Braucht Nominalphrasenerkennung linguistisches Wissen?

Sandra Kübler

SFB 441 "Linguistische Datenstrukturen", Eberhard-Karls-Universität Tübingen,
Nauklerstr. 35, 72074 Tübingen
kuebler@sfs.nphil.uni-tuebingen.de

Zusammenfassung Maschinelles Lernen wird häufig zur effizienten Annotation großer Datenmengen eingesetzt. Die Forschung zu maschinellen Lernverfahren bschränkt sich i.a. darauf, unterschiedliche Lernverfahren zu vergleichen oder die optimale Größe der Trainingsdaten zu bestimmen. Bisher wurde jedoch nicht untersucht, inwieweit sich linguistisches Wissen bei der Aufgabendefinition positiv auswirken kann. Dies soll hier anhand des Lernens von Base-Nominalphrasen mit drei unterschiedlichen Definitionen untersucht werden. Die Definitionen unterschieden sich im Grad der linguistisch motivierten Erweiterungen, die zu einer eher praktisch motivierten ersten Definition hinzukamen. Die Untersuchungen ergaben, daß sich die Anzahl der falsch klassifizierten Wörter um ein Drittel reduzieren läßt.

1 Einleitung

Maschinelle Lernverfahren werden häufig als Mittel angesehen, wenn sich die Aufgabe stellt, große Ressourcen auf dem Gebiet der syntaktischen Annotation mit vertretbarem Aufwand zu erstellen. In den letzten Jahren haben sich auf dem Gebiet des Maschinellen Lernens Verfahren zur Klassifikation wie z.B. Entscheidungsbäume, Hidden-Markov-Modelle oder ähnlichkeitsbasiertes Lernen durchgesetzt; Implementierungen dieser Verfahren sind frei erhältlich.

Ein Vorteil von Klassifikationsverfahren wird oft darin gesehen, daß für die Aufbereitung der Korpora für die Lernaufgabe kein linguistisches Wissen benötigt wird, daß also linguistisches Wissen nur bei der Annotation des Trainingsmaterials nicht jedoch bei der Anwendung von Lernverfahren eingesetzt werden muß. Die Forschung im Bereich der maschinellen Lernverfahren beschränkt sich i.a. auf den Vergleich verschiedener Lernverfahren oder auf Untersuchungen, welche Größe bei Trainingskorpora benötigt wird bzw. welche Merkmale den Lernerfolg für eine bestimmte Aufgabe und ein bestimmtes Lernverfahren maximieren. Es gibt jedoch keine Literatur darüber, wie sich die Definition einer Lernaufgabe auf das Ergebnis auswirkt.

In diesem Beitrag soll gezeigt werden, daß für die Definition der Lernaufgabe linguistisches Wissen von Vorteil sein kann. Die hier gewählte Lernaufgabe besteht in der Erkennung von Nominalphrasen (NPs), eine Aufgabe, die auch schon zum Vergleich verschiedener Lernverfahren verwendet wurde (vgl. [8]).

Als Lernverfahren wurde das "Memory-Based Learning" in der Implementierung von Daelemans et al. [1] verwendet, als Trainings- und Testdaten die deutsche Verbmobil-Baumbank [7].

2 Das Lernen von Base-Nominalphrasen

2.1 Die Lernaufgabe

Die Aufgabe besteht darin, Base-Nominalphrasen zu erkennen und als solche zu kennzeichnen. Unter Base-Nominalphrasen versteht man nicht-rekursive Nominalphrasen, d.h. Base-NPs enden nach dem Head, Postmodifikationen werden als nicht zur Base-NP gehörend betrachtet.

