen enorm hohen Wert in Bezug auf die Types aufweisen (je ca. 30,5% - 40,5% - 38,8%). Aufgrund dieser Werte und der Werte bezüglich der Tokens kann angenommen werden, dass Kinder im Alter von 3;5 bis 5;5 Jahren Nomen nicht viel häufiger gebrauchen als Erwachsene. Jedoch sind die Nomen, die von den Kindern geäußert werden, äußerst verschieden. Im Umkehrschluss könnte man sagen, dass Erwachsene häufig dieselben Nomen in einer untersuchten Textpassage verwenden. An dieser Stelle kann nicht eindeutig geklärt werden, warum dies so ist. In Abschnitt 6.4 wurden bereits Vermutungen dahingehend angesetzt, dass sich Kinder häufig in Spielsituationen befinden und zudem viele Fantasiebegriffe verwenden, wodurch ein derart hoher Anteil zustande kommt. Ein Vergleich zu den Daten der Erwachsenen aus Kauschkes Analyse ist hier mit großer Vorsicht anzustellen, da die Inhalte der Gespräche nicht bekannt sind ([Kauschke (2007)])). So ist es durchaus möglich, dass ein bestimmtes Thema vorgegeben wurde und aufgrund dessen wiederholt dieselben Nomen verwendet wurden, was in einem geringeren Types-Anteil resultiert. Wie bereits erwähnt, wäre eine Inhaltsanalyse beider Parteien sehr hilfreich. Für die Daten der vorliegenden Arbeit wurden deshalb im Anschluss daran die Lemmata aller Kategorien im Hinblick auf inhaltliche Aspekte untersucht. Aufgrund der bisher erhobenen Daten kann folgende Hypothese angeführt werden:
**Annahme 2:** Kinder im Alter von 3;5 bis 5;5 Jahren verwenden nicht mehr Nomen als Erwachsene. Jene Nomen, die sie verwenden, sind jedoch von einer größeren Verschiedenartigkeit geprägt.

Zuletzt soll ein Blick auf die Anteile der Verben (VA, VM, VV) in den Altersgruppen geworfen werden. Hier gibt es zwischen den Gruppen keine gravierenden Unterschiede (je 23,7% - 22,6% - 22,9%). Ob und wann weitere Veränderungen nach dem sechsten Lebensjahr auftauchen, kann an dieser Stelle nicht gesagt werden. Es ist aber offensichtlich, dass im hier untersuchten Zeitraum von 3;5 bis 5;5 Jahren keine Veränderungen bezüglich der Verbanteile am Gesamtvokabular zu verzeichnen sind.

In einem nächsten Schritt wurden nicht nur die Daten aller Kinder in einer Altersgruppe untersucht, sondern die Daten jedes Kindes in jeder Altersgruppe. Im Vorfeld war nicht klar, ob die Ergebnisse zielführend und aussagekräftig sein würden, da jedes Kind unterschiedlich viele Tokens hervorbrachte. Nach der Analyse stellte sich jedoch heraus, dass die Verteilung der Wortarten bei jedem Kind in jeder Altersgruppe erstaunlicherweise ein ähnliches Muster aufweist und für eine Analyse herangezogen werden kann. Es reichen also bereits wenige Tokens aus, um ein aussagekräftiges Bild über die Verteilung der Wortarten zu erhalten. In der ersten Gruppe 1281 bis 1495 waren die Kinder MA und MK vertreten. Ähnlich der Ergebnisse aus der Analyse der Altersgruppen waren sowohl bei den Tokens als auch bei den Types Adverbien und Pronominaladverbien (ADV+PAV) stark vertreten (Tokens: 10,3% bis 15,7% Types: 9,0% bis 12,1%). Auch hier sind im Vergleich der Tokens zu den Types ähnliche Verteilungsmuster wie in der Altersgruppenanalyse ersichtlich. Die Anteile der Adverbien und Pronominaladverbien (ADV+PAV) sind bei den Tokens und Types relativ nah beieinander und weichen nicht derartig stark voneinander ab, wie es bei den Nomen beobachtet werden kann. Ein analoges Bild zeigt sich bei der Betrachtung der Adverbien und Pronominaladverbien (ADV+PAV) der Kinder in der zweiten Altersgruppe 1495 bis 1708 (Tokens: 14,3% bis 18,4% Types: 11,3% bis 12,9%) sowie in der Altersgruppe 1709 bis 1983 (Tokens: 9,1% bis 16,6% Types: 10,1% bis 16,7%). Die gewonnenen Ergebnisse unterstützen die bisherigen Erkenntnisse, dass Kinder im Alter von 3;5 bis 5;5 Jahren keine gravierenden Änderungen in der Verteilung der Adverbien aufweisen. Auffällig ist die Beziehung der Tokens und Types im Hinblick auf die Nomen (N). In allen Altersgruppen und bei jedem einzelnen Kind ist der Anteil am Gesamtwortschatz bezüglich der Verwendungshäufigkeit relativ klein (Gruppe 1: 8,7% bis 10,5%, Gruppe 2: 6,7% bis 11,8%, Gruppe 3: 5,9% bis 15,9%). In Anbetracht der Types ist der Anteil der Nomen sehr viel
KAPITEL 8. DISKUSSION

