Christiane Pankow (Göteborg)
Helena Pettersson (Göteborg)

Auswertung der Leistung von zwei frei zugänglichen POS-Taggern für die Annotation von Korpora des gesprochenen Deutsch

1. Einleitung

In den letzten Jahren ist die Erstellung von linguistischen Korpora in verschiedenen Bereichen der empirischen Sprachanalyse immer wichtiger geworden. Die in der philologischen Forschung bisher üblichen Beispielsammlungen werden inzwischen häufig durch digitalisierte Sprachkorpora ersetzt oder wenigstens ergänzt. Linguistische Korpora sind Datensammlungen, die aus schriftlichen oder gesprochenen Äußerungen in einer oder mehreren Sprachen bestehen. Viele Forschungsprojekte zu sprachlichen Phänomenen in der Linguistik und angrenzenden Fachbereichen wie Psycholinguistik, Soziolinguistik, Fremdsprachendidaktik usw. beginnen damit, dass zuerst ein maschinenlesbares Korpus erstellt wird. Solche Korpora werden dann möglicherweise durch Metadaten und durch linguistische Annotationen ergänzt. Dabei muss man sich im Klaren sein, dass Transkriptionen gesprochener Äußerungen immer Vereinfachungen sind. Das trifft auch für ihre linguistischen Annotationen zu, die jeweils linguistische Interpretationen der gegebenen Daten sind. Im Unterschied zu anderen empirischen Forschungen befindet sich die Entwicklung von Standards für quantitative Sprachanalysen noch in der Diskussion. Obwohl die Kategorisierung von Korpusdaten sowohl theoretisch als auch für eine entsprechende empirische Analyse oft nicht unproblematisch ist, wird in vielen Untersuchungen immer mehr von kategorisierten Korpusdaten ausgegangen, ohne weiter darauf hinzuweisen, nach welchen linguistischen und anderen Maßstäben die Korpusdaten kategorisiert wurden.[1]

Bisher am meisten verbreitet und auch automatisiert ist die Kategorisierung morphosyntaktischer Information, d.h. es handelt sich hierbei um die Markierung der Wortart für jede vorkommende Wortform im Korpus. Diese Wortartenannotation gibt zwar einerseits recht basale Ergebnisse, sie ist aber andererseits ein erster Interpretationsschritt von Korpora mit sehr verschiedenen Untersuchungszielen. Zum Beispiel kann in Lernerkorpora interessant sein, ob eine hohe oder niedrige Anzahl von Verben verwendet wird, in welchem quantitativen Verhältnis Konjunktionen, Adjektive und Präpositionen stehen, ob mehr Substantive als Pronomen verwendet werden usw. Wortartenkategorien sind außerdem nicht nur morphosyntaktische Einheiten, sondern sie sagen auch etwas über semantische und pragmatische Zusammenhänge im Kontext aus. Daher ist die morphosyntaktische Annotation von Korpusdaten die Grundlage vielfältiger korpuslinguistischer Analysen.

Die Wortartenannotation bzw. das Part-of-Speech Tagging (POS-Tagging) bezeichnet die automatische Zuweisung von Wortartenkategorien zu einzelnen Wortformen. Automatische Programme zur Wortartenannotation werden auch Tagger genannt. In der folgenden Untersuchung wird zuerst die Verfügbarkeit von POS-Taggern für das Deutsche aufgezeigt. Danach wird getestet, wie erfolgreich ausgewählte Tagger für eine Annotation transkribierter

[1] Vgl. Lüdeling 2006.

gesprochener Sprache eingesetzt werden können[2]. Für den Test sind zum Vergleich zwei POS-Tagger ausgewählt worden. Das Testkorpus besteht aus zwei transkribierten Aufnahmen aus der Datenbank Gesprochenes Deutsch (DGD)[3] am Institut für deutsche Sprache in Mannheim.

2. POS-Tagging

Bei der Annotierung von Wortarten in einem Korpus kann unterschiedlich vorgegangen werden. Da eine manuelle Annotierung in der Regel sehr zeitaufwendig ist, scheint eine automatische Annotierung mit Hilfe eines Programms von großem Nutzen zu sein. Ein solches Tagging-Programm besteht aus Regeln, nach denen den einzelnen Wortformen im Korpus entsprechende Wortartenkennzeichnungen zugewiesen werden. Bei sämtlichen Tagger-Programmen liegt der Ausgangspunkt der automatischen Zuweisung von Wortarten in einem Lexikon, d.h. in einer Auflistung von Wortformen. In einem ersten Annotierungsschritt wird im Lexikon nachgeschlagen. Dabei kann es bereits geschehen, dass ein Wort nicht vorgefunden wird. Mit Hilfe einer morphologischen Heuristik wird die Wortklasse dann erraten. Das nächste Problem entsteht gewöhnlich dann, wenn bei einer Wortform mehrere Wortklassen möglich sind. Um die jeweils richtige Klasse zu bestimmen, führt das Programm eine Disambiguierung durch. (Vgl. dazu Lemnitzer & Zinsmeister, 2006:72) Die Disambiguierung kann mit verschiedenen Methoden erfolgen. Die Tagger werden – je nach dem, wie sie dieses Problem lösen – in symbolische, stochastische und hybride Tagger eingeteilt, auf deren Arbeitsweise wir hier nicht weiter eingehen wollen. (Vgl. dazu Lemnitzer & Zinsmeister, 2006:73)

2.1 Auswahl der POS-Tagger für die Untersuchung

Die Tagger sind hauptsächlich durch Internet-Suche ermittelt worden. Kriterium für die Suche war, dass die Tagger frei zur Verfügung stehen sollten. Insgesamt sind fünf verschiedene Tagger für das Deutsche gefunden worden:

TreeTagger[4]

TreeTagger ist ein sprachunabhängiger POS-Tagger, der am Institut für maschinelle Sprachverarbeitung (IMS) an der Universität Stuttgart entwickelt wurde. Der Tagger kann kostenlos heruntergeladen werden.

Morphy 3.0[5]

Das Programm wurde an der Universität Paderborn entwickelt. Es ist frei verfügbar, und nur für das Operativsystem Windows zugänglich.

Brill-Tagger

Der Tagger kann kostenlos von der Homepage von Eric Brill[6] heruntergeladen werden. Die Computerlinguistik-Gruppe an der Universität Zürich trainiert den Brill-Tagger für das Deutsche. Der trainierte Tagger kann auf der Homepage getestet werden[7].

---

[2] Die Untersuchung ist ein Teilprojekt von TEXTIL (*Textteknologi i forskning och lärande*), das als fakultätsfinanziertes Projekt an der humanistischen Fakultät der Universität Göteborg entstanden ist. Teilgenommen haben Morten Hunke, Christiane Pankow und Helena Pettersson. Morten Hunke und Helena Pettersson haben jeweils einen Tagger ausgewertet.

[3] http://dsav-oeff.ids-mannheim.de/DSAv/DSAVINFO.HTM (8.4.2006)

[4] http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/DecisionTreeTagger-de.html (5.4.2006)

[5] http://www.wolfganglezius.de/morphy/ (4.4.2006)

[6] http://www.cs.jhu.edu/~brill/ (8.4.2006)

[7] http://www.ifi.unizh.ch/cl/tagger/ (8.4.2006)

Morphix[8]

Morphix kann frei heruntergeladen werden. Die Homepage informiert darüber, dass der Tagger vor allem für Deutsch und Italienisch ausgewertet wurde.

TnT - Statistical Part-of-Speech Tagging[9]

Das Programm wurde an der Universität des Saarlandes in Saarbrücken entwickelt. Für die Verwendung von TnT ist eine kostenlose Lizenz notwendig.

Wir haben uns für die beiden Tagger Morphy und TreeTagger aus verschiedenen, zum Teil praktischen Gründen entschieden. Sämtliche Programme empirisch zu prüfen, wäre für unser Vorhaben zu umfassend gewesen.

2.2 TreeTagger

TreeTagger ist ein stochastischer Tagger. Ein stochastischer Tagger errechnet die Wahrscheinlichkeit einer bestimmten Wortklasse mithilfe von Lexikon und Kontext. (Vgl. Lemnitzer & Zinsmeister, 2006:73) Im Unterschied zu anderen stochastischen Taggern arbeitet TreeTagger mit Beschlussbäumen (decision trees). Der Beschlussbaum bestimmt automatisch die angemessene Größe des Kontexts, die für die Berechnung der Wahrscheinlichkeit der Zuweisung verwendet wird. TreeTagger erreicht bei den englischen Penn-Treebank-Daten eine Korrektheit von 96,36%. (Vgl. Schmid, 1994)

2.3 Morphy

Morphy ist ein Programm sowohl zur Morphologieanalyse (POS-Tagging) als auch zur kontextabhängigen Lemmatisierung. Es hat ein Lexikon von 324 000 Wortformen, das auf 50 500 Stammformen basiert. Das Lexikon ist sehr kompakt, weil nur Stammformen für jedes Wort und dessen Flexionsklassen gespeichert werden. Der jeweilige Benutzer kann dem Lexikon nach und nach neue Wörter hinzufügen. (Vgl. Lezius et al., 1998:1)

Für jedes Wort im Text bestimmt das Analysesystem bei Morphy den Stamm, die Wortart und falls notwendig auch Genus, Kasus, Numerus, Person, Tempus und Komparationsgrad. Gegebenenfalls wird der Kontext berücksichtigt. Falls eine Wortform nicht erkannt werden kann, wird die Wortart durch statistische Methoden erraten. (Vgl. Lezius et al., 1998:2) Der Disambiguator berechnet hier die Wahrscheinlichkeit einer bestimmter Wortform mit Hilfe einer Statistik[10.] (Vgl. Lezius et al., 1998:3) Morphy ist demzufolge ein stochastischer Tagger. Ursprünglich wurde Morphy mit einem großen Tagset morphosyntaktischer Informationen versehen. Um die Fehlerquote zu verringern, wurde auch ein kleineres Tagset durch Ausschluss der morphosyntaktischen Merkmale erstellt. Die Korrektheit beim kleinen Tagset beträgt 96%. (Vgl. Lezius et al., 1998:4)

2.4 Die Tagsets bei TreeTagger und Morphy

Für die Annotierung von Texten ist ein Tagset notwendig. Das Tagset gibt an, mit welchen Kennzeichnungen die Wortformen versehen werden sollen. Als allgemeine Regel gilt, dass genau jede Wortform einen Tag enthält. Neben üblichen Wortformen werden auch Zahlen, Satzzeihen, abgetrennte Wortteile oder Kompositionserstglieder getaggt. (Vgl. Schiller et al.,

[8] http://www.dfki.de/~neumann/morphix/morphix.html (5.4.2006)
[9] http://www.coli.uni-saarland.de/~thorsten/tnt/ (4.4.2006)
[10] Der verwendete Algoritmus ist der Church-Algoritmus, der bei Church (1998) beschrieben wird.

1995:4) Unter einem Tag versteht man demnach die entsprechende Kennzeichnung, mit der die Wortform beim Tagging versehen wird. Die Zusammenstellung der einzelnen Tags macht ein Tagset aus. (Die Tagsets von TreeTagger und Morphy können in der Anlage 1 eingesehen werden.) Die Wortarten-Tags bestehen aus einer Mischung unterschiedlicher Kategorien. Hier spielen positionelle Eigenschaften, syntaktische Funktionen, morphologische Merkmale so wie semantische Merkmale eine Rolle. (Vgl. Lemnitzer & Zinsmeister, 2006:66)

Das STTS-Tagset[11], das in TreeTagger verwendet wird, umfasst 11 so genannte Hauptwortarten. (Vgl. Schiller et al., 1995:6) Die Hauptwortarten und ihre Tags gehen aus der folgenden Übersicht hervor:

Tab. 1: Hauptwortarten aus dem STTS-Tagset

1.	Nomina (N)
2.	Verben (V)
3.	Artikel (ART)
4.	Adjektive (ADJ)
5.	Pronomina (P)
6.	Kardinalzahlen (CARD)
7.	Adverbien (ADV)
8.	Konjunktionen (KO)
9.	Adpositionen (AP)
10.	Interjektionen (ITJ)
11.	Partikeln (PTK)

Diese Hauptwortarten werden unterschiedlich tief subklassifiziert. Die Pronomen werden z.B. in acht Subklassen unterschieden. Insgesamt enthält das STTS-Tagset 54 Tags, davon sind 48 reine POS-Tags. Die sechs zusätzlichen Tags umfassen fremdsprachliches Material, so genannte Nichtwörter und Satzzeichen. (Für eine ausführliche Beschreibung des Tagsets siehe Schiller et al., 1995).