Bsp: *dann bräuchte [NP ich] noch [NP ein paar Tips] für [NP das Abendprogramm]; [NP ich] würde [NP Ihnen] [NP die Auswahl] [NP des Hotels] überlassen.*

Die meisten linguistischen Theorien des Deutschen, dagegen, würden die komplette Sequenz *ein paar Tips für das Abendprogramm* inklusive der postmodifizierenden Präpositionalphrase als Nominalphrase bezeichnen. Ebenso die Sequenz aus Nominalphrase und Genitiv *die Auswahl des Hotels.*

Es ist offensichtlich, daß diese Definition der Base-NP gewissen Beschränkungen unterliegt, wobei die Nichtrekursivität dieser Strukturen gerade im Deutschen zu Problemen führt. Dies zeigt sich an NPs mit prämodifizierenden Partizipialkonstruktionen wie *das billigste der drei zur Debatte stehenden Hotels.* Laut Definition wird diese komplexe NP folgendermaßen in Base-NPs gegliedert: *[NP das billigste] der drei zur [NP Debatte] stehenden Hotels.* Die NP *der drei zur Debatte stehenden Hotels* ist in ihrer Gesamtheit keine Base-NP, da sie eine weitere NP enthält. Aufgrund der eingebetteten Struktur, kann jedoch auch keine Base-NP *der drei Hotels* gebildet werden, da dies den linearen Ablauf der Struktur stören würde und daher nicht dargestellt werden kann.

Die Base-Nominalphrasen-Erkennung wird häufig als repräsentative Aufgabe zum Vergleich von Lernverfahren verwendet. In diesen Fällen werden meist die Daten von Ramshaw und Marcus [4] benutzt (siehe [8]). Ramshaw und Marcus verwenden eine Heuristik, die aus der Penn Treebank [3] nur solche NPs extrahiert, die keine weitere NP enthalten. Sie merken an, "This heuristic fails in some cases", machen jedoch keine Angaben, wie häufig diese Fälle sind. Anhand der deutschen Verbmobil-Baumbank soll untersucht werden, ob sich durch linguistisches Wissen in Form einer linguistisch motivierten Definition von Base-NPs Klassifikationsfehler vermeiden lassen.

Als Darstellungsformat für die syntaktische Information wurde das IOB-Format gewählt, das jedem Wort eines der drei Tags {I,O,B} zuweist. I kennzeichnet Wörter innerhalb einer NP, O Wörter, die nicht zu einer NP gehören, und B das erste Wort einer NP, die direkt an eine vorausgehende NP anschließt. Bsp: *kann/O ich/I den/B Termin/I bestätigen/O.* Da die beiden NPs *ich* und *den Termin* direkt aneinander anschließen, muß der Beginn der zweiten NP durch ein besonderes Tag, B, markiert werden. Dieses Format wurde bei einem Vergleich von mehreren Repräsentationsformen [9] als das erfolgreichste ermittelt.

2.2 Die deutsche Verbmobil-Baumbank

Die deutsche Verbmobil-Baumbank [7] enthält ca. 38 000 syntaktisch annotierte Äußerungen aus den Domänen Terminabsprachen, Reiseplanung und Hotelreservierung. Es handelt sich dabei um Transliterationen gesprochener Dialoge.

Da für die Definition der Base-NP vor allem die Annotationsstrategie bezüglich der Anbindung von prämodifizierenden und postmodifizierenden Elementen in der NP von Bedeutung ist, sollen diese Phänomene hier kurz erklärt werden. Das Annotationsschema sieht eine flache Anbindung für pränominale Modifikationen vor (siehe Abb. 1), sie werden direkt an das Head-Nomen angebunden. Postnominale Modifikationen werden i.d.R. hoch angebunden (siehe Abb. 2), d.h. das Nomen wird mit den pränominalen Modifikatoren zu einer Nominalphrase projeziert, dann wird die postnominale Phrase auf der nächsten Ebene angebunden. Eine Ausnahme zu dieser Annotationsstrategie bilden Appositionen, die flach in der NP angebunden werden (siehe Abb. 3), da ihr Skopus nicht ambig ist.