höher (Gruppe 1: 24,7% bis 27,1%, Gruppe 2: 24,7% bis 35,0%, Gruppe 3: 16,7% bis 30,7%), aber auch sehr viel individueller. So gibt es einige Kinder, die mit rund 17% nicht sehr viele verschiedene Nomen verwenden, MK (Gruppe 1496_1708) hingegen mit 35,0% und LL (Gruppe 1709_1983) mit 30,7% umso mehr. Dies bedeutet, wie bereits oben erwähnt, dass Kinder im Alter von 3;5 bis 5;5 Jahren eher wenige Nomen verwenden. Jene Nomen, die gebraucht werden, können aber individuell sehr verschieden sein. Ein ähnliches Bild konnte bereits bei der Analyse der Nomen in den Altersgruppen beobachtet werden. Im Hinblick auf die Tokens sind die Anteile durchgängig recht niedrig, in Bezug auf die Types wurden Werte über 30% beobachtet. Die unterschiedliche Verwendung von Nomen wurde durch Einzelanalysen jedes Kindes in jeder Altersgruppe zum Teil bestätigt. Zusätzlich wird ersichtlich, dass gerade im Hinblick auf die Types sehr individuelle Unterschiede zu verzeichnen sind, die erst durch die Einzelanalysen deutlich gemacht werden konnten. Die Anteile der Verben (VA, VM, VV) sind sowohl bei den Tokens als auch bei den Types annähernd gleich bei allen Kindern in allen Altersgruppen (Gruppe 1: 20,4% bis 21,3% Tokens, 22,4% bis 25,0% Types, Gruppe 2: 20,8% bis 23,2% Tokens, 22,5% bis 24,3% Types, Gruppe 3: 21,1% bis 24,0% Tokens, 21,0% bis 28,1% Types).


Generell kommt es zwischen 1;9 und 3;6 Jahren zu einer sprunghaften Ausweitung des Wortschatzes, woraufhin ab dem 4. Lebensjahr eine Verlangsamung zu verzeichnen ist ([Klann-Delius (1999)], S. 36). Ob und wie stark sich der Wortschatz zwischen 3;5 und 5;5 Jahren ausweitet, kann anhand der vorliegenden Daten nur zum Teil erörtert werden, da nicht die Menge der produzierten Wörter in einem gegebenen Zeitraum untersucht wurde. Ab 3;0 Jahren muss es aber zu einer Verlangsamung der Entwicklung der Wortartenverteilung gekommen sein, da in der hier untersuchten Altersspanne keine großen Veränderungen zu verzeichnen sind. Szagun (2006) fand heraus, dass die ersten Wörter im Spracherwerbsprozess folgenden Wortarten zugeordnet werden können: 60,5% Nomen, 28,6%