Das Tagset bei Morphy stützt sich auf die Klassifikation der Duden-Grammatik von 1984 (vgl. Lezius, 1996:4). Die Wortarten im Duden (1984) umfassen Verb, Substantiv, Adjektiv, Adverb, Präposition, Konjunktion, Interjektion und Begleiter des Substantivs (Artikel und Pronomen). Partikeln sind in dieser Auflage der Duden-Grammatik keine eigene Wortart. Die Zahlwörter werden als Zahladjektive zu den Adjektiven gezählt. Morphy folgt dieser Einteilung mit den Ausnahmen, dass Artikel und Pronomen für sich bezeichnet und Zahlwörter als eigene Klasse behandelt werden. Bei Morphy sind auch zwei weitere Klassen, nämlich SKZ (Sonderklasse für zu) und ZUS (Verbzusatz) hinzugefügt worden. Auch Abkürzungen und Satzzeichen machen eigene Klassen aus. Das kleine Tagset besteht aus 51 Part-of-Speech-Tags. Das große Tagset besteht aus 456 Tags und kann direkt auf das kleine gelegt werden. (Vgl. Lezius et al., 1996:4)

Die Tagsets bei TreeTagger und Morphy stimmen in vielem miteinander überein. In einigen Wortarten unterscheiden sie sich aber. Bei Morphy wird u.a. zwischen unbestimmten und bestimmten Artikeln unterschieden, diese Unterscheidung wird von TreeTagger nicht vorgenommen. Partikeln als eigene Klasse kommen nur in der TreeTagger-Analyse vor, da diese bei Morphy als Adverbien klassifiziert werden. Die Kategorien Negationspartikel und Antwortpartikel sind also spezifisch für TreeTagger. Das STTS-Tagset bei TreeTagger macht

[11] Das STTS (Stuttgart-Tübingen Tagset) hat sich offensichtlich als Standard für deutschsprachige Korpora etabliert. (Vgl. Lemnitzer & Zinsmeister, 2006:66)

weiter eine Unterscheidung nach Präpositionen in verschiedenen Positionen. Für diese Untersuchung sind jedoch nur zwei Klassen der Präpositionen, die Zirkumposition links und die Präposition mit Artikel aktuell.

Einige der Klassen sind für den jeweiligen Taggingzweck spezifisch. Bespiele hierfür sind die Klassen Verbzusatz und Sonderklasse für *zu*, die in beiden Tagsets vorhanden sind. Der abgetrennte Verbzusatz wird normalerweise als Teil des Verbs betrachtet. Beim automatischen Tagging wird aber jede Wortform für sich beachtet und muss einer Kategorie zugeordnet werden können.

Satzzeichen und Abkürzungen sind in beiden Tagsets vorhanden. Diese werden jedoch in unserer Untersuchung nicht weiter beachtet, sondern sind in der Korrektur wegsortiert worden. (Die Tagsets mit den in dieser Untersuchung vorkommenden Tags aus beiden Programmen werden in Anlage 2 gegenübergestellt.)

2.5 Andere Untersuchungen zu TreeTagger und Morphy

Innerhalb des ETAP-Projekts[12] wurden an der Universität Uppsala die Tagger Morphy und TreeTagger miteinander kombiniert. Die Texte, die als Testmaterial dienten, waren zwei technische Texte und zwei deutsche Übersetzungen einer schwedischen Regierungserklärung. Ein wichtiges Ergebnis dieser Untersuchung besteht darin, dass die Kombination beider Tagger offensichtlich eine höhere Korrektheit aufweist als bei jedem für sich. (Vgl. Bengtsson et al., 2000)

3. Auswertung des POS-Tagger

3.1 Auswahl der Texte

Für die Auswertung wurden zwei Transkripte aus der Datenbank Gesprochenes Deutsch (DGD)[13] ausgewählt. Die Transkripte sind in der Datenbank in verschiedene Gesprächstypen eingeteilt worden. Da wir uns in diesem Projekt vorgenommen haben, spontane gesprochene Sprache zu untersuchen, haben wir den Gesprächstyp ‚Interview' gewählt. In den ausgewählten Transkripten erfolgt die Kommunikation sowohl medial als auch konzeptionell mündlich. Die Gesprächsteilnehmer sprechen frei und gehen also nicht von einem vorliegenden Manuskript aus. Beide Transkripte stammen aus dem Freiburger Teilkorpus. Die Gespräche im Freiburger Korpus wurden im Zeitraum 1960-1974 aufgenommen. Ein kurzer Überblick der Transkripte wird in der Tabelle 2 gegeben.

Tab. 2: Beschreibung der Korpus-Dokumente

Text	Inhalt	Ort	Datum	Anzahl Sprecher
Text 1 (FR015)[14]	Gespräch mit einem Lotsen auf dem Rhein	Bad Godesberg	-	2
Text 2 (FR090)[15]	Aussprache über einen Unfall	-	1970-01-24	2

[12] ETAP = Etablering och annotering av parallellkorpus för igenkänning av översättningsekvivalenter (Creating an annotating a parallel corpus for the recognition of translation equivalents). (Vgl. Bengtsson et al., 2000)
[13] http://dsav-oeff.ids-mannheim.de/DSAv/DSAVINFO.HTM (8.4.2006)
[14] http://dsav-wiss.ids-mannheim.de/DSAv/KORPORA/FR/FR0/FR015/FR015DOK.HTM (8.4.2006)
[15] http://dsav-wiss.ids-mannheim.de/DSAv/KORPORA/FR/FR0/FR090/FR090DOK.HTM (8.4.2006)

Die Anzahl der Tokens und Types in den Auswahltranskripten geht aus der Tabelle 3 hervor.

Tab. 3: Tokens und Types der Korpus-Dokumente

Text	**Tokens**	**Types**
Text 1 (FR015)	2119	640
Text 2 (FR090)	1099	353
Summe Tokens	**3218**	

3.2 Vorgehensweise beim Tagging und Auswertung

Die Transkripte sind von Transkriptionsnotationen gesäubert und als Textfiles von beiden Programmen getaggt worden[16]. Da beide Programme eigene Tokenisierer eingebaut haben, konnten die Text-Dokumente ohne vorbereitende Segmentierungen (vgl. Lemnitzer & Zinsmeister, 2006:64f) eingegeben werden. Bei Morphy wurde das kleine Tagset gewählt, um den Vergleich mit TreeTagger zu erleichtern. Die Tagging-Ergebnisse werden von beiden Taggern wie folgt dargestellt:

Tab. 4: Darstellung der Tagging-Ergebnisse durch die Programme

Morphy:

FR120.SMA

```
: SZD
ja
ja ADV
ja
ja ADV
sie
ich PRO PER PRO
habe
haben VER AUX
ich
ich PRO PER PRO
glaube
glauben VER
ich
ich PRO PER PRO
mindestens
mindestens ADV
mit
mit PRP
vollem
voll ADJ
```

TreeTagger:

FR120wind_taggad.txt

```
:       $.      :
ja      PTKANT  ja
ja      ADV     ja
sie     PPER    sie
habe    VAFIN   haben
ich     PPER    ich
glaube  VVFIN   glauben
ich     PPER    ich
mindestens      ADV     mindestens
mit     APPR    mit
vollem  ADJA    voll
Bewußtsein      NN      Bewußtsein
```

Bei Morphy steht das Token auf einer eigenen Zeile, danach folgt darunter das Lemma und die zugeordnete Wortklasse:

sie (Token)
ich (Lemma) *PRO PER PRO* (POS)

[16] TreeTagger haben wir auf OS X getestet. Wichtig für das Tagging ist, dass das Text-Dokument mit dem Format Windows Latin 1 gespeichert wird. Die Sonderzeichen können sonst von TreeTagger nicht gedeutet werden.

Bei TreeTagger wird das Ergebnis auf einer Zeile präsentiert. Zuerst steht das Token, danach die morphologische Klasse und dann das Lemma:

sie (Token) *PPER* (POS) *ich* (Lemma)

Das Lemma bei den persönlichen Pronomina ist, wie aus dem Beispiel hervorgeht, in beiden Programmen *ich*. Die Art der Lemmatisierung der Wortformen wird in dieser Untersuchung jedoch nicht weiter verfolgt.

Nach dem Tagging mit dem jeweiligen Programm haben wir die Text-Dokumente in ein Excel- Dokument eingefügt[17]. Mit Hilfe der Excel-Dokumente ist dann die Klassifizierung der Programme manuell kontrolliert worden. Nach der Kontrolle wurden die Excelfiles ins Datenbank-Programm Filemaker importiert. Hier sind die Morphy-Dokumente und TreeTagger-Dokumente noch miteinander verglichen worden. Mit Hilfe von Filemaker konnte nun problemlos eine Statistik über die Anzahl der Fehler in den beiden Programmen zusammengestellt werden.

4. Fehleranalyse der Programme

Bei der Analyse der Fehler gehen wir von den nach der Korrektur vorhandenen Wortarten aus und werten pro Wortart die Fehlerquote und Art der Fehler aus. Bei der Analyse von Morphy sind 634 Fehler ermittelt worden, die Fehlerquote bei Morphy beträgt damit 19,7%. (Korrektheitsrate 80,3%). Die Anzahl Fehler beträgt bei Treetagger 462, die Fehlerquote liegt bei 14,3% (Korrektheitsrate: 85,7%). Die Fehlerquoten von sämtlichen Tags kann in der Anlage 2 eingesehen werden. Hier sind auch die Ergebnisse der beiden Programme gegenübergestellt worden. Eine Liste sämtlicher Fehler der beiden Programme liegt in der Anlage 3 vor.

Welche Fehlertypen entstehen nun beim Tagging? Liegen die Fehler in der Zuweisung einer Wortart oder machen die Programme eher Fehler in der Zuweisung einer korrekten Subklasse? Um dies zu beantworten, sind die Fehler nach den jeweiligen Wortarten und ihren Subklassen eingeteilt worden. Unter „Fehler Subklasse“ verstehen wir die Fehler, wo die Subklassifizierung nicht korrekt ist, die Wortart aber richtig getaggt wurde. Das betrifft zum Beispiel solche Fälle, wo das Programm ein Demonstrativpronomen als Relativpronomen getaggt hat. Bei „Fehler Wortart“ hat das Programm eine Klassifizierung außerhalb der Wortart gewählt, zum Beispiel bei den Fällen, wo das Programm ein Verb als Adjektiv getaggt hat.

Neben diesen Fehlertypen wurde auch beachtet, ob der Fehler auf einer sprechsprachlichen Orthographie beruht, z.B. *nicht* in der Schreibweise *nich*. In den Transkripten sind an einigen Stellen Schreibweisen gewählt worden, die dem mündlichen Sprachgebrauch entsprechen sollen. Folgende Formen werden im getaggten Korpus verwendet: *se (sie), s (es), is (ist), isch (ist), n (ein), ne (eine), mein (meine), net (nicht)*. Beide Programme haben mit diesen Schreibweisen deutliche Probleme (siehe Tab. 7).

[17] Bei Morphy bereitete das Einfügen Probleme, da dort das Token auf einer eigenen Zeile steht. Hier mussten wir über die Suche-Ersatz-Funktion in Word gehen, um das Ergebnis auf eine Zeile zu bekommen und damit die Auswertung in Excel zu ermöglichen.

4.1 Analyse der Fehler bei Morphy

In Tabelle 5 wird eine Übersicht der einzelnen Fehlerquoten pro Wortart gegeben. Hier wird gezeigt, wie viele Fehler, die auf einer fehlerhaften Zuweisung der Subklasse bzw. Wortart beruhen, auftreten.

Tab. 5: Fehler pro Wortart bei Morphy

	Fehler	**Tokens**	**Fehler-quote**	**Fehler Subklasse**	**Fehler Wortart**	**Fehler-quote Wortart**
ADJEKTIVE	22	141	15,60%	2	20	14,18%
ADVERBIEN	73	543	13,44%	13	60	11,05%
ARTIKEL	12	303	3,96%	0	12	3,96%
INTERJEKTIONEN	4	4	100,00%	0	4	100,00%
KONJUNKTIONEN	18	215	8,37%	3	15	6,98%
PRONOMEN	216	535	40,37%	66	150	28,04%
PRÄPOSITIONEN	12	272	4,41%	0	12	4,41%
EIGENNAMEN	58	93	62,37%	0	58	62,37%
SUBSTANTIVE	22	474	4,64%	0	22	4,64%
VERBEN	158	571	27,67%	95	63	11,03%
VERBZUSATZ	17	21	80,95%	0	17	80,95%
ZU-KLASSE	2	10	20,00%	0	2	20,00%
ZAHLWORT	20	36	55,56%	0	20	55,56%
INSGESAMT:	**634**	**3218**	**19,70%**		**455**	**14,14%**
KORREKTHEITS-RATE:			**80,30%**			**85,86%**

4.1.1 Adjektiv

Bei den Adjektiven beträgt die Fehlerquote in den Subklassen ‚Adjektivformen' (ADJ) 0% und bei den ‚adverbiellen Adjektiven' (ADJ ADV) 31%. Es handelt sich hier vor allem um Fehler in der Zuweisung der Wortart. Nur zwei der Fehler sind Zuweisungen der Subklasse, d.h. (ADJ) statt (ADJ ADV). Bei den ‚adverbiellen Adjektiven' treten 22 Fehler auf, davon sind 14 Tokens, bei denen vom Programm die Klassifizierung Verbzusatz (ZUS) gewählt wurde, vor allem beim Lexem *klar*:

(1) S1: sondern entscheidend is daß überhaupt jemand da drin gesessen ist der gefährdet wurde net
S2: ja ja ja das isch s genau das isch klar [ZUS => ADV ADJ][18]
(FR090)

Insgesamt beträgt die Fehlerquote bei den Adjektiven 15,60%. Betrachten wir Adjektive als eine Wortart ohne Subklassifizierung ist die Fehlerquote etwas geringer, sie beträgt 14,18%.