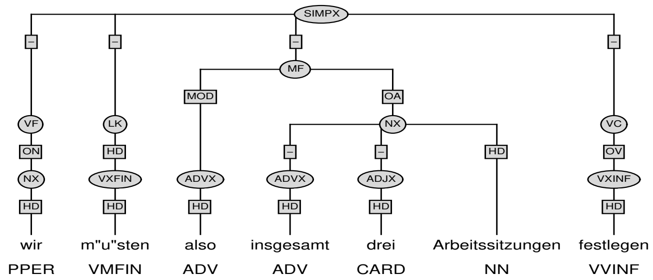

Abbildung 1. Eine prämodifizierte NP aus der deutschen Verbmobil-Baumbank

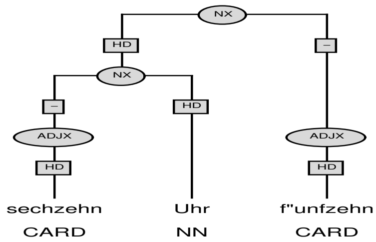

Abbildung 2. Eine postmodifizierte NP aus der deutschen Verbmobil-Baumbank

Verwendet man die Definition der Base-NPs von Ramshaw und Marcus, wird deutlich, daß sowohl bei Postmodifikationen als auch bei Appositionen zwar die

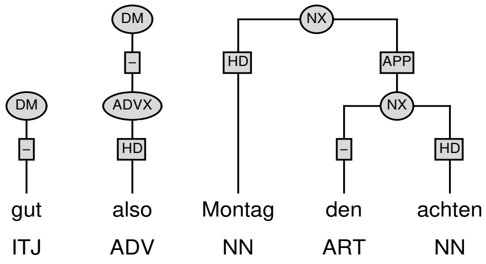

Abbildung 3. Eine NP mit Apposition aus der deutschen Verbmobil-Baumbank

interne NP richtig markiert wird; die Postmodifikation bzw. das durch die Apposition modifizierte Nomen werden jedoch durchgehend als nicht zu einer Base-NP gehörend markiert.

2.3 Drei Definitionen für Base-NPs

Aus dem Annotationsschema der deutschen Verbmobil-Baumbank wird deutlich, daß eine Definition von Base-NPs wie die von Ramshaw und Marcus einen Teil der Annotationen systematisch ausspart. Da bei der Nominalphrasenerkennung die zu klassifizierenden Wörter selbst einen sehr hohen Informationswert besitzen, treten Klassifikationsfehler nicht nur in den Fällen auf, bei denen eine Apposition oder eine Postmodifikation in dem zu klassifizierenden Beispiel vorliegt, sondern auch bei solchen Wörtern im allgemeinen, die häufig mit solchen Konstruktionen auftreten. Da in der deutschen Verbmobil-Baumbank gerade Wochentage gehäuft mit Appositionen auftreten, kann für Wörter wie *Montag* aus Abb. 3 nur sehr schwer entschieden werden, wann sie als I und wann als O annotiert werden müssen. Beide Fälle lassen sich in den Trainingsdaten finden, und die Information, ob sich eine Apposition anschließt, ist bei der Klassifikation nicht verfügbar.

Diese nicht abgedeckten Beispiele, die hier anhand des Annotationsschema der deutschen Verbmobil-Baumbank exemplarisch verdeutlicht wurden, finden sich ebenfalls in anderen Korpora, sofern das Annotationsschema nicht eine komplett flache Annotation der Nominalphrasen vorsieht oder von einer Anbindung postnominaler Modifikationen völlig absieht. In der Negra-Baumbank [6] z.B. ist das Annotationsschema sehr flach gewählt. Dort finden sich keine Strukturen, die in der deutschen Verbmobil-Baumbank den Postmodifikationen entsprechen, da diese Strukturen flach in die Nominalphrase gruppiert werden. Das Äquivalent zu Appositionen in der Verbmobil-Baumbank ist jedoch sehr häufig zu finden, u.a. bei Genitiv-Konstruktionen. Abb. 4 zeigt ein Beispiel für eine komplexe Genitiv-Konstruktion aus der Negra-Baumbank, bei der nach der Original-Definition der Base-NP weder *die Betreuung* noch *behinderter* als zu einer Base-NP gehörend annotiert würden.