Sehr aussagekräftige Ergebnisse über das kindliche Lexikon erzielte Kauschke (1999), die bereits an zahlreichen Stellen dieser Arbeit zitiert wurde. Nach Kauschke ist bis zu einem Alter von 36 Monaten ein Types-Zuwachs bezüglich aller Wortarten zu verzeichnen, wobei die Anzahl unterschiedlicher Wörter in Abhängigkeit vom Alter steigt. Dabei beobachtete sie zunächst ein exponentielles Wachstum im 2. Lebensjahr, das im 3. Lebensjahr abnimmt und in einen anschließenden linearen Verlauf der Types übergeht. Ähnliches ermittelte Kauschke für die Verteilung der Tokens. Mit zunehmendem Alter seien keine bedeutenden Veränderungen mehr feststellbar, was sich unter anderem in einer gleichbleibenden Type-Token-Relation äußert ([Kauschke (1999)]). Dies kann ebenfalls für die Daten dieser Arbeit behauptet werden, wobei die Type-Token-Relation über die Altersgruppen hinweg nahezu gleich bleibt, wenn auch andere Werte annimmt als in der Auswertung Kauschkes (Abschnitt 6.2.3). Die Analyse der hier vorliegenden Daten schließt in vielen Punkten an die bisherigen Ergebnisse zum Lexikonerwerb an. So können ab 3;0 Jahren keine starken Veränderungen im Hinblick auf die Tokens festgestellt werden. Auch im Hinblick auf die Types ergibt sich ein relativ konstantes Bild - bis auf die Nomen, die individuell sehr unterschiedlich verwendet werden. Ferner wurde bereits an anderen Stellen in dieser Arbeit erwähnt, dass Kauschke zu der Erkenntnis gelangt, dass im 3. Lebensjahr kein Kind mehr als 25% Nomen verwendet und Verben den größten Anteil des Lexikons ausmachen. Im Hinblick auf die Verwendungshäufigkeit der Lemmata kann dies auch für die untersuchten Kinder zwischen 3;5 und 5;5 Jahren behauptet werden. Bei den Types sind, wie schon erwähnt, stärkere
individuelle Unterschiede zu verzeichnen, was jedoch auch Kauschke in ihren Ergebnissen herausstellte. Demzufolge fügen sich die in dieser Arbeit gewonnenen Ergebnisse bezüglich der Verteilung der Wortarten ein in das Bild der bisher gewonnenen Ergebnisse anderer Arbeiten, welche Kinder bis zu einem Alter von meist 3;0 Jahren untersuchten. Bis zu einem Alter von 5;5 Jahren sind keine großen Veränderungen in der Verteilung der Wortarten feststellbar. Ab einem Alter von ca. 4;1 Jahren (ab Gruppe 1496 bis 1708) muss es den hiesigen Daten zufolge jedoch zu einer inhaltlichen Umstrukturierung des Lexikons kommen, was mit einer qualitativen Analyse der Inhaltswörter und zu Teilen der Funktionswörter gezeigt werden konnte (Abschnitte 7.2 und 7.3).


\textsuperscript{4}Das auf Basis der hier analysierten Daten erstellte Lexikon befindet sich im Anhang dieser Arbeit. Es besteht im Wesentlichen aus drei Korpuslexika: 1) Zusammenfassung_Lemmata_aller_Gruppen_mit_Beispielen, 2) Zusammenfassung_Lemmata_in_den_Gruppen, 3) Zusammenfassung_alle_Gruppen_mit_POS_Tags.
6 vorgenommen wurde, wäre die Erstellung des Lexikons nicht möglich gewesen. Denn nur aufgrund des Wissens um die einzelnen Wortarten konnten auch semantische Aspekte der geäußerten Lexeme ermittelt werden.


Das auf diese Weise entstandene Lexikon (Anhang) mit seinen ergänzenden Dateien bildet in umfassender Weise die Spontansprache der hier untersuchten Kinder im Alter von 3;5 bis 5;5 Jahren ab und kann für zahlreiche weitere Analysen sowie als Nachschlagewerk für unterschiedliche Zwecke genutzt werden.

An dieser Stelle sollen nun die in Kapitel 4 formulierten Fragen beantwortet werden.

stabile Verteilung weiterhin bestehen bleibt, wobei Nomen nur relativ selten, aber in sehr verschiedener Form verwendet werden können. Der Anteil an Types ist in den Altersgruppen 2 und 3 höher als in Gruppe 1. Durch Einzelanalysen konnte gezeigt werden, dass der hohe Anteil an verschiedenartigen Nomen individuell derartig unterschiedlich ist, dass diesbezüglich keine allgemeinen Schlussfolgerungen getroffen werden können. Verben treten mit einem Anteil von etwa einem Fünftel bezüglich der Tokens und der Types gleichermaßen in den hier untersuchten Altersgruppen auf auf.


Die eigentliche Komposition des Lexikons der hier untersuchten Kinder wurde durch eine Analyse der Wortbedeutungen ermöglicht. In Kapitel
7 wurde beschrieben, nach welchen semantischen Kategorien alle hier geäußerten Lemmata, neben ihrer Bezeichnung mit POS-Tags, klassifiziert werden. Auf diese Weise ergaben sich neben bereits vorhandenen Kategorien, auch neue Kategorien, die wichtige Informationen enthalten. Die Klasse der Verben wurde beispielsweise unter semantischen Aspekten in **Handlungs-**, **Vorgangs-**, **Zustands-**, **Modal-**, **Auxiliar-** und **Kopulaverben** differenziert. Eine ähnlich detaillierte Differenzierung erfuhren die **Adjektive**, **Adverbien**, **Präpositionen**, **Junktionen**, **Nomen** und **Partikeln**. Alle derart klassifizierten Lemmata sind mit ihren jeweiligen semantischen Kategorien und Beispielen aus den Daten dieser Arbeit im Anhang in Form eines **Lexikons** aufgeführt. Dieses Lexikon ist das Ergebnis aller bisher getätigten Recherchen und Analysen und spiegelt anschaulich das Vokabular der hier untersuchten Kinder im Alter von 3;5 bis 5;5 Jahren wider.
Kapitel 9