4.1.2 Adverb

Zur Wortart Adverbien werden ‚Adverbien' (ADV) und ‚Pronominaladverbien' (PRO ADV) gezählt. Die Fehlerquote bei den ‚Adverbien' beträgt 12,50% und bei den ‚Pronominaladverbien' 46,67%. Zu den Adverbien sind viele Lexeme als ‚adverbielle Adjektive' klassifiziert worden und in der Ergebnisübersicht als Wortartenfehler kategorisiert worden. Insgesamt beträgt die Fehlerquote ohne Subklassifizierung 11,05%.

4.1.3 Artikel

[18] In der Beschreibung der Fehler wird die ausgeführte Korrektur innerhalb von eckigen Klammern angegeben.

Bei den Artikeln macht Morphy eine Einteilung in ‚bestimmte und unbestimmte Artikel'. Bei den ‚bestimmen Artikeln' (ART DEF) beträgt die Fehlerquote nur 0,88% und bei den ‚unbestimmten Artikeln' (ART IND) 13,16%. Insgesamt beträgt für die Wortart Artikel die Fehlerquote nur 3,96%. Hier beruhen sämtlicher Fehler auf einer fehlerhaften Zuweisung der Wortart.

4.1.4 Interjektion

Interjektionen kommen im Korpus sehr selten vor; die Fehlerquote von 100% bei vier Vorkommen besagt deshalb kaum etwas. Morphy's Lexikon könnte hier leicht durch weitere Interjektionen ergänzt werden, falls es für ein entsprechendes Korpus von Bedeutung wäre.

4.1.5 Konjunktion

Konjunktionen haben im Tagset bei Morphy vier Subklassen. Die Mehrheit der Konjunktionen sind ‚nebenordnende Konjunktionen' (141 Vorkommen). Bei der Zuweisung dieser Subklasse macht das Programm keine Fehler. Bei den ‚unterordnenden Konjunktionen' beträgt die Fehlerquote 12,50%. ‚Infinitivkonjunktionen' (KON INF) und ‚Vergleichskonjunktionen' (KON VGL) kommen sehr selten im Korpus vor, hier sind aber sämtliche Vorkommen falsch getaggt worden. Insgesamt beträgt die Fehlerquote bei den Konjunktionen 8,37%. Sehen wir von Fehlern in der Subklassifizierung ab, beträgt die Fehlerquote 6,98%.

4.1.6 Pronomen

Bei den Pronomen gibt das Tagset bei Morphy 12 Subklassen an. Die meisten Fehler finden wir bei den Demonstrativpronomen, die Fehlerquote beträgt hier 100%. Bei 73,68% (70 Vorkommen) der Vorkommen wird der Artikel statt Pronomen gewählt.

(2) das is ihr gutes Recht das [ART => PRO DEM PRO] kann ihnen niemand verübeln
(FR090)

Bei den Demonstrativpronomen wurden 18 Wortformen als Personalpronomen (PRO PER PRO) klassifiziert.

(3) S2: [...] und diese Mädchenlehen wurde die Mädchen die in dem Jahr achtzehn Jahre alt wurden, öffentlich meistbietend versteigert das [PRO PER => PRO DEM PRO] war so gesagt
(FR015)

Bei den Relativpronomen (PRO REL PRO) liegt die Fehlerquote bei 76,19%. Auch hier werden viele Pronomen (30 Vorkommen) als Artikel (ART DEF) getaggt.

(4) sie sind doch sicher auch Lotse für Passagierschiffe die [ART => PRO REL PRO] hier durchkommen nich
(FR015)

Auch bei den Indefinitpronomen (PRO IND PRO) sind 21 Fehler entdeckt worden. Die Fehlerquote beträgt 44,68%. Hier wird u.a. *was* achtmal fehlerhaft als Relativpronomen annotiert:

(5) $S1: gut ja sie wissen jetzt kommen wir also zur eigentlichen Sache
$S2: ja ja
$S1: sie wissen um was [PRO REL PRO => PRO IND PRO] es geht
(FR090)

Bei den Personalpronomen (PRO PER PRO) überwiegen Fehler, die auf der Schreibweise der Wortformen beruhen: hier bei *s* (*es*), *se* (*sie*) und *i* (*ich*). Insgesamt beträgt die Fehlerquote bei den Pronomen 40,37%. Sehen wir aber von der Subklassifizierung ab, beträgt die Fehlerquote 28,04%. In vielen Fällen wird also eine falsche Subklasse gewählt.

4.1.7 Präposition

Bei den Präpositionen (PRP) ist die Fehlerquote niedrig, sie liegt nur bei 4,41%. Diese Fehler sind teils als ZUS und teils als ADV getaggt worden.

4.1.8 Eigennamen

Bei den Eigennamen handelt es sich um Tokens, die dem Programm nicht bekannt sind. Die Fehler werden deshalb auch mehrmals wiederholt: *Lorelei* wird acht Mal als Substantiv klassifiziert, *Sankt Goar* wird hingegen achtmal als Verb oder Substantiv klassifiziert und *Schwarz* wird elfmal als ‚adverbielles Adjektiv' klassifiziert. Die Fehlerquote beträgt bei den Eigennamen 62,37%.

4.1.9 Substantiv

Bei den Substantiven (SUB) beträgt die Fehlerquote 4,64%. Auch unter den Substantiven handelt es sich zum Teil um im Lexikon nicht vorhandene Lexeme: *Lotse* wird z.B. sechsmal als Verb klassifiziert.

4.1.10 Verb

Bei den Verben beträgt die Fehlerquote 27,67%, ohne Subklassifizierung 11,03%. Bei den finiten Vollverben (VER) sind 52 Fehler (35,62%) korrigiert worden. 15-mal werden die finiten Verben als Infinitive getaggt:

(6) S1: gut ja sie wissen [VER] jetzt kommen [VER] wir also zur eigentlichen Sache
S2: ja ja
S1: sie wissen [VER INF => VER] um was es geht
(FR090)

Im Beispiel (7) wird beim ersten Vorkommen von *wissen* richtig getaggt; beim zweiten Mal jedoch als ‚infinit' getaggt. Finite Vollverben werden auch 14-mal als Perfektpartizip getaggt:

(7) wenn man überlegt [VER PA2> VER] daß ein Schubschiff bis zu sechstausend Tonnen in vier Backs befördert [VER PA2=> VER] und nur eine Länge aufweist von circa zweihundert Metern gegenüber der Länge von einem Kilometer in der Schleppschiffahrt wo auch nur sechstausend Tonnen befördert wurde [...]
(FR015)

Die finiten Vollverben werden auch neunmal als Possessivpronomen (PRO POS ATT) getaggt, es handelt sich hier um die Wortform *mein*. Hier ist es die sprechsprachliche Orthographie, die vom Programm nicht beherrscht wird.

(8) $S2: dann ist eine Gefährdung trotzdem gegeben gewesen ich muß sie darauf hinweisen ich mein [PRO POS ATT => VER] an ihrem Fahrzeug ist ja in erster Linie nur Sachschaden entstanden am LKW ist weiter nichts eingetreten soweit ich s aus den Akten entnehme.
(FR090)

Die Fehlerquote beträgt bei den Infinitiven (VER INF) 28,57%. Die 14 korrigierten Fehler bestehen aus finiten Vollverben (VER).

(9) ja dann wird die Sache auch nicht so nicht so kraß für sie abgehen . ich darf ihnen nur das eine sagen [VER => VER INF] natürlich geht es nicht mit zwanzig Mark
(FR090)

Unter den Perfektpartizip-Vorkommen (VER PA2) werden 31 von insgesamt 92 Vorkommen falsch getaggt. Von den 31 Fehlern werden 26 als Adjektiv klassifiziert. Beispiele hierfür sind:

(10) und sie sind von der Polizei angezeigt [ADJ => VER PA2] worden
(FR090)

(11) dann kennen sie wahrscheinlich auch haargenau all das was hier passiert [ADJ => VER PA2] ist und was sich ereignet hat [ADJ => VER PA2]
(FR015)

Die Hilfsverben (VER AUX) wurden 42-mal korrigiert, die Fehlerquote beträgt 18,75%. Hier handelt es sich ausschließlich um Lexeme in einer sprechsprachlichen Schreibweise, d.h. die Formen *is*, *isch* und *hab* werden vom Programm falsch klassifiziert.

(12) da is [VER PA2 => VER AUX] nichts zu machen für sie
(FR090)

(13) von Bingen bis nach Sankt Goar das is [ADJ => VER AUX] rheinabwärts und Sankt Goar bis Bingen ist rheinaufwärts
(FR015)

Die Klassifizierung Hilfsverb als Perfektpartizip (VER AUX PA2) scheint vom Programm nicht beherrscht zu werden. Hier werden sämtliche neun Vorkommen als ‚Vollverb, Perfektpartizip' getaggt.

(14) [...] dann ist eine Gefährdung trotzdem gegeben gewesen [VER PA2 => VER AUX PA2] ich muß sie darauf hinweisen ich mein an ihrem Fahrzeug ist ja in erster Linie nur Sachschaden entstanden [...]

Verbzusätze werden vom Programm als eigene Klasse aufgestellt. Die Fehlerquote beträgt hier 80,95%. Die Wortformen wurden fälschlich als PRP oder ADV getaggt.

4.1.11 Zahlwort

Bei den ‚Zahlwörtern' (ZAL) hat Morphy jedoch Probleme. Im 14 von 20 Fällen taggt Morphy das ‚Zahlwort' als Adjektiv.

(15) die Lorelei wurde auch einmal achtzehn [ADJ => ZAL] Jahre alt und wurde auch öffentlich meistbietend versteigert

4.2 Analyse der Fehler bei TreeTagger

In Tabelle 6 wird eine Übersicht der Fehler pro Wortart gegeben.

Tab. 6: Fehler pro Wortart bei TreeTagger

	Fehler	**Tokens**	**Fehler-quote**	**Fehler Subklasse**	**Fehler Wortart**	**Fehler-quote Wortart**
ADJEKTIVE	4	141	2,84%	1	3	2,13%
ADVERBIEN	105	543	19,34%	23	82	15,10%
ARTIKEL	11	303	3,63%	0	11	3,63%
INTERJEKTIONEN	3	4	75,00%	0	3	75,00%
KONJUNKTIONEN	10	215	4,65%	4	6	2,79%
PRONOMEN	133	535	24,86%	23	110	20,56%
PRÄPOSITIONEN	9	272	3,31%	1	8	2,94%
EIGENNAMEN	22	93	23,66%	0	22	23,66%
SUBSTANTIVE	13	474	2,74%	0	13	2,74%
VERBEN	129	571	22,59%	69	60	10,51%
VERBZUSATZ	15	21	71,43%	0	15	71,43%
ZU-KLASSE	1	10	10,00%	0	1	10,00%
ZAHLWÖRTER	7	36	19,44%	0	5	13,89%
INSGESAMT:	**462**	**3218**	**14,36%**		**339**	**10,53%**
KORREKTHEITS-RATE:			**85,64%**			**89,47%**

4.2.1 Adjektiv

Bei den Adjektiven macht TreeTagger bei den ‚adverbiellen Adjektiven' nur vier Fehler. Die Fehlerquote beträgt für die Adjektive insgesamt 2,84%.

4.2.2 Adverb

Zu den Adverbien sind auch ‚Negationspartikeln' (PTKNEG), ‚Antwortpartikeln' (PTKANT), ‚adverbielle Interrogativ- oder Relativpronomina' (PWAW) und ‚Pronominaladverbien' (PAV) gerechnet worden. Bei den Adverbien (ADV) sind 67 Fehler festgestellt worden, die Fehlerquote beträgt 17,54 %. Bei den Fehlern handelt es sich bei 47 Vorkommen um Adjektive.

‚Antwortpartikeln' (PTKANT) sind nach den Anweisungen von Schiller et al. (1995) „ [...] die Wortformen *ja*, *nein*, *danke*, *bitte*, die im Allgemeinen nur in direkter Rede vorkommen und dann alleine einen Satz bilden oder in einem Antwortsatz als Bejahung, Verneinung oder Verstärkung verwendet werden." (Schiller et al., 1996: 69) In der Korrektur sind wir aber so vorgegangen, dass auch die Verwendungen von *ja* und *nein* als ‚Gliederungspartikel' PTKANT gebilligt worden sind. *Ja* wird besonders im Korpus FR090 sehr häufig als Gliederungssignal eingesetzt, wie im folgenden Abschnitt sichtbar wird:

(16) S2: ja jetzt hab ich zuerst noch ne Frage
S1: bitte
S2: ob sie mir sagen können was gegen mich vorliegt wie
S1: nein nein das steht.
S2: ich mein das steht da drauf und
S1: ja
S2: zu dem kann ich mich äußern.
S1: ja das mein ich sonst nichts
S2: ja und das möcht ich ja ja ja
S1: ja gut dann is es also klar .sie wollen also dann aussagen
S2: ja
S1: gut ja sie wissen jetzt kommen wir also zur eigentlichen Sache)
S2: ja ja
S1: sie wissen um was es geht
(FR090)

Das Programm kann sämtliche Vorkommen von *ja* als ‚Abtönungspartikel' erkennen, diese sind dann als Adverb getaggt worden. Jedoch werden auch einige der ‚Antwortpartikeln' als Adverb getaggt. Die Fehlerquote bei den ‚Antwortpartikeln' beträgt 23,33%. Als ‚Negationspartikel' wird nur die Wortform *nicht* getaggt. Dem Programm gelingt es aber nicht, die Varianten *nich* und *net* korrekt zu taggen, hier sind deshalb Korrekturen vorgenommen worden, die die Fehlerquote 28,6% ergeben haben. Insgesamt beträgt die Fehlerquote bei den Adverbien 19,34% mit Subklassifizierung. Ohne Subklassifizierung beträgt die Fehlerquote 15,10%.