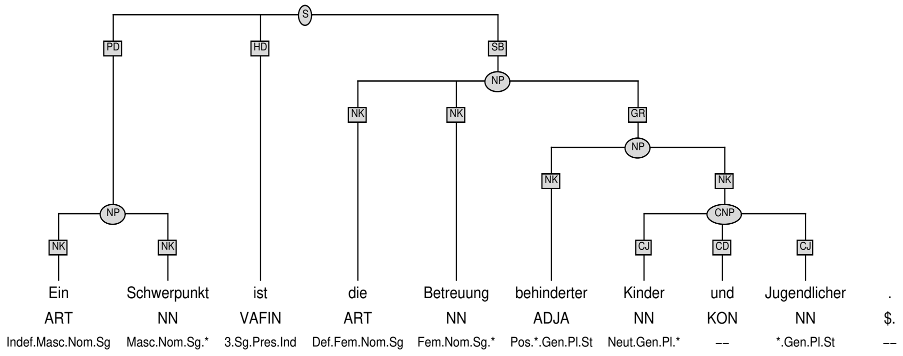

Abbildung 4. Satz 459 aus der Negra-Baumbank

Aus diesem Grund erscheint es sinnvoll, die Definition linguistisch motiviert so auszuweiten, daß auch Postmodifikationen und Appositionen systematisch abgedeckt werden.

Im folgende sollen die drei verschiedenen Definitionen vorgestellt werden, die in den Versuchsreihen verwendet wurden.

1. Die Original-Definition von Ramshaw und Marcus [4]: es wurden nur solche NPs als I markiert, die keine weiteren NPs enthalten.
   Damit wurden die Beispiele aus Abb. 1-3 folgendermaßen annotiert: *wir/I müßten/O also/O insgesamt/I drei/I Arbeitssitzungen/I festlegen/O*; *sechzehn/I Uhr/I fünfzehn/O*; *gut/O also/O Montag/O den/I achten/I*.
2. Erweiterung um NPs mit Appositionen: in diesem Fall werden NPs, die eine Apposition enthalten, abzüglich der Apposition, als Base-NP annotiert. Prämodifizierende Präpositionalphrasen werden jedoch als nicht zur NP gehörend gewertet.
   Das Appositionsbeispiel wurde daher folgendermaßen annotiert:
   gut/O also/O Montag/I den/B achten/I.
3. Erweiterung um Postmodifikationen: jetzt werden auch Postmodifikationen als eigenständige Base-NPs annotiert (dies ist im Annotationsschema nicht vorgesehen). Ausgenommen sind hier postmodifizierende PPs, da sie zu rekursiven Strukturen führen würden.
   Das Postmodifikationsbeispiel wurde daher folgendermaßen annotiert:
   sechzehn/I Uhr/I fünfzehn/B.

Auch bei dieser letzten Definition bleiben Problemfälle, die sich mit der IOB-Darstellung nicht zufriedenstellend lösen lassen. Zum einen gehören NPs mit prämodifizierenden Partizipialkonstruktionen dazu, z.B. *das billigste der drei zur Verfügung stehenden Hotels*. Hier läßt sich die äußere NP nicht ohne Rekursion annotieren. In solchen Fällen wird die äußere NP aufgespalten: *das/I billigste/I der/B drei/I zur/O Verfügung/I stehenden/B Hotels/I*. Die so entstandene NP *der drei* entspricht der Annotation einer elliptischen NP.

Ein zusätzliches Problem bei der Annotation von Base-NPs ergibt sich aus der Natur der hier verwendeten Korpusdaten, da in der deutschen Verbmobil-Baumbank transliterierte gesprochene Sprache annotiert wurde. Aus diesem Grund sind Wiederholungen und Satzabbrüche sehr häufig. Vor allem Beispiele, wie in Abb. 5 mit der Wiederholung *fünfzehn fünfzehn*, bei der die eigentliche Nominalphrase aus der Wortsequenz *nur fünfzehn Minuten* bestehen sollte, können nicht adäquat als Base-NPs abgebildet werden. Hier kann das wiederholte Wort zu einer NP projeziert werden, die den linearen Ablauf der sie umgebenden NP unterbricht. Auch hier wird die äußere NP aufgespalten: *man/I hat/O nur/I fünfzehn/B fünfzehn/B Minuten/I zum/O Zentrum/I.* In anderen Fällen, bei denen das wiederholte Wort nicht projeziert werden kann, führt dies zu Trainingsinstanzen, die einem Wort, das typischerweise in einer Nominalphrase auftritt, das IOB-Tag O zuweisen.