Ausblick


Da bis zu einem Alter von 5;5 Jahren keine großen Veränderungen in


Stellt man sich die Frage nach dem Sinn und Zweck der hier durchgeführten Analyse, ist das nicht ganz eindeutig zu beantworten. In der linguistischen Forschung wird ein Großteil der experimentellen Studien und Beobachtungsstudien häufig deshalb gemacht, weil zum jeweiligen

\[1\] Ein eindeutiger Entwicklungsschritt bezüglich der Nomen kann hier nicht ausgemacht werden, ist aber nicht auszuschließen.
auditive Lernzwecke eingesetzt werden. Das abschließend erstellte *Lexikon* bildet, zusammen mit all seinen zusätzlichen Informationen (sprachliche Beispiele aus den EXMARaLDA-Dateien, Audio-Dateien, Wortartenanalyse, semantische Analyse, Häufigkeitsanalyse), ein nützliches Nachschlagewerk, das den Wortschatz von Kindern im Alter von 3;5 bis 5;5 Jahren zu großen Teilen abbildet.

Ich hoffe, dass ich mit der hier erstellten Arbeit einen Beitrag leisten kann, der sich in die bisherige, sehr umfangreiche Forschungslandschaft des Spracherwerbs, aber auch der Lexikologie und Lexikographie sowie der korpusbasierten Linguistik eingliedert.
Abbildungsverzeichnis

4.1 Folker-Datei: ic_a_31.flk ................................. 68
4.2 Folker-Datei: ic_a_16.flk ................................. 70
4.3 Beispielhafte EXMARaLDA-Datei mit sieben Sprecherinnen und
    Sprechern ............................................. 72
4.4 Separierte Tonspur in einer EXMARaLDA-Datei ............... 72
4.5 Tonspur eines einzelnen Sprechers (RD) mit Annotationsspur 73
4.6 EXMARaLDA-Datei: MM_lem.exb .......................... 73

5.1 CA: project folder structure ............................... 91
5.2 CA: Installation ......................................... 91
5.3 CA: copy and save files .................................. 92
5.4 CA: convert wav-files ..................................... 92
5.5 CA: create database ....................................... 93
5.6 CA: import exmaralda files ................................. 93
5.7 CA: generate all results .................................. 94
5.8 CA: generate tag count results ............................. 95
5.9 CA: type-token results ................................... 95
5.10 CA: type-token-lemma results ............................. 96
5.11 CA: lemma tag results ................................... 96
5.12 CA: lemma bedeutung results .............................. 96
5.13 CA: type-token bedeutung results ......................... 97
5.14 CA: table structure sprecher ............................ 98
5.15 CA: table structure aufnahme ............................ 99
5.16 CA: table structure: textpostags ......................... 100
5.17 CA: table structure wav_datei ........................... 100
5.18 CA: Wortliste 1281 bis 1495 ART .......................... 107
5.19 CA: Beispiel lemma_tag 1496 bis 1708 ..................... 118
5.20 CA: Beispiel all_lemma_tag_MA 1281 bis 1495 ............ 119
5.21 CA: Skript db_query_all_lemma_tag_name LL_ 1709_1983 .. 120

6.1 POS-Tags: tokens_types_gruppiert_1281_1495 ............. 131
6.2 Gruppierte POS-Tags, Verteilung Tokens, 1281 bis 1495 Tage . . 135
6.3 Gruppierte POS-Tags, Verteilung Tokens, 1496 bis 1708 Tage . . 136
6.4 Gruppierte POS-Tags, Verteilung Tokens, 1709 bis 1983 Tage . . 137
6.5 Beispiel: tokens_VVIMP_1709_1983 . . . . . . . . . . . . . . . . 137
6.6 Beispiel: types_VVFIN_1281_1495 . . . . . . . . . . . . . . . . . 138
6.7 Gruppierte POS-Tags, Verteilung Types, 1281 bis 1495 Tage . . 140
6.8 Gruppierte POS-Tags, Verteilung Types, 1496 bis 1708 Tage . . 141
6.9 Gruppierte POS-Tags, Verteilung Types, 1709 bis 1983 Tage . . 142
6.10 Beispiel: JS_lemmata_ADJA_1709_1983 . . . . . . . . . . . . . 145
6.11 Beispiel: LEO_lemmata_NN_1709_1983 . . . . . . . . . . . . . 146