4.2.3 Artikel

Bei den Artikeln ist die Fehlerquote gering; nur 3,63% der Vorkommen sind falsch bestimmt.

4.2.4 Interjektion

Da Interjektionen nur viermal im Korpus vorkommen, ist die Fehlerquote von 75% nicht von Belang.

4.2.5 Konjunktion

Die Fehlerquote bei den Konjunktionen beträgt insgesamt 4,65%. Ohne Subklassifizierung würde die Fehlerquote bei 2,79% liegen. Bei den ‚Vergleichskonjunktionen' (KOKOM) hat TreeTagger weniger Probleme als Morphy. Hier wurden sieben von acht Vorkommen korrekt klassifiziert.

4.2.6 Pronomen

Bei den Pronomen hat TreeTagger Probleme mit den Demonstrativpronomen (PDS). Die Fehlerquote beträgt 44,21%. Von den 42 Fehlern unter den Demonstrativpronomen wurden 34 als Artikeln getaggt. Auch bei den Relativpronomen (PRELS) wird der Tag ART bevorzugt. Unter den 39 Fehlern bei den Relativpronomen sind 30 als Artikel annotiert. Die Fehlerquote bei den Relativpronomen liegt bei 92,86%. Bei den Personalpronomen (PPER) überwiegen Fehler, die auf der sprechsprachlichen Orthographie beruhen, hier durch *s (es), se (sie)* und *i (ich)*. Die Fehlerquote beträgt 15,77%. Bei den Indefinitpronomen schneidet aber TreeTagger besser ab als Morphy. Die Fehlerquote bei den substituierenden Indefinitpronomen beträgt 14,89%. Insgesamt beträgt sie bei den Pronomen 24,86%, ohne Subklassifizierung bei 20,56%.

4.2.7 Präposition

Bei der Annotierung von Präpositionen hat TreeTagger keine großen Probleme, die Fehlerquote beträgt hier nur 3,31% mit Subklassifizierung.

4.2.8 Eigennamen

Bei den Eigennamen hat Treetagger weniger Probleme als Morphy. Zum Beispiel erkennt Treetagger Schwarz im Unterschied zu Morphy als Eigenname. Die Fehlerquote liegt hier bei 23,66%.

4.2.9 Substantiv

Die Fehlerquote ist hier sehr gering, sie liegt bei 2,74%.

4.2.10 Verb

Bei den Verben werden Infinitive (VVINF) 31-mal falsch getaggt, hiervon werden 29 Vorkommen als ‚finites Verb' analysiert: Fehlerquote 63,27%.

(17) [...] wir müssen immer zur Verfügung stehen [VVFIN => VVINF]
(FR015)

Auch bei den finiten Verben kommen Fehler vor. Die Anzahl Fehler beträgt 22, die Fehlerquote liegt bei 15,07%. Bei den Fehlern handelt es sich um Verbstämme, die vom Programm nicht verstanden werden, wie z.B. die sprechsprachlichen Varianten *mein, mach*. Bei den ‚finiten Hilfsverben' (VAFIN) handelt es sich um Fehler, die aufgrund der Orthographie entstanden sind.

(18) ja das is [NN => VAFIN] ihr gutes Recht das kann ihnen niemand verübeln
(FR090)

Auch bei TreeTagger kommen Fehler bei den Perfektpartizipien (VVPP) vor. Bei 23 Fehlern wird vorwiegend der Tag VVFIN gewählt. Die Fehlerqoute beträgt hier 25%.

(19) das is auch weiter nicht schlimm nur noch eine Frage die ich im Moment von Flensburg noch nicht zurück habe, haben sie schon Vorstrafen oder Ordnungswidrigkeiten in Flensburg eingetragen bekommen [VVFIN =>VVPP]
(FR090)

Insgesamt liegt die Fehlerquote bei den Verbformen bei 22,24%. Sehen wir von der Subklassifizierung ab, beträgt die Fehlerquote nur 10,16%. Bei den Verbzusätzen liegt die Fehlerquote bei 71,43%. Hier werden die Wortformen vorwiegend als Präposition oder Adverb getaggt.

4.2.11 Zahlwort

Bei den Zahlwörtern liegt die Fehlerquote bei 19,44%. Hier schneidet TreeTagger besser ab als Morphy.

5. Zusammenfassung der Fehler beim Tagging

Beide Tagger haben Probleme, zwischen verschiedenen Pronomen und Artikeln zu unterscheiden. Die Zuweisung als Demonstrativpronomen scheint besonders schwierig zu sein. Morphy verwendet die Klassifizierung PRO DEM PRO kein einziges Mal. TreeTagger schneidet hier etwas besser ab. Eine Erklärung für die Schwierigkeiten bei den Pronomen könnte sein, dass der Satzbau in der spontan gesprochenen Sprache anders aussieht als in der geschriebenen Sprache. Im Beispiel (20) wird *das* bei Morphy als Artikel getaggt. Bei TreeTagger wird *das* hingegen in den drei Vorkommen im Ausschnitt als Demonstrativpronomen annotiert.

(20) S2: ob sie mir sagen können, was gegen mich vorliegt, wie
S1: nein nein das steht.
S2: ich mein das steht da drauf und
S1: ja
S2: zu dem kann ich mich äußern.
S1: ja das mein ich sonst nichts
(FR090)

Auch bei den Relativpronomen haben die Programme Probleme. Sämtliche Relativpronomen (*der, die, das*) werden bei Morphy als Artikel getaggt. Eine mögliche Erklärung wäre, dass das Programm ein Komma vor einem Relativpronomen erwartet. Auch bei TreeTagger werden mit einer Ausnahme sämtliche Relativpronomen falsch annotiert. Im Beispiel (21) kann die fehlerhafte Zuweisung von *der* auch auf Besonderheiten der Gesprächssequenzen beruhen. Die Äußerungen von S1 überlappen sich hier mit den Äußerungen von S2. Im Transkript sind die Äußerungen jedoch aneinander gereiht, was bedeutet, dass die Aussage von S1 zweimal unterbrochen wird und dies dem Tagging-Programm Schwierigkeiten bereitet.

(21) $S1: also allein schon dadurch die herrschende Rechtsprechung sagt in dem Moment wo ein Kraftfahrer durch das Verhalten eines andern
$S2: ja das ist mir völlig klar
$S1: der schuldhaft handelt
$S2: ja ja
$S1: gezwungen wird zu bremsen in dem Moment tritt schon ein Gefährdung ein

Bei der Annotierung von Possessivpronomen treten Fehler aber nur bei Morphy auf.

Bei den Verben haben beide Programme Schwierigkeiten, zwischen finiten und infiniten Formen zu unterscheiden. Morphy annotiert finite Verben als ‚infinit' oder Perfektpartizip. Diese Fehler kommen in TreeTagger nicht vor. TreeTagger dagegen annotiert die Infinitive oft als finite Verben. Dieser Fehlertyp kommt zwar auch im Morphy vor, ist aber häufiger bei TreeTagger, wie z.B. im folgenden Ausschnitt:

(22) Sie können ohne Anwalt die Sache nicht bearbeiten oder ohne Anwalt nicht durchstehen

Hier wird *durchstehen* bei Morphy richtig getaggt, bei TreeTagger jedoch nicht.

Weitere häufige Fehler bestehen darin, dass Morphy oft ein Perfektpartizip als Adjektiv taggt, TreeTagger dagegen die Perfektpartizipien eher als finite Vollverben. Beide Programme klassifizieren Adverbien als Adjektive. Auch bei den Verbzusätzen haben beide Programme Probleme.

Vergleichen wir die beiden Programme, ist die Fehlerquote in sämtlichen Annotationsklassen bei TreeTagger niedriger als bei Morphy, mit Ausnahme von infiniten Vollverben und finiten Hilfsverben.

Die Programme haben auch mit unbekannten Lexemen Probleme. Hier handelt sich teils um Eigenamen, Morphy hat damit größere Probleme als TreeTagger, teils sind es Lexeme, die

in sprechsprachlicher Orthographie transkribiert wurden, z.B. *is*, *ne*, *net*. In Tabelle 7 ist ein Überblick dieses Phänomens zusammengestellt worden. Hier wird gezeigt, bei welchen Lexemen an die Aussprache angepasste Schreibweisen auftreten.

Tab. 7: Sprechsprachliche Orthographie bei Morphy und TreeTagger

	Tokens	**Fehler Morphy**	**Fehler TreeTagger**	**Korrekte Klassifizierung Morphy**	**Korrekte Klassifizierung TreeTagger**
n (ein)	5	5	5	ART IND	ART
ne (eine)	3	3	0	ART IND	ART
se (sie)	5	5	5	PRO PER	PPR
s (es)	24	20	24	PRO PER	PPR
i (ich)	3	3	3	PRO PER	PPR
mein (meine)	9	9	9	PRO POS ATT, VER	PPOSAT, VVFIN
is (ist)	28	24	28	VER AUX	VAFIN
isch (ist)	13	12	13	VER AUX	VAFIN
hab (habe)	5	4	5	VER AUX	VAFIN
net	11	11	11	ADV	ADV
SUMME:	**106**	**96**	**103**		

Fassen wir die Fehlerquellen zusammen, können drei Gruppen festgestellt werden:

1. Sprechsprachliche Orthographie: In beiden Programmen treten Probleme auf, wenn Wortformen getaggt werden müssen, die nicht schriftsprachlich kodifiziert sind.

2. Besonderheiten des Satzbaus der gesprochenen Sprache: Die Programme können u.a. die dialogische Form der gesprochenen Sprache nicht ausreichend gut bewältigen. Ein Sprecher kann z.B. einen Neuanfang in seiner Äußerung machen oder vom Partner unterbrochen werden wie im Ausschnitt (23), wo sich der Sprecher in *über äußern* selbst korrigiert. Diese Äußerungsstruktur weicht zu stark von der Satzstruktur der Schriftsprache ab, was sich in den Annotationen widerspiegelt.

(23) $S2: ich mein ich kann au net viel dazu äußern über äußern ich wollt bloß ich hab jetzt bloß gedacht daß seien jetzt irgendwelche andere Personen da waren nämlich weit und breit waren einfach keine Personen

3. Unspezifische Fehlerquellen: Ohne genauere Kenntnisse darüber, wie die Programme technisch arbeiten, ist es in einigen Fällen schwer, die Ursache der Fehler zu ermitteln.

5.1 Möglichkeiten zur Erlangung höherer Korrektheitsraten

Die Korrektheitsraten in den untersuchten Teilkorpora betragen bei TreeTagger 85,7% und bei Morphy 80,3%. Diese Raten liegen unter den Angaben, die in der Dokumentation der Programme zu finden sind. Die niedrigeren Korrektheitsraten beruhen wahrscheinlich darauf, dass die Tagger nicht auf unserem Probekorpus trainiert wurden. Auch der Kommunikationstyp ‚spontan gesprochene Sprache' bereitet Probleme, weil der Satzbau größtenteils nicht den Regelmäßigkeiten der Schriftsprache folgt.

Die Korrektheitsraten sind auf dem ersten Blick zu niedrig, um ohne manuelle Kontrolle verwendet werden zu können. Gibt es nun Möglichkeiten den Korrektheitsgrad zu erhöhen?

Eine Möglichkeit, die schon in der Fehleranalyse dargestellt worden ist, besteht darin, die Anzahl der Tags zu verringern und nur eine Einteilung in Wortarten vorzunehmen. Die Subklassifizierung lässt man einfach beiseite. Insgesamt weist dann Morphy die Korrektheitsrate 85,9% und TreeTagger die Rate 89,6% auf.

Tab. 8: Korrektheitsraten mit und ohne Subklassifizierung

	mit Subklassifizierung	**ohne Subklassifizierung**
Morphy	80,3% (19,7%)	85,9% (14,1%)
TreeTagger	85,6% (14,4%)	89,5% (10,5%)

Aus der Tabelle 8 geht hervor, dass die Korrektheitsraten bei Ausschluss der Subklassen erheblich erhöht werden können.

Selbstverständlich sind auch weitere Zusammenlegungen möglich, z.B. können die Artikel mit den Pronomen gruppiert werden oder eine Zusammenlegung von Eigennamen und Substantiven ist möglich.

Eine weitere Möglichkeit, die Korrektheitsrate zu erhöhen, wäre eine Erweiterung des Lexikons der Tagging-Programme durch neue Lexeme. In Morphy kann das Lexikon ohne große Umstände ausgebaut werden. Man könnte dann die sprechsprachlichen Formen eingeben. Die Korrektheitsrate würde dann bei Morphy 92,1% betragen. Bei TreeTagger besteht keine Möglichkeit neue Lexeme hinzuzufügen.