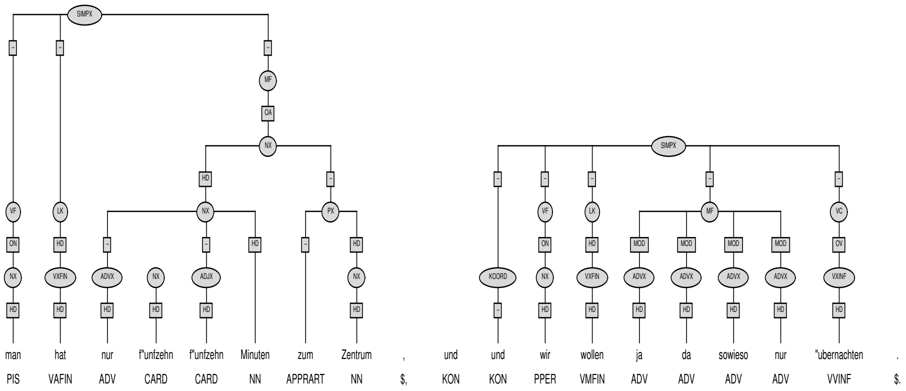

Abbildung 5. Ein Beispiel für Wiederholungen

2.4 Das Testsetting

Beim MBL wird die Klassifizierung einer Testinstanz durch die Ähnlichkeit zu Trainingsinstanzen erreicht. Eine Instanz besteht aus einer Reihe von Merkmalen, über die die Ähnlichkeit errechnet wird. Als Merkmale zur Bestimmung der Ähnlichkeit beim MBL dienen jeweils das zu klassifizierende Wort, sein POS-Tag und ein Kontext von einem, zwei oder drei Wörtern, mit ihren POS-Tags, jeweils rechts und links des Wortes. Die POS-Tags wurden aus der Baumbank extrahiert, es wird das STTS [5] verwendet.

Der in Abb. 3 gezeigte Satz wird (bei einem Kontext von einem Wort) in der Instanzenbasis wie in Abb. 6 dargestellt. Jedes zu klassifizierende Wort ergibt eine Instanz; ist kein Kontext vorhanden (am Anfang oder am Ende eines Satzes), wird -NONE- eingetragen. Die Spalten enthalten jeweils den Wert eines

Merkmals, die letzte Spalte enthält die Klasse, der diese Instanz angehört. In der ersten Zeile in Abb. 6 z.B. erhält das Wort *gut* im angegebenen Kontext das IOB-Tag O.

LK Wort	**akt. Wort**	RK Wort	LK POS-Tag	**akt. POS-Tag**	RK POS-Tag	IOB-Tag
-NONE-	gut	also	-NONE-	ITJ	ADV	O
gut	also	Montag	ITJ	ADV	NN	O
also	Montag	den	ADV	NN	ART	I
Montag	den	achten	NN	ART	NN	B
den	achten	-NONE-	ART	NN	-NONE-	I

Abbildung 6. Instanzen für die Wörter des Satzes aus Abb. 3

Als Lernsystem wurde TiMBL [1] verwendet, als Lernalgorithmus IB1 mit dem Ähnlichkeitsmaß *overlap metric*.

Es wurden für jede der drei Definitionen und für jede Kontextgröße ein Testlauf durchgeführt. Zur Absicherung der Ergebnisse wurde für die einzelnen Testläufe eine “tenfold cross-validation” angewandt, d.h. das Korpus wurde in 10 Teile aufgeteilt und jeweils neun Teile zum Training und ein Teil zum Testen verwendet. Dadurch ergaben sich Testsets von jeweils ca. 41 000 Wörtern.

3 Ergebnisse

Bei den Tests mit den drei Definitionen einer Base-NP ergaben sich die in Tabelle 1 gezeigten Werte, es handelt sich jeweils um die gemittelten Werte über die zehn Versuche. Precision, Recall und F-Score wurden nicht bezüglich der Wörter sondern bezüglich der Phrasen berechnet. Dies bedeutet, daß eine komplette Nominalphrase als falsch gerechnet wird, sobald mindestens eines der dazugehörenden Wörter falsch IOB-getaggt wurde. Accuracy, dagegen, wird auf Wort-Basis berechnet.