7.1 Verwendung von „Idee“: 039_MA . . . . . . . . . . . . . . . . . . 201
7.2 „wissen“ im Kontext „weißt du was?“ (042_MK und 044_MK) . . 205
7.3 „wissen“: 020_MM, 036_LAR . . . . . . . . . . . . . . . . . . . . 205
7.4 „glauben“: 024_LUA, 031_AV, 031_LAR, 031_SO . . . . . . . . . 206
7.5 „denken“: 033_JS, 045_MK . . . . . . . . . . . . . . . . . . . . . 206
7.6 Adjektive in den drei Altersgruppen (Ausschnitt) . . . . . . . . . 209
7.7 tag count_NN_1496_1708 . . . . . . . . . . . . . . . . . . . . . . . 217
7.8 tag count_JS_ADV_1709_1983 . . . . . . . . . . . . . . . . . . . . 218
7.9 type token_VVINF_1281_1495 . . . . . . . . . . . . . . . . . . . . 219
7.10 type token lemma_ADJA+ADJD_1709_1983 . . . . . . . . . . . . 220
7.11 lemma tag_VMFIN_1496_1708 . . . . . . . . . . . . . . . . . . . . 221
7.12 lemma bedeutung_v-vorgang_1281_1495 . . . . . . . . . . . . . 221
7.13 type token bedeutung_adv-lok_1709_1983 . . . . . . . . . . . . 222
Tabellenverzeichnis

3.1 Wortartenklassifikation nach Kauschke ([Kauschke (1999)], S. 140) ................................................................. 43
3.2 STTS Tag Table (1995/1996) .................................................. 44
4.2 Gruppe der Testpersonen ...................................................... 88
6.1 POS-Tags zusammengefasst ................................................ 131
6.2 Types und Tokens in den Altersgruppen ................................. 133
6.3 Vorkommen der POS-Tags (Tokens) in allen Altersgruppen im Überblick - gruppiert .............................................. 134
6.4 Vorkommen der POS-Tags (Types) in allen Altersgruppen im Überblick - gruppiert ............................................... 139
6.5 Type-Token-Verhältnis in den Altersgruppen ............................ 140
6.6 MA: Verteilung der Types, gruppiert, 1281 bis 1495 .................. 148
6.7 MK: Verteilung der Types, gruppiert, 1281 bis 1495 ................. 149
6.8 AV: Verteilung der Types, gruppiert, 1496 bis 1708 ............... 150
6.9 LEO: Verteilung der Types, gruppiert, 1496 bis 1708 ............. 151
6.10 LUA: Verteilung der Types, gruppiert, 1496 bis 1708 .......... 152
6.11 MK: Verteilung der Types, gruppiert, 1496 bis 1708 ............. 153
6.12 AV: Verteilung der Types, gruppiert, 1709 bis 1983 .............. 154
6.13 JK: Verteilung der Types, gruppiert, 1709 bis 1983 .............. 155
6.14 JS: Verteilung der Types, gruppiert, 1709 bis 1983 .............. 156
6.15 LAR: Verteilung der Types, gruppiert, 1709 bis 1983 ............ 157
6.16 LEO: Verteilung der Types, gruppiert, 1709 bis 1983 ............ 158
6.17 LL: Verteilung der Types, gruppiert, 1709 bis 1983 .............. 159
6.18 LUA: Verteilung der Types, gruppiert, 1709 bis 1983 ............ 160
6.19 MM: Verteilung der Types, gruppiert, 1709 bis 1983 ............ 161
6.20 RD: Verteilung der Types, gruppiert, 1709 bis 1983 ............. 162
6.21 SO: Verteilung der Types, gruppiert, 1709 bis 1983 ............. 163
7.1 Nomen in den Altersgruppen .............................................. 199
7.2 Verben in den Altersgruppen .............................................. 204
7.3 Adjektive in den Altersgruppen 207
Literaturverzeichnis


[Backscheider und Shatz (1993)] A. Backscheider und M. Shatz. Children’s acquisition of the lexical domain of color. In Beals, K. et al., (Hrsg.), *What we think, what we mean, and how we say it. Papers from the parasession on the correspondence of conceptual, semantic and


[Schmoe (2002)] F. Schmoe. Folglich trat Hubert barfuß und dennoch ungemein heftig gegen die zue Tür - Über einige Eigenschaften der deutschen Adverbien. In P. Wiesinger, (Hrsg.), *Akten des X.*