Inwiefern ein automatisches Tagging mit den verwendeten Programmen sinnvoll ist, hängt daher zum großen Teil davon ab, welche Ergebnisse man vom Tagging erwartet. Bei einer gröberen Einteilung in Wortarten ohne Subklassifizierung wäre die Fehlerquote wahrscheinlich für einen gegebenen Untersuchungszweck gering genug.

Literatur

Bengtsson, Camilla, Lars Borin and Henrik Oxhammar (2000): Comparing and combining part-of-speech taggers for multilingual parallel corpora. In: Working papers in Computational Linguistics & Language Engineering 22. Department of Linguistics, Uppsala University. S. 11-30.

Church, K.W. (1998): A stochastic parts program and noun phrase parser for unrestricted text. In: Second Conference on Applied Natural Language Processing, Austin, Texas. S. 136-143.

Duden. Grammatik der deutschen Standardsprache (1984). (Hrsg.) Drosdowski, Günter. Duden Band 4.

Lemnitzer, Lothar und Heike Zinsmeister (2006): Korpuslinguistik. Eine Einführung. Narr Studienbücher. Tübingen.

Lezius, Wolfgang, Reinhard Rapp und Manfred Wettler (1998): A Freely Available Morphological Analyzer, Disambiguator, and Context Sensitive Lemmatizer for German in Proceedings of the COLING-ACL 1998 pp. 743-747.

Lezius, Wolfgang; Rapp, Reinhard; Wettler, Manfred (1996): A Morphology-System and Part-of-Speech Tagger for German. In: D. Gibbon (ed.), Natural Language Processing and

Speech Technology. Results of the 3rd KONVENS Conference. Mouton de Gruyter, pp. 369-378.

Lezius, Wolfgang (1998): Die Wortklassensysteme von Morphy (Vollständiges Klassensystem, großes und kleines Tag Set.)
http://www.wolfganglezius.de/doku.php?id=public:cl:morphy

Lüdeling, Anke (2006): Das Zusammenspiel von qualitativen und quantitativen Methoden in der Korpuslinguistik. In: Zifonun, Gisela und Wernder Kallmayer (Hrsg.): IDS-Jahrbuch 2006. Im Druck.

Schiller, Anne, Simone Teufel und Christine Stöckert (1995): Vorläufige Guidelines für das Tagging deutscher Textcorpora. Universität Stuttgart. Institut für maschinelle Sprachverarbeitung.

Schmid, Helmut (1994): Probabilistic Part-of-Speech Tagging Using Decision Trees. In: Proceedings of the International Conference on New Methods in Language Processing, Manchester, UK, pp. 44-49.

Anlage 1: Tagsets von Morphy und TreeTagger

a) Morphy (Lezius, 1998)

Table 3: The small tag set (51 tags)

tag name	explanation of the tag	example
SUB	Substantiv	*(der) Mann*
EIG	Eigenname	*Egon, (Herr) Hansen*
VER	finite Verbform	*spielst, läuft*
VER INF	Infinitiv	*spielen, laufen*
VER PA2	Partizip Perfekt	*gespielt, gelaufen*
VER EIZ	erweiterter Infinitiv mit *zu*	*abzuspielen*
VER IMP	Imperativ	*lauf', laufe*
VER AUX	finite Hilfsverbform	*bin, hast*
VER AUX INF	Infinitiv	*haben, sein*
VER AUX PA2	Partizip Perfekt	*gehabt, gewesen*
VER AUX IMP	Imperativ	*sei, habe*
VER MOD	finite Modalverbform	*kannst, will*
VER MOD INF	Infinitiv	*können, wollen*
VER MOD PA2	Partizip Perfekt	*gekonnt, gewollt*
VER MOD IMP	Imperativ	*könne*
ART IND	unbestimmter Artikel	*ein, eines*
ART DEF	bestimmter Artikel	*der, des*
ADJ	Adjektivform	*schnelle, kleinstes*
ADJ ADV	Adjektiv, adverbiell	*(Er läuft) schnell.*
PRO DEM ATT	Demonstrativpronomen, attributiv	*diese (Frau)*
PRO DEM PRO	Demonstrativpronomen, pronominal	*diese*
PRO REL ATT	Relativpronomen, attributiv	*,dessen (Frau)*
PRO REL PRO	Relativpronomen, pronominal	*,welcher*
PRO POS ATT	Possesivpronomen, attributiv	*mein (Buch)*
PRO POS PRO	Possesivpronomen, pronominal	*(Das ist) meiner.*
PRO IND ATT	Indefinitpronomen, attributiv	*alle (Menschen)*
PRO IND PRO	Indefinitpronomen, pronominal	*(Ich mag) alle.*
PRO INR ATT	Interrogativpronomen, attributiv	*Welcher (Mann)?*
PRO INR PRO	Interrogativpronomen, pronominal	*Wer?*
PRO PER	Personalpronomen	*er, wir*
PRO REF	Reflexivpronomen	*sich, uns*
ADV	Adverb	*schon, manchmal*
ADV PRO	Pronominaladverb	*damit, dadurch*
KON UNT	unterordnende Konjunktion	*daß, da*
KON NEB	nebenordnende Konjunktion	*und, oder*
KON INF	Infinitivkonjunktion	*um (zu spielen)*
KON VGL	Vergleichskonjunktion	*als, denn, wie*
KON PRI	Proportionalkonjunktion	*desto, um so, je*
PRP	Präposition	*durch, an*
SKZ	Sonderklasse für *zu*	*(,um) zu (spielen)*
ZUS	Verbzusatz	*(spielst) ab*
INJ	Interjektion	*Wau, Oh*
ZAL	Zahlwörter	*eins, tausend*
ZAN	Zahlen	*100, 2*
ABK	Abkürzung	*Dr., usw.*
SZD	Doppelpunkt	:
SZE	Satzendezeichen	.!?
SZG	Gedankenstrich	-
SZK	Komma	,
SZS	Semikolon	;
SZN	sonstige Satzzeichen	()/

b) TreeTagger (Schiller et al, 1995)

POS =	Beschreibung	Beispiele
ADJA	attributives Adjektiv	*[das] große [Haus]*
ADJD	adverbiales oder prädikatives Adjektiv	*[er fährt] schnell* *[er ist] schnell*
ADV	Adverb	*schon, bald, doch*
APPR	Präposition; Zirkumposition links	*in [der Stadt], ohne [mich]*
APPRART	Präposition mit Artikel	*im [Haus], zur [Sache]*
APPO	Postposition	*[ihm] zufolge, [der Sache] wegen*
APZR	Zirkumposition rechts	*[von jetzt] an*
ART	bestimmter oder unbestimmter Artikel	*der, die, das,* *ein, eine*
CARD	Kardinalzahl	*zwei [Männer], [im Jahre] 1994*
FM	Fremdsprachliches Material	*[Er hat das mit "]* *A big fish [" übersetzt]*
ITJ	Interjektion	*mhm, ach, tja*
KOUI	unterordnende Konjunktion mit "zu" und Infinitiv	*um [zu leben],* *anstatt [zu fragen]*
KOUS	unterordnende Konjunktion mit Satz	*weil, daß, damit,* *wenn, ob*
KON	nebenordnende Konjunktion	*und, oder, aber*
KOKOM	Vergleichspartikel, ohne Satz	*als, wie*
NN	normales Nomen	*Tisch, Herr, [das] Reisen*
NE	Eigennamen	*Hans, Hamburg, HSV*
PDS	substituierendes Demonstrativ–pronomen	*dieser, jener*
PDAT	attribuierendes Demonstrativ–pronomen	*jener [Mensch]*
PIS	substituierendes Indefinit–pronomen	*keiner, viele, man, niemand*
PIAT	attribuierendes Indefinit–pronomen ohne Determiner	*kein [Mensch],* *irgendein [Glas]*
PIDAT	attribuierendes Indefinit–pronomen mit Determiner	*[ein] wenig [Wasser],* *[die] beiden [Brüder]*
PPER	irreflexives Personalpronomen	*ich, er, ihm, mich, dir*
PPOSS	substituierendes Possessiv–pronomen	*meins, deiner*
PPOSAT	attribuierendes Possessivpronomen	*mein [Buch], deine [Mutter]*
	Relativpronomen	
PRELS	substituierend	*[der Hund,] der*
PRELAT	attribuierend	*[der Mann ,] dessen [Hund]*

POS =	Beschreibung	Beispiele
	Relativpronomen	
PRF	reflexives Personalpronomen	*sich, einander, dich, mir*
PWS	substituierendes Interrogativpronomen	*wer, was*
PWAT	attribuierendes Interrogativpronomen	*welche [Farbe], wessen [Hut]*
PWAV	adverbiales Interrogativ– oder Relativpronomen	*warum, wo, wann, worüber, wobei*
PAV	Pronominaladverb	*dafür, dabei, deswegen, trotzdem*
PTKZU	"zu" vor Infinitiv	*zu [gehen]*
PTKNEG	Negationspartikel	*nicht*
PTKVZ	abgetrennter Verbzusatz	*[er kommt] an, [er fährt] rad*
PTKANT	Antwortpartikel	*ja, nein, danke, bitte*
PTKA	Partikel bei Adjektiv oder Adverb	*am [schönsten], zu [schnell]*
TRUNC	Kompositions–Erstglied	*An– [und Abreise]*
VVFIN	finites Verb, voll	*[du] gehst, [wir] kommen [an]*
VVIMP	Imperativ, voll	*komm [!]*
VVINF	Infinitiv, voll	*gehen, ankommen*
VVIZU	Infinitiv mit "zu", voll	*anzukommen, loszulassen*
VVPP	Partizip Perfekt, voll	*gegangen, angekommen*
VAFIN	finites Verb, aux	*[du] bist, [wir] werden*
VAIMP	Imperativ, aux	*sei [ruhig !]*
VAINF	Infinitiv, aux	*werden, sein*
VAPP	Partizip Perfekt, aux	*gewesen*
VMFIN	finites Verb, modal	*dürfen*
VMINF	Infinitiv, modal	*wollen*
VMPP	Partizip Perfekt, modal	*[er hat] gekonnt*
XY	Nichtwort, Sonderzeichen enthaltend	*D2XW3*
$,	Komma	,
$.	Satzbeendende Interpunktion	. ? ! ; :
$(	sonstige Satzzeichen; satzintern	– [ ]()

Anlage 2: Fehlerquoten der verschiedenen Tags bei Morphy und TreeTagger

Morphy				TreeTagger			
	Fehler	**Tokens**	**Anteil Fehler**		**Fehler**	**Tokens**	**Anteil Fehler**
ADJEKTIVE				**ADJEKTIVE**			
ADJ	0	70	0,00%	ADJA	0	70	0,00%
ADJADV	22	71	30,99%	ADJD	4	71	5,63%
ADVERBIEN				**ADVERBIEN**			
ADV	66	528	12,50%	ADV	67	382	17,54%
		0		PTKNEG	13	46	28,26%
		0		PTKANT	21	90	23,33%
		0		PWAW	2	10	20,00%
PRO ADV	7	15	46,67%	PAV	2	15	13,33%
ARTIKEL				**ARTIKEL**			
ART DEF	2	227	0,88%	ART	11	303	3,63%
ART IND	10	76	13,16%			0	
INTERJEKTIONEN				**INTERJEKTIONEN**			
INJ	4	4	100,00%	ITJ	3	4	75,00%
KONJUNKTIONEN				**KONJUNKTIONEN**			
KON INF	2	2	100,00%	KOUI	2	2	100,00%
KON NEB	0	141	0,00%	KON	0	141	0,00%
KON UNT	8	64	12,50%	KOUS	7	64	10,94%
KON VGL	8	8	100,00%	KOKOM	1	8	12,50%
PRONOMEN				**PRONOMEN**			
PRO DEM ATT	2	51	3,92%	PDAT	5	51	9,80%
PRO DEM PRO	95	95	100,00%	PDS	42	95	44,21%
PRO IND ATT	6	22	27,27%	PIAT	1	22	4,55%
PRO IND PRO	21	47	44,68%	PIS	7	47	14,89%
PRO INR ATT	0	0		PWAT	0	0	
PRO INR PRO	5	5	100,00%	PWS	4	5	80,00%
PRO PER PRO	34	222	15,32%	PPER	35	222	15,77%
PRO POS ATT	16	21	76,19%	PPOSAT	0	21	0,00%
PRO POS PRO	0	0		PPOSS	0	0	
PRO REF PRO	5	30	16,67%	PRF	0	30	0,00%
PRO REL ATT	0	0		PRELAT	0	0	
PRO REL PRO	32	42	76,19%	PRELS	39	42	92,86%
PRÄPOSITIONEN				**PRÄPOSITIONEN**			
PRP	12	272	4,41%	APPR	9	244	3,69%
		0		APPRART	0	28	0,00%
EIGENNAMEN				**EIGENNAMEN**			
EIG	58	93	62,37%	NE	22	93	23,66%
SUBSTANTIVE				**SUBSTANTIVE**			
SUB	22	474	4,64%	NN	13	474	2,74%
VERBEN				**VERBEN**			
VER AUX	42	224	18,75%	VAFIN	50	224	22,32%
VER AUX INF	4	7	57,14%	VAINF	1	7	14,29%