Die hohen Werte schon bei der Original-Definition lassen sich zum einen durch die hohe Anzahl an Trainingsinstanzen von ca. 335 000 Wörtern erklären, zum anderen durch die geringe durchschnittliche Länge der Base-NPs von 1,3 Wörtern. Da es sich bei den hier verwendeten Daten um gesprochene Sprache handelt, sind komplexe NPs nicht sehr häufig vertreten, während Personal- und Demonstrativpronomina überproportional häufig vorkommen.

Die Verbesserung, die sich durch die Erweiterung der Definition ergibt, wird deutlicher, wenn man einen Blick auf die Anzahl der falsch klassifizierten Wörter wirft. Tabelle 2 gibt die Zahlen für die drei Definitionen bei drei Wörtern Kontext an.

Hier wird deutlich, daß sich die durchschnittliche Anzahl der falsch IOB-getaggten Wörter durch die linguistisch motivierte Definition der Base-NP, bei der sowohl Appositionen als auch Postmodifikationen systematisch eingeschlossen werden, um ungefähr ein Drittel verringert hat.

Tabelle 1. Die quantitative Evaluation

NP Definition	Kontextgröße	Accuracy	Precision	Recall	F-Score
Original-Definition	je 1 Wort	99,23 %	97,44 %	97,33 %	97,38
	je 2 Wörter	99,27 %	97,62 %	97,47 %	97,55
	je 3 Wörter	99,26 %	97,64 %	97,45 %	97,54
+ Appositionen	je 1 Wort	99,37 %	97,73 %	97,78 %	97,75
	je 2 Wörter	99,38 %	97,81 %	97,81 %	97,81
	je 3 Wörter	99,36 %	97,79 %	97,75 %	97,77
+ Postmodifikationen	je 1 Wort	99,54 %	98,33 %	98,36 %	98,34
	je 2 Wörter	99,52 %	98,33 %	98,32 %	98,33
	je 3 Wörter	99,50 %	98,23 %	98,18 %	98,21

Tabelle 2. Absolute Anzahl der falsch IOB-getaggten Wörter

Definition	minimale Anzahl	maximale Anzahl	Durchschnitt
Original-Definition	254	302	275,1
+ Appositionen	217	262	237,8
+ Postmodifikationen	168	206	188,9

3.1 Fehleranalyse

Um zu sehen, wie sich die verschiedenen Definitionen der Base-NP auf die Art der Fehler auswirken, wurde für jede Definition eine Fehleranalyse für ein Testset gemacht, bei der eine Klassifizierung der Fehler in sieben Kategorien vorgenommen wurde. Die Kategorien spiegeln soweit möglich die Gründe der Fehlklassifikation wider. Als Fehlerkategorien wurden die folgenden gewählt: Fgold (Fehler im Gold Standard), Idef (Fehler durch Base-NP-Definition), Koord (Koordination), Prae (adverbialer Prämodifikator), Traindef (Fehler in der Trainingsinstanz durch Base-NP-Definition), Unpass (unpassende Instanz ausgewählt) und Wiederh (Wiederholung, gesprochene Sprache). Bei einer Koordination ist aus dem Kontext von maximal 3 Wörtern oft nicht zu erkennen, welchen Skopus die Koordination umfaßt. Adverbiale Prämodifikatoren stellen ein häufiges Problem dar, da der Klassifizierungsalgorithmus sie aufgrund der gegebenen Informationen nicht von satzmodifizierenden Adverbien unterscheiden kann. "Unpassende" Trainingsinstanzen sind häufig solche, die sich nur minimal dadurch von der zu klassifizierenden Instanz unterscheiden, daß das direkt vorhergehende Wort nicht derselben IOB-Klasse angehört. Soll das Wort *die* in der Instanz *im Juni sich die Galerie ansehen* durch die Instanz *im Juni gemeinsam die Galerie ansehen* klassifiziert werden, ergibt sich das falsche IOB-Tag I statt B aus dem Unterschied zwischen der Nominalphrase *sich* und dem Adverb *gemeinsam*.