VER AUX PA2	9	9	100,00%	VAPP	0	9	0,00%
VER MOD	4	39	10,26%	VMFIN	1	39	2,56%
VER MOD INF	1	1	100,00%	VMINF	0	1	0,00%
VER	52	146	35,62%	VVFIN	22	146	15,07%
VER IMP	1	2	50,00%	VVIMP	1	2	50,00%
VER INF	14	49	28,57%	VVINF	31	49	63,27%
VER EIZ	0	2	0,00%	VVIZU	0	2	0,00%
VER PA2	31	92	33,70%	VVPP	23	92	25,00%
ZUS	17	21	80,95%	PTKVZ	15	21	71,43%
SKZ	2	10	20,00%	PTKZU	1	10	10,00%
ZAL	20	36	55,56%	CARD	7	36	19,44%
INSGESAMT:	**634**	**3218**	**19,70%**	**INSGESAMT:**	**462**	**3218**	**14,36%**
KORREKTHEIT:			**80,30%**	**KORREKTHEIT:**			**85,64%**

Anlage 3: Liste sämtlicher Fehler (sortiert nach den korrigierten Wortklassen)

a) Morphy

Korr. POS	POS Morphy	Token
ADJ ADV	ADJ	deutschstämmig
ADJ ADV	ADJ	schuldhaft
ADJ ADV	ADV	ganz
ADJ ADV	ADV	offenbar
ADJ ADV	ADV	selbstverständlich
ADJ ADV	VER PA2	bekannt
ADJ ADV	VER PA2	bekannt
ADJ ADV	VER PA2	erbost
ADJ ADV	ZUS	gut
ADJ ADV	ZUS	klar
ADJ ADV	ZUS	klar
ADJ ADV	ZUS	klar
ADJ ADV	ZUS	klar
ADJ ADV	ZUS	klar
ADJ ADV	ZUS	klar
ADJ ADV	ZUS	klar
ADJ ADV	ZUS	klar
ADJ ADV	ZUS	klar
ADJ ADV	ZUS	klar
ADJ ADV	ZUS	schön
ADJ ADV	ZUS	wahr
ADJ ADV	ZUS	voll
ADV	ADJ	au
ADV	ADJ	lange
ADV	ADJ	net
ADV	ADJ	net
ADV	ADJ	nich
ADV	ADJ	nich
ADV	ADJ ADV	breit
ADV	ADJ ADV	breit
ADV	ADJ ADV	einfach
ADV	ADJ ADV	erfahrungsgemäß
ADV	ADJ ADV	früher
ADV	ADJ ADV	gut
ADV	ADJ ADV	natürlich
ADV	ADJ ADV	natürlich
ADV	ADJ ADV	natürlich
ADV	ADJ ADV	natürlich
ADV	ADJ ADV	natürlich
ADV	ADJ ADV	natürlich

ADV	ADJ ADV	net
ADV	ADJ ADV	net
ADV	ADJ ADV	rheinabwärts
ADV	ADJ ADV	sicher
ADV	ADJ ADV	sicher
ADV	ADJ ADV	ungefähr
ADV	ADJ ADV	weit
ADV	ADJ ADV	weit
ADV	ADJ ADV	weiter
ADV	ADJ ADV	weiter
ADV	ADJ ADV	weiter
ADV	ADJ ADV	weiter
ADV	ADJ ADV	weiter
ADV	ADJ ADV	verhältnismäßig
ADV	ADJ ADV	ziemlich
ADV	KON NEB	denn
ADV	KON NEB	denn
ADV	KON NEB	nur
ADV	KON NEB	nur
ADV	KON NEB	wie
ADV	PRP	a
ADV	PRP	a
ADV	PRP	plus
ADV	PRP	zu
ADV	VER AUX	grade
ADV	VER AUX	net
ADV	VER AUX INF	grade
ADV	VER AUX PA2	net
ADV	VER AUX PA2	net
ADV	VER AUX PA2	net
ADV	VER AUX PA2	nich
ADV	VER MOD	circa
ADV	VER MOD	net
ADV	VER MOD	rheinaufwärts
ADV	VER PA2	au
ADV	VER PA2	jawohl
ADV	VER PA2	jawohl
ADV	VER PA2	net
ADV	VER PA2	net
ADV	VER PA2	rheinaufwärts
ADV	VER PA2	soviel
ADV	VER PA2	vielmals
ADV	ZUS	halt
ADV	ZUS	halt
ADV	ZUS	halt

ADV	ZUS	herein
ADV	ZUS	hinauf
ADV	ZUS	hinauf
ADVPRO	ADJ ADV	drinne
ADVPRO	ADV	dran
ADVPRO	ADV	drauf
ADVPRO	ADV	drin
ADVPRO	ADV	drin
ADVPRO	ADV	drüben
ADVPRO	ZUS	darauf
ART DEF	PRO PER PRO	das
ART DEF	PRO PER PRO	das
ART IND	ADJ	n
ART IND	ADJ	n
ART IND	ADJ	ne
ART IND	ADJ	ne
ART IND	ADJ	ne
ART IND	ADJ ADV	n
ART IND	ADJ ADV	n
ART IND	ADJ ADV	n
ART IND	VER	eine
ART IND	VER AUX INF	e
EIG	ADJ ADV	Schwarz
EIG	ADJ ADV	Schwarz
EIG	ADJ ADV	Schwarz
EIG	ADJ ADV	Schwarz
EIG	ADJ ADV	Schwarz
EIG	ADJ ADV	Schwarz
EIG	ADJ ADV	Schwarz
EIG	ADJ ADV	Schwarz
EIG	ADJ ADV	Schwarz
EIG	SUB	Bingen
EIG	SUB	Bingen
EIG	SUB	Bingen
EIG	SUB	Bingen
EIG	SUB	Bingen
EIG	SUB	Bingen
EIG	SUB	Christianstraße
EIG	SUB	Flensburg
EIG	SUB	Flensburg
EIG	SUB	Flensburg
EIG	SUB	Goar
EIG	SUB	Goar
EIG	SUB	Goar

EIG	SUB	Goar
EIG	SUB	Goar
EIG	SUB	Goar
EIG	SUB	Goar
EIG	SUB	Goar
EIG	SUB	Goar
EIG	SUB	Hasenbach
EIG	SUB	Hasenbach
EIG	SUB	Hasenbach
EIG	SUB	Heidenheim
EIG	SUB	Julius
EIG	SUB	Kaub
EIG	SUB	Lorelei
EIG	SUB	Lorelei
EIG	SUB	Lorelei
EIG	SUB	Lorelei
EIG	SUB	Lorelei
EIG	SUB	Lorelei
EIG	SUB	Lorelei
EIG	SUB	Lorelei
EIG	SUB	Lorelei
EIG	SUB	Rheines
EIG	SUB	Schönburg
EIG	SUB	Wolf
EIG	VER	Sankt
EIG	VER	Sankt
EIG	VER	Sankt
EIG	VER	Sankt
EIG	VER	Sankt
EIG	VER	Sankt
EIG	VER	Sankt
EIG	VER	Sankt
EIG	VER	Sankt
EIG	VER	Sankt
EIG	VER	Sankt
EIG	ZUS	Schwarz
INJ	SUB	na
INJ	VER	bitte
INJ	VER AUX	bitteschön
INJ	VER PA2	gel
KON INF	ADV	um
KON INF	ADV	um
KON UNT	ADV	bevor
KON UNT	ADV	bevor
KON UNT	ADV	wie
KON UNT	ADV	wo

KON UNT	ADV	wo
KON UNT	KON NEB	wie
KON UNT	KON NEB	wie
KON UNT	PRP	als
KON VGL	ADV	wie
KON VGL	ADV	wie
KON VGL	ADV	wie
KON VGL	ADV	wie
KON VGL	ADV	wie
KON VGL	ADV	wie
KON VGL	KON NEB	als
KON VGL	PRP	als
PRO DEM ATT	ADJ	m
PRO DEM ATT	PRO IND PRO	jedem
PRO DEM PRO	ART DEF	das
PRO DEM PRO	ART DEF	das
PRO DEM PRO	ART DEF	das

PRO DEM PRO	ART DEF	das
PRO DEM PRO	ART DEF	das
PRO DEM PRO	ART DEF	das
PRO DEM PRO	ART DEF	das
PRO DEM PRO	ART DEF	das
PRO DEM PRO	ART DEF	das
PRO DEM PRO	ART DEF	das
PRO DEM PRO	ART DEF	das
PRO DEM PRO	ART DEF	das
PRO DEM PRO	ART DEF	das
PRO DEM PRO	ART DEF	das
PRO DEM PRO	ART DEF	das
PRO DEM PRO	ART DEF	das
PRO DEM PRO	ART DEF	das
PRO DEM PRO	ART DEF	das
PRO DEM PRO	ART DEF	das
PRO DEM PRO	ART DEF	das
PRO DEM PRO	ART DEF	dem
PRO DEM PRO	ART DEF	dem
PRO DEM PRO	ART DEF	dem
PRO DEM PRO	ART DEF	den
PRO DEM PRO	ART DEF	der
PRO DEM PRO	ART DEF	der
PRO DEM PRO	ART DEF	der
PRO DEM PRO	ART DEF	der
PRO DEM PRO	ART DEF	der
PRO DEM PRO	ART DEF	des
PRO DEM PRO	ART DEF	des
PRO DEM PRO	ART DEF	des
PRO DEM PRO	ART DEF	des
PRO DEM PRO	ART DEF	des
PRO DEM PRO	ART DEF	des
PRO DEM PRO	ART DEF	die
PRO DEM PRO	ART DEF	die
PRO DEM PRO	ART DEF	die
PRO DEM PRO	ART DEF	die
PRO DEM PRO	ART DEF	die
PRO DEM PRO	ART DEF	die
PRO DEM PRO	ART DEF	die
PRO DEM PRO	PRO DEM ATT	diese
PRO DEM PRO	PRO DEM ATT	diese
PRO DEM PRO	PRO DEM ATT	diese
PRO DEM PRO	PRO DEM ATT	diese
PRO DEM PRO	PRO DEM ATT	diesem
PRO DEM PRO	PRO DEM ATT	dieses

PRO DEM PRO	PRO PER PRO	das
PRO DEM PRO	PRO PER PRO	das
PRO DEM PRO	PRO PER PRO	das
PRO DEM PRO	PRO PER PRO	das
PRO DEM PRO	PRO PER PRO	das
PRO DEM PRO	PRO PER PRO	das
PRO DEM PRO	PRO PER PRO	das
PRO DEM PRO	PRO PER PRO	das
PRO DEM PRO	PRO PER PRO	das
PRO DEM PRO	PRO PER PRO	das
PRO DEM PRO	PRO PER PRO	das
PRO DEM PRO	PRO PER PRO	das
PRO DEM PRO	PRO PER PRO	das
PRO DEM PRO	PRO PER PRO	das
PRO DEM PRO	PRO PER PRO	das
PRO DEM PRO	PRO PER PRO	das
PRO DEM PRO	PRO PER PRO	das
PRO DEM PRO	PRO PER PRO	das
PRO DEM PRO	PRO PER PRO	die
PRO DEM PRO	PRO PER PRO	dies
PRO IND ATT	ADJ	all
PRO IND ATT	ADJ ADV	irgendwelche
PRO IND ATT	ADV	viel
PRO IND ATT	ADV	viel
PRO IND ATT	VER PA2	all
PRO IND ATT	VER PA2	irgendwelche
PRO IND PRO	ADJ	andere
PRO IND PRO	ADJ	anderes
PRO IND PRO	ADJ	andern
PRO IND PRO	ADV	etwas
PRO IND PRO	ADV	etwas
PRO IND PRO	ADV	etwas
PRO IND PRO	ADV	etwas
PRO IND PRO	ART IND	eine
PRO IND PRO	PRO IND ATT	nichts
PRO IND PRO	PRO IND ATT	nichts
PRO IND PRO	PRO IND ATT	nichts
PRO IND PRO	PRO IND ATT	nichts
PRO IND PRO	PRO IND ATT	nichts
PRO IND PRO	PRO REL PRO	was
PRO IND PRO	PRO REL PRO	was
PRO IND PRO	PRO REL PRO	was
PRO IND PRO	PRO REL PRO	was
PRO IND PRO	PRO REL PRO	was
PRO IND PRO	PRO REL PRO	was
PRO IND PRO	PRO REL PRO	was

PRO IND PRO	PRO REL PRO	was
PRO INR PRO	ADV	wie
PRO INR PRO	ADV	wie
PRO INR PRO	ADV	wie
PRO INR PRO	PRO REL PRO	was
PRO INR PRO	PRO REL PRO	was
PRO PER PRO	ADJ	i
PRO PER PRO	ADJ	s
PRO PER PRO	ADJ	s
PRO PER PRO	ADJ	s
PRO PER PRO	ADJ	s
PRO PER PRO	ADJ	s
PRO PER PRO	ADJ ADV	s
PRO PER PRO	ADJ ADV	s
PRO PER PRO	ADJ ADV	se
PRO PER PRO	ADJ ADV	se
PRO PER PRO	ADJ ADV	se
PRO PER PRO	PRO REF PRO	mich
PRO PER PRO	PRO REF PRO	mir
PRO PER PRO	VER AUX	s
PRO PER PRO	VER AUX	s
PRO PER PRO	VER AUX	s
PRO PER PRO	VER AUX	s
PRO PER PRO	VER AUX	s
PRO PER PRO	VER AUX	s
PRO PER PRO	VER AUX	se
PRO PER PRO	VER AUX INF	s
PRO PER PRO	VER AUX PA2	s
PRO PER PRO	VER AUX PA2	s
PRO PER PRO	VER AUX PA2	s
PRO PER PRO	VER INF	s
PRO PER PRO	VER PA2	i
PRO PER PRO	VER PA2	i
PRO PER PRO	VER PA2	s
PRO PER PRO	VER PA2	s
PRO PER PRO	VER PA2	s
PRO PER PRO	VER PA2	s
PRO PER PRO	VER PA2	s
PRO PER PRO	VER PA2	s
PRO PER PRO	VER PA2	se
PRO POS ATT	PRO PER PRO	ihr
PRO POS ATT	PRO PER PRO	ihr
PRO POS ATT	PRO PER PRO	ihr
PRO POS ATT	PRO PER PRO	ihr
PRO POS ATT	PRO PER PRO	ihr

PRO POS ATT	PRO POS PRO	ihre
PRO POS ATT	PRO POS PRO	ihrem
PRO POS ATT	PRO POS PRO	ihrem
PRO POS ATT	PRO POS PRO	ihren
PRO POS ATT	PRO POS PRO	ihren
PRO POS ATT	PRO POS PRO	seinem
PRO POS ATT	PRO POS PRO	unseren
PRO POS ATT	PRO REF PRO	meiner
PRO POS ATT	PRO REF PRO	seiner
PRO POS ATT	PRO REF PRO	unser
PRO POS ATT	PRO REF PRO	unser
PRO REF PRO	PRO PER PRO	mich
PRO REF PRO	PRO PER PRO	mich
PRO REF PRO	PRO PER PRO	mich
PRO REF PRO	PRO PER PRO	mich
PRO REF PRO	PRO PER PRO	uns
PRO REL PRO	ART DEF	das
PRO REL PRO	ART DEF	der
PRO REL PRO	ART DEF	der
PRO REL PRO	ART DEF	der
PRO REL PRO	ART DEF	der
PRO REL PRO	ART DEF	der
PRO REL PRO	ART DEF	der
PRO REL PRO	ART DEF	der
PRO REL PRO	ART DEF	der
PRO REL PRO	ART DEF	die
PRO REL PRO	ART DEF	die
PRO REL PRO	ART DEF	die
PRO REL PRO	ART DEF	die
PRO REL PRO	ART DEF	die
PRO REL PRO	ART DEF	die
PRO REL PRO	ART DEF	die
PRO REL PRO	ART DEF	die
PRO REL PRO	ART DEF	die
PRO REL PRO	ART DEF	die
PRO REL PRO	ART DEF	die
PRO REL PRO	ART DEF	die
PRO REL PRO	ART DEF	die
PRO REL PRO	ART DEF	die
PRO REL PRO	ART DEF	die
PRO REL PRO	ART DEF	die
PRO REL PRO	ART DEF	die
PRO REL PRO	ART DEF	die
PRO REL PRO	ART DEF	die
PRO REL PRO	ART DEF	die

PRO REL PRO	ART DEF	die
PRO REL PRO	ART DEF	die
PRO REL PRO	PRO PER PRO	das
PRO REL PRO	PRO PER PRO	die
PRO REL PRO	PRO PER PRO	die
PRP	ADV	bis
PRP	ADV	bis
PRP	ADV	bis
PRP	ADV	bis
PRP	ADV	bis
PRP	ADV	unter
PRP	ADV	vor
PRP	ADV	zu
PRP	ADV PRO	z
PRP	VER PA2	vorm
PRP	ZUS	aus
PRP	ZUS	durch
SKZ	ADV	zu
SKZ	PRP	zu
SUB	ADJ	Folgendes
SUB	ADJ	ganzen
SUB	ADJ	ganzen
SUB	ADJ	großen
SUB	ADJ	großen
SUB	ADJ ADV	Leichter
SUB	ADJ ADV	Personal
SUB	ADJ ADV	Personal
SUB	ADJ ADV	xbeliebige
SUB	EIG	Felsenriffe
SUB	EIG	Mädchenlehen
SUB	EIG	Mädchenlehen
SUB	EIG	Mär
SUB	EIG	NN
SUB	VER	bitte
SUB	VER	Lotse
SUB	VER	Lotse
SUB	VER	Lotse
SUB	VER	Lotse
SUB	VER	Lotse
SUB	VER	Lotse
SUB	VER IMP	Stocher
VER	ADJ	besagt
VER	ADJ	hinunterfahren
VER	ADJ ADV	bedeutet
VER	ADJ ADV	weiß

VER	ADJ ADV	weiß
VER	ADJ ADV	weiß
VER	ADJ ADV	weiß
VER	PRO POS ATT	mein
VER	PRO POS ATT	mein
VER	PRO POS ATT	mein
VER	PRO POS ATT	mein
VER	PRO POS ATT	mein
VER	PRO POS ATT	mein
VER	PRO POS ATT	mein
VER	PRO POS ATT	mein
VER	PRO POS ATT	mein
VER	VER AUX	kullern
VER	VER AUX	möcht
VER	VER AUX INF	hineinsteigen
VER	VER IMP	gelt
VER	VER IMP	glaub
VER	VER IMP	mach
VER	VER IMP	mach
VER	VER INF	durchkommen
VER	VER INF	erklären
VER	VER INF	gehen
VER	VER INF	halten
VER	VER INF	importieren
VER	VER INF	kommen
VER	VER INF	kommen
VER	VER INF	kommen
VER	VER INF	kommen
VER	VER INF	machen
VER	VER INF	machen
VER	VER INF	machen
VER	VER INF	sagen
VER	VER INF	sitzen
VER	VER INF	wissen
VER	VER PA2	bedeutet
VER	VER PA2	befördert
VER	VER PA2	befördert
VER	VER PA2	bekommen
VER	VER PA2	geb
VER	VER PA2	gefällt
VER	VER PA2	gehört
VER	VER PA2	intoniert
VER	VER PA2	nehm
VER	VER PA2	verkörpert
VER	VER PA2	verlangt

VER	VER PA2	überlegt
VER	VER PA2	überlegt
VER	VER PA2	überlegt
VER AUX	ADJ	is
VER AUX	ADJ	is
VER AUX	ADJ	is
VER AUX	ADJ	is
VER AUX	ADJ	is
VER AUX	ADJ	is
VER AUX	ADJ	is
VER AUX	ADJ	isch
VER AUX	ADJ	isch
VER AUX	ADJ	isch
VER AUX	ADJ	isch
VER AUX	ADJ	isch
VER AUX	ADJ	isch
VER AUX	ADJ	isch
VER AUX	ADJ ADV	is
VER AUX	ADJ ADV	is
VER AUX	ADJ ADV	is
VER AUX	ADJ ADV	is
VER AUX	ADJ ADV	is
VER AUX	ADJ ADV	isch
VER AUX	VER AUX INF	haben
VER AUX	VER AUX PA2	is
VER AUX	VER AUX PA2	is
VER AUX	VER INF	is
VER AUX	VER INF	is
VER AUX	VER INF	isch
VER AUX	VER INF	isch
VER AUX	VER PA2	hab
VER AUX	VER PA2	hab
VER AUX	VER PA2	hab
VER AUX	VER PA2	hab
VER AUX	VER PA2	is
VER AUX	VER PA2	is
VER AUX	VER PA2	is
VER AUX	VER PA2	is
VER AUX	VER PA2	is
VER AUX	VER PA2	is
VER AUX	VER PA2	is
VER AUX	VER PA2	isch
VER AUX	VER PA2	isch
VER AUX	VER PA2	ischt

VER AUX INF	VER AUX	haben
VER AUX INF	VER AUX	haben
VER AUX INF	VER INF	haben
VER AUX INF	VER INF	sein
VER AUX PA2	VER PA2	gehabt
VER AUX PA2	VER PA2	gewesen
VER AUX PA2	VER PA2	gewesen
VER AUX PA2	VER PA2	gewesen
VER AUX PA2	VER PA2	gewesen
VER AUX PA2	VER PA2	gewesen
VER AUX PA2	VER PA2	geworden
VER AUX PA2	VER PA2	worden
VER AUX PA2	VER PA2	worden
VER IMP	VER INF	entschuldigen
VER INF	VER	abgehen
VER INF	VER	bedeuten
VER INF	VER	eröffnen
VER INF	VER	fragen
VER INF	VER	sagen
VER INF	VER	sagen
VER INF	VER	sagen
VER INF	VER	sagen
VER INF	VER	sagen
VER INF	VER	sagen
VER INF	VER	verübeln
VER INF	VER	wissen
VER INF	VER	äußern
VER INF	VER PA2	hintendranhängen
VER MOD	VER MOD INF	können
VER MOD	VER MOD INF	können
VER MOD	VER MOD INF	müssen
VER MOD	VER MOD INF	wollen
VER MOD INF	VER MOD	müssen
VER PA2	ADJ	aufgewickelt
VER PA2	ADJ ADV	angezeigt
VER PA2	ADJ ADV	ausgestiegen
VER PA2	ADJ ADV	beachtet
VER PA2	ADJ ADV	befahren
VER PA2	ADJ ADV	bekommen
VER PA2	ADJ ADV	eingefahren
VER PA2	ADJ ADV	ereignet
VER PA2	ADJ ADV	erteilt
VER PA2	ADJ ADV	erzählt
VER PA2	ADJ ADV	festgemacht
VER PA2	ADJ ADV	gedacht

VER PA2	ADJ ADV	gefahren
VER PA2	ADJ ADV	gefährdet
VER PA2	ADJ ADV	gegeben
VER PA2	ADJ ADV	gemacht
VER PA2	ADJ ADV	gemerkt
VER PA2	ADJ ADV	gesessen
VER PA2	ADJ ADV	hingelegt
VER PA2	ADJ ADV	kredenzt
VER PA2	ADJ ADV	passiert
VER PA2	ADJ ADV	unterschrieben
VER PA2	ADJ ADV	verletzt
VER PA2	ADJ ADV	verletzt
VER PA2	ADJ ADV	verliebt
VER PA2	ADJ ADV	zugenommen
VER PA2	VER	kommen
VER PA2	VER AUX PA2	dabeigehabt
VER PA2	VER AUX PA2	eingetragen
VER PA2	VER INF	bekommen
VER PA2	VER MOD	raufgezogen
ZAL	ADJ	achtundzwanzigsten
ZAL	ADJ	achtzehn
ZAL	ADJ	achtzehn
ZAL	ADJ	fuffzig
ZAL	ADJ	fünfzehnUhrfünfzig
ZAL	ADJ	fünfzig
ZAL	ADJ	fünfzig
ZAL	ADJ	neunzehnhundertelfzwölf
ZAL	ADJ	sechstausend
ZAL	ADJ	sechstausend
ZAL	ADJ	zweihundert
ZAL	ADJ	zweihundertfünfzig
ZAL	ADJ ADV	neunzehnhundertfünfundsechzig
ZAL	VER AUX	neunzehnhundertzehn
ZAL	VER MOD	dreihundertfünfundsechzig
ZAL	VER MOD	vierundzwanzig
ZAL	ADJ	achtzig
ZAL	ART IND	ein
ZAL	VER INF	sieben
ZAL	VER PA2	siebenundachtzig
ZUS	ADV	ab
ZUS	ADV	über
ZUS	ADV	über
ZUS	ADV	zurück
ZUS	ADV PRO	dabei
ZUS	ADV PRO	hinzu

ZUS	PRP	an
ZUS	PRP	auf
ZUS	PRP	auf
ZUS	PRP	aus
ZUS	PRP	vor
ZUS	ADV	zu
ZUS	ADV PRO	dazu
ZUS	ADV PRO	hinzu
ZUS	ART IND	ein
ZUS	ART IND	ein
ZUS	PRP	auf

b) TreeTagger

Korr POS	**POS TreeTagger**	**Token**
ADJD	ADJA	voller
ADJD	ADV	weniger
ADJD	VVFIN	bewegt
ADJD	VVPP	bekannt
ADV	ADJA	net
ADV	ADJD	annähernd
ADV	ADJD	bloß
ADV	ADJD	bloß
ADV	ADJD	bloß
ADV	ADJD	bloß
ADV	ADJD	breit
ADV	ADJD	breit
ADV	ADJD	einfach
ADV	ADJD	erfahrungsgemäß
ADV	ADJD	früher
ADV	ADJD	früher
ADV	ADJD	früher
ADV	ADJD	gar
ADV	ADJD	gar
ADV	ADJD	gleich
ADV	ADJD	gut
ADV	ADJD	natürlich
ADV	ADJD	natürlich
ADV	ADJD	natürlich
ADV	ADJD	natürlich
ADV	ADJD	natürlich
ADV	ADJD	natürlich
ADV	ADJD	net
ADV	ADJD	nämlich
ADV	ADJD	nämlich
ADV	ADJD	nämlich
ADV	ADJD	sicher
ADV	ADJD	sicher
ADV	ADJD	später
ADV	ADJD	später
ADV	ADJD	ungefähr