Aus den Zahlen in Tabelle 3 wird deutlich, daß Fehlklassifikationen, die auf zu eingeschränkt definierte Base-NPs zurückzuführen sind (d.h. die Kategorien Idef und Traindef), nach der Anpassung fast vollständig beseitigt wurden. Die noch

Tabelle 3. Die Fehlerkategorien bei verschiedenen Base-NP-Definitionen

Fehlerart	Original-Definition	+ Appositionen	+ Postmodifikationen
Fgold	19 (6.29%)	16 (5,11%)	13 (6,60%)
Idef	51 (16,89%)	32 (12,21%)	2 (1,02%)
Koord	6 (1,99%)	6 (2,29%)	6 (3,05%)
Prae	60 (19,87%)	59 (22,52%)	60 (30,46%)
Traindef	59 (19,54%)	32 (12,21%)	8 (4,06%)
Unpass	70 (23,18%)	80 (30,53%)	70 (35,53%)
Wiederh	37 (12,25%)	37 (14,12%)	38 (19,29%)

bestehenden Fehler der Kategorien Idef und Traindef sind auf Fälle zurückzuführen, bei denen zwei Base-NPs durch einen hoch angebundenen Artikel prämodifiziert werden. Beispiele für solche Konstruktionen sind *[NP die [NP Frau] [NP Ministerin]]* oder Koordinationen wie *[NP meine [NP Telefonnummer] und [NP Email-Adresse]]*. Sie wurden in der Definition nicht berücksichtigt.

4 Schluß

Mithilfe der hier beschriebenen Versuche zum Maschinellen Lernen von Base-Nominalphrasen konnte nachgewiesen werden, daß bei der Definition der Lernaufgabe der Einsatz von linguistischem Wissen zu einer Verbesserung der Ergebnisse führen kann. In den hier beschriebenen Experimenten konnte der F-Score-Wert bei einer Baseline von rund 97,5 % um ca. 1 % erhöht werden. Die Anzahl der fehlklassifizierten Wörter konnte um ungefähr ein Drittel gesenkt werden.

Es ist zu erwarten, daß sich diese Verbesserung in solchen Fällen noch deutlicher zeigt, bei denen die Baseline niedriger liegt. Diese Annahme soll in Zukunft anhand der englischen Verbmobil-Baumbank [2], die eine größere Anzahl solcher Konstruktionen enthält, und der (deutschen) Negra-Baumbank [6], in der die NPs komplexer und länger sind, überprüft werden.

Literatur

1. W. Daelemans, J. Zavrel, K. van der Sloot, and A. van den Bosch. TIMBL: Tilburg memory based learner - version 3.0 - reference guide. Technical Report ILK 00-01, Induction of Linguistic Knowledge, Computational Linguistics, Tilburg University, 2000. http://ilk.kub.nl/ ilk/papers/ilk0001.ps.
2. V. Kordoni. Stylebook for the English Treebank in VERBMOBIL. Technical Report 241, Verbmobil, 2000.
3. M. Marcus, B. Santorini, and M. A. Marcinkiewicz. Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics*, 19(2):313 – 330, 1993.
4. L. Ramshaw and M. Marcus. Text chunking using transformation-based learning. In *Proceedings of the ACL Third Workshop on Very Large Corpora*, pages 82 – 94, 1995.

5. A. Schiller, S. Teufel, and C. Thielen. Guidelines für das Tagging deutscher Textkorpora mit STTS. Technical report, Universität Stuttgart and Universität Tübingen, September 1995. (URL: http://www.sfs.nphil.uni-tuebingen.de/Elwis/stts/stts.html).
6. W. Skut, B. Krenn, T. Brants, and H. Uszkoreit. An annotation scheme for free word order languages. In *Proceedings of the Fifth Conference on Applied Natural Language Processing (ANLP), Washington, D.C.*, 1997.
7. R. Stegmann, H. Telljohann, and E. W. Hinrichs. Stylebook for the German Treebank in VERBMOBIL. Technical Report 239, Verbmobil, 2000.
8. E. Tjong Kim Sang and S. Buchholz. Introduction to the CoNLL shared task: Chunking. In *Proceedings of CoNLL and LLL 2000, Lisbon, Portugal*, pages 127 – 132, 2000.
9. E. Tjong Kim Sang and J. Veenstra. Representing chunks. In *Proceedings of the Ninth EACL, Bergen, Norway*, pages 173 – 179, 1999.