ADV	ADJD	wahrscheinlich
ADV	ADJD	wahrscheinlich
ADV	ADJD	wahrscheinlich
ADV	ADJD	weit
ADV	ADJD	weit
ADV	ADJD	weiter
ADV	ADJD	weiter
ADV	ADJD	weiter
ADV	ADJD	weiter
ADV	ADJD	weiter
ADV	ADJD	verhältnismäßig
ADV	ADJD	vielleicht
ADV	ADJD	vielleicht
ADV	ADJD	ziemlich
ADV	ADJD	ziemlich
ADV	APPR	zu
ADV	APPR	zu
ADV	APPR	zu
ADV	APPR	zu
ADV	FM	a
ADV	FM	a
ADV	ITJ	au
ADV	ITJ	au
ADV	KOUS	da
ADV	NN	grade
ADV	NN	grade
ADV	PAV	trotzdem
ADV	PIAT	viel
ADV	PIAT	viel
ADV	PIS	soviel
ADV	PTKA	zu
ADV	PTKANT	nein
ADV	PTKVZ	hinauf
ADV	PTKVZ	nebeneinander
ADV	PTKVZ	zurück
APPR	APPO	gegenüber
APPR	KON	bis
APPR	KON	bis
APPR	KON	bis
APPR	NN	s
APPR	NN	z
APPR	PTKA	zu
APPR	PTKVZ	an
APPR	PTKVZ	über
ART	ADJA	einen
ART	NN	e
ART	NN	n
ART	NN	n
ART	NN	n
ART	NN	n
ART	NN	n
ART	PDS	das

ART	PDS	das
ART	PIS	einer
ART	PIS	eines
CARD	ADJA	achtundzwanzigsten
CARD	ADJA	neunzehnhundertelfzwölf
CARD	NN	dreihundertfünfundsechzig
CARD	NN	fuffzig
CARD	NN	fünfzehnUhrfünfzig
CARD	NN	neunzehnhundertfünfundsechzig
CARD	NN	neunzehnhundertzehn
ITJ	ADJD	bitteschön
ITJ	NN	gel
ITJ	VVFIN	bitte
KOKOM	KOUS	wie
KOUI	APPR	um
KOUI	APPR	um
KOUS	ADV	ob
KOUS	ADV	soweit
KOUS	KOKOM	wie
KOUS	KOKOM	wie
KOUS	KON	wie
KOUS	PWAV	wo
KOUS	PWAV	wo
NE	NN	Bingen
NE	NN	Bingen
NE	NN	Bingen
NE	NN	Bingen
NE	NN	Bingen
NE	NN	Bingen
NE	NN	Bingen
NE	NN	Christianstraße
NE	NN	Hasenbach
NE	NN	Hasenbach
NE	NN	Hasenbach
NE	NN	Kaub
NE	NN	Lorelei
NE	NN	Lorelei
NE	NN	Lorelei
NE	NN	Lorelei
NE	NN	Lorelei
NE	NN	Lorelei
NE	NN	Lorelei
NE	NN	Lorelei
NE	NN	Lorelei
NE	NN	Schönburg
NN	ADJA	ganzen
NN	ADJA	ganzen
NN	ADJA	großen
NN	ADJA	großen
NN	NE	Fischer
NN	NE	Fischer
NN	NE	LKW

NN	NE	LKW
NN	PTKANT	bitte
NN	VVFIN	Legitimieren
NN	VVFIN	NN
NN	VVIMP	Mär
NN	VVIMP	Mär
PAV	ADV	drüben
PAV	NN	drinne
PDAT	NN	m
PDAT	PDS	diese
PDAT	PDS	diese
PDAT	PDS	diese
PDAT	PIDAT	jedem
PDS	ART	das
PDS	ART	das
PDS	ART	das
PDS	ART	das
PDS	ART	das
PDS	ART	das
PDS	ART	das
PDS	ART	das
PDS	ART	das
PDS	ART	das
PDS	ART	das
PDS	ART	das
PDS	ART	das
PDS	ART	das
PDS	ART	das
PDS	ART	das
PDS	ART	das
PDS	ART	das
PDS	ART	das
PDS	ART	das
PDS	ART	das
PDS	ART	dem
PDS	ART	dem
PDS	ART	den
PDS	ART	der
PDS	ART	der
PDS	ART	der
PDS	ART	der
PDS	ART	des
PDS	ART	des
PDS	ART	des
PDS	ART	des
PDS	ART	des
PDS	ART	des
PDS	ART	die
PDS	ART	die
PDS	ART	die
PDS	ART	die
PDS	PDAT	diese
PDS	PDAT	diese
PDS	PDAT	diesem

PDS	PDAT	dieses
PDS	PRELS	die
PDS	PRELS	die
PIAT	PIDAT	jeder
PIS	ADJA	andern
PIS	ADV	etwas
PIS	ADV	etwas
PIS	ADV	etwas
PIS	ADV	etwas
PIS	PRELS	was
PIS	PWS	was
PPER	ADJA	s
PPER	ADJD	se
PPER	ADJD	se
PPER	APPR	in
PPER	FM	i
PPER	FM	i
PPER	NN	i
PPER	NN	s
PPER	NN	s
PPER	NN	s
PPER	NN	s
PPER	NN	s
PPER	NN	s
PPER	NN	s
PPER	NN	s
PPER	NN	s
PPER	NN	s
PPER	NN	s
PPER	NN	s
PPER	NN	s
PPER	NN	s
PPER	NN	s
PPER	NN	s
PPER	NN	s
PPER	NN	s
PPER	NN	s
PPER	NN	s
PPER	NN	s
PPER	NN	s
PPER	PRF	mich
PPER	PRF	mir
PPER	PRF	mir
PPER	VVFIN	se
PPER	VVFIN	se
PPER	VVFIN	se
PRELS	ART	das
PRELS	ART	der
PRELS	ART	der
PRELS	ART	der
PRELS	ART	der
PRELS	ART	der

PRELS	ART	der
PRELS	ART	der
PRELS	ART	der
PRELS	ART	die
PRELS	ART	die
PRELS	ART	die
PRELS	ART	die
PRELS	ART	die
PRELS	ART	die
PRELS	ART	die
PRELS	ART	die
PRELS	ART	die
PRELS	ART	die
PRELS	ART	die
PRELS	ART	die
PRELS	ART	die
PRELS	ART	die
PRELS	ART	die
PRELS	ART	die
PRELS	ART	die
PRELS	ART	die
PRELS	ART	die
PRELS	ART	die
PRELS	ART	die
PRELS	ART	die
PRELS	ART	die
PRELS	ART	die
PRELS	PDS	das
PRELS	PIS	was
PRELS	PIS	was
PRELS	PWS	was
PRELS	PWS	was
PRELS	PWS	was
PTKANT	ADV	ja
PTKANT	ADV	ja
PTKANT	ADV	ja
PTKANT	ADV	ja
PTKANT	ADV	ja
PTKANT	ADV	ja
PTKANT	ADV	ja
PTKANT	ADV	ja
PTKANT	ADV	ja
PTKANT	ADV	ja
PTKANT	ADV	ja
PTKANT	ADV	ja
PTKANT	ADV	ja
PTKANT	ADV	ja
PTKANT	ADV	ja
PTKANT	ADV	ja
PTKANT	ADV	ja
PTKANT	ADV	ja
PTKANT	ADV	ja

PTKANT	ADV	ja
PTKANT	ADV	ja
PTKNEG	ADJA	net
PTKNEG	ADJA	net
PTKNEG	ADJA	nich
PTKNEG	ADJA	nich
PTKNEG	ADJA	nich
PTKNEG	ADJD	net
PTKNEG	PIAT	keine
PTKNEG	VVFIN	net
PTKNEG	VVFIN	net
PTKNEG	VVFIN	net
PTKNEG	VVFIN	net
PTKNEG	VVFIN	net
PTKNEG	VVFIN	net
PTKVZ	ADV	hinauf
PTKVZ	ADV	hinzu
PTKVZ	ADV	runter
PTKVZ	APPR	ab
PTKVZ	APPR	an
PTKVZ	APPR	auf
PTKVZ	APPR	auf
PTKVZ	APPR	vor
PTKVZ	APPR	über
PTKVZ	APPR	über
PTKVZ	ART	ein
PTKVZ	ART	ein
PTKVZ	PAV	dabei
PTKVZ	PAV	dazu
PTKVZ	PTKA	zu
PTKZU	PTKA	zu
PWAV	KOKOM	wie
PWAV	KOUS	wie
PWS	KOKOM	wie
PWS	KOKOM	wie
PWS	KOKOM	wie
PWS	PRELS	was
VAFIN	ADJA	isch
VAFIN	ADJA	isch
VAFIN	ADJA	isch
VAFIN	ADJA	isch
VAFIN	ADJA	isch
VAFIN	ADJA	isch
VAFIN	ADJA	isch
VAFIN	ADJA	isch
VAFIN	ADJA	isch
VAFIN	ADJA	isch
VAFIN	ADJA	isch
VAFIN	ADJA	ischt
VAFIN	NN	is
VAFIN	NN	is
VAFIN	NN	is

VAFIN	NN	is
VAFIN	NN	is
VAFIN	NN	is
VAFIN	NN	is
VAFIN	NN	is
VAFIN	NN	is
VAFIN	NN	is
VAFIN	NN	is
VAFIN	NN	is
VAFIN	NN	is
VAFIN	NN	is
VAFIN	NN	is
VAFIN	NN	is
VAFIN	NN	is
VAFIN	NN	is
VAFIN	NN	is
VAFIN	NN	is
VAFIN	NN	is
VAFIN	NN	is
VAFIN	NN	is
VAFIN	NN	is
VAFIN	NN	is
VAFIN	NN	is
VAFIN	NN	is
VAFIN	NN	is
VAFIN	VAIMP	hab
VAFIN	VAIMP	hab
VAFIN	VAIMP	hab
VAFIN	VAIMP	hab
VAFIN	VAIMP	hab
VAFIN	VAINF	haben
VAFIN	VAINF	haben
VAFIN	VAINF	haben
VAFIN	VVFIN	isch
VAFIN	VVFIN	isch
VAINF	VAFIN	haben
VMFIN	VMINF	können
VVFIN	ADJA	gelt
VVFIN	ADJD	nehm
VVFIN	ADJD	überlegt
VVFIN	PPOSAT	mein
VVFIN	PPOSAT	mein
VVFIN	PPOSAT	mein
VVFIN	PPOSAT	mein
VVFIN	PPOSAT	mein
VVFIN	PPOSAT	mein
VVFIN	PPOSAT	mein
VVFIN	PPOSAT	mein
VVFIN	PPOSAT	mein
VVFIN	PPOSAT	meine
VVFIN	PTKANT	danke
VVFIN	VVIMP	glaub

VVFIN	VVIMP	mach
VVFIN	VVIMP	mach
VVFIN	VVIMP	schiebt
VVFIN	VVINF	machen
VVFIN	VVINF	sagen
VVFIN	VVPP	befördert
VVFIN	VVPP	verlangt
VVIMP	VVFIN	entschuldigen
VVINF	VVFIN	abgehen
VVINF	VVFIN	aussagen
VVINF	VVFIN	bedeuten
VVINF	VVFIN	bringen
VVINF	VVFIN	durchlesen
VVINF	VVFIN	durchstehen
VVINF	VVFIN	einlegen
VVINF	VVFIN	erinnern
VVINF	VVFIN	eröffnen
VVINF	VVFIN	fragen
VVINF	VVFIN	hintendranhängen
VVINF	VVFIN	hinweisen
VVINF	VVFIN	kommen
VVINF	VVFIN	kriegen
VVINF	VVFIN	sagen
VVINF	VVFIN	sagen
VVINF	VVFIN	sagen
VVINF	VVFIN	sagen
VVINF	VVFIN	sagen
VVINF	VVFIN	sagen
VVINF	VVFIN	sagen
VVINF	VVFIN	sitzen
VVINF	VVFIN	stehen
VVINF	VVFIN	unterstellen
VVINF	VVFIN	verübeln
VVINF	VVFIN	wissen
VVINF	VVFIN	vorhalten
VVINF	VVFIN	äußern
VVINF	VVFIN	äußern
VVINF	VVFIN	äußern
VVINF	VVFIN	äußern
VVPP	ADJA	aufgewickelt
VVPP	ADJA	raufgezogen
VVPP	ADJD	eingefahren
VVPP	ADJD	eingefahren
VVPP	ADJD	verliebt
VVPP	NN	Eingetragen
VVPP	PTKVZ	gefangen
VVPP	VVFIN	beansprucht
VVPP	VVFIN	bekommen
VVPP	VVFIN	bekommen
VVPP	VVFIN	bewegt
VVPP	VVFIN	entstanden
VVPP	VVFIN	erklärt

VVPP	VVFIN	erzählt
VVPP	VVFIN	fortbewegt
VVPP	VVFIN	gehört
VVPP	VVFIN	kommen
VVPP	VVFIN	verdoppelt
VVPP	VVFIN	verlebt
VVPP	VVFIN	verletzt
VVPP	VVFIN	versehen
VVPP	VVFIN	versteigert
VVPP	VVFIN	verurteilt