

# NeGra und TüBa-D/Z – Ein Vergleich

Wolfgang Maier

Universität Tübingen  
wmaier@sfs.uni-tuebingen.de

**Zusammenfassung** Baumbanken sind nicht nur für die Linguistik von Wert, sie dienen auch als Basis und Datengrundlage für die Entwicklung von NLP-Applikationen wie z.B. Parsern. Wenn verschiedene Baumbanken zur Auswahl stehen, wie im Deutschen, stellt sich die Frage, inwieweit sie sich unterscheiden, insbesondere im Hinblick auf eine bestimmte Anwendung wie Parsing. In dieser Arbeit werden zwei Baumbanken des Deutschen, NeGra und TüBa-D/Z, verglichen. Das Annotationsschema einer Baumbank wird an das Annotationsschema der jeweils anderen angenähert, während das Verhalten der veränderten Versionen mit einem PCFG-Parser geprüft wird. Zum einen kann auf diese Weise darauf geschlossen werden, welche Rolle die Baumbankannotation bei PCFG-Parsing spielt, zum anderen macht der Zugewinn an Information über die beiden Baumbanken aus eben diesen wertvollere Ressourcen.

## 1 Einleitung

Baumbanken, syntaktisch annotierte Korpora, sind nicht nur für die Linguistik von Wert, sie dienen auch als Basis und Datengrundlage für die Entwicklung von NLP-Applikationen, wie z.B. Parsern. Heute verfügen wir über eine große Zahl von Baumbanken großen Umfangs in verschiedenen Sprachen. Daher stellt sich uns die Frage, welche Baumbank die für eine bestimmte Anwendung geeignetste darstellt, und welche Rolle die einzelnen Eigenschaften einer Baumbank im Kontext einer solchen Anwendung spielen.

Mehrere Arbeiten beschäftigen sich aus diesem Grund mit dem Vergleich von Baumbanken, wobei sich ein großer Teil der Forschung auf einen einzelnen Bereich konzentriert: das Parsing. Haben die Arbeiten zu statistischem baumbankbasiertem Parsing lange Zeit nur eine einzige Baumbank herangezogen (den Wall-Street-Journal-Teil (WSJ) der Penn Treebank), so sind in letzter Zeit doch Arbeiten entstanden, die sich insbesondere mit dem Zusammenhang zwischen verschiedenen Baumbanken (auch nicht-englischen) befassen. Gildea (2001) erforscht die Auswirkungen, die bestimmte Baumbankeigenschaften, wie die Verteilung von Subkategorisierungsrahmen, auf Parsingergebnisse haben. Er führt Experimente mit dem WSJ und dem Brown-Korpus durch, wobei er eine Baumbank für das Training seines Parsers verwendet und auf der jeweils anderen parst. Dabei zieht er den Schluss, dass eine kleine Menge „passender“ Trainingsdaten besser ist als eine große Menge „unpassender“ Daten. Dubey and Keller (2003) analysieren die Schwierigkeiten, die sich beim Parsen des Deutschen ergeben. Sie

benutzen die NeGra-Baumbank für ihre Experimente und zeigen, dass Lexikalisierung, trotz ihres Nutzens für das Parsen des Englischen, für das Deutsche nicht zu besseren Ergebnissen führt. Basierend auf dieser Erkenntnis entwickeln sie ein Parsing-Modell für das Deutsche, das auf Schwester-Kopf-Abhängigkeiten basiert. Corazza et al. (2004) führen Experimente mit Modell 2 von Collins' Parser (Collins, 1999) sowie mit dem Stanford-Parser (Klein and Manning, 2002, 2003) auf zwei italienischen Baumbänken durch. Sie erreichen keine guten Ergebnisse, was sie auf den unterschiedlichen Schwierigkeitsgrad unterschiedlicher Parsing-Teilaufgaben in Englisch und Italienisch, sowie auf verschiedene Annotationsstile in beiden Baumbanken zurückführen.

In der vorliegenden Arbeit werden zwei deutsche Baumbanken, NeGra und TüBa-D/Z, auf ihr Verhalten bei PCFG-Parsing untersucht. Dazu orientieren wir uns an der Methode von Kübler (2005): Die Annotation einer Baumbank wird an die Annotation der jeweils anderen angenähert, indem schrittweise einzelne Annotationskomponenten entfernt bzw. hinzugefügt werden. Das Verhalten jeder (veränderten) Baumbankversion wird mit einem PCFG-Parser evaluiert. Wir erreichen damit zwei Ziele. Zum einen erlangen wir neue Erkenntnisse über PCFG-Parsing im Allgemeinen, da wir mit einer anderen Sprache als Englisch und einer anderen Baumbank als dem WSJ arbeiten. Zum anderen erhöhen wir den Nutzwert beider Baumbanken im Hinblick auf Parsing, da wir mehr über die Eigenschaften der einzelnen Teile ihrer Annotation erfahren.

In Abschnitt 2 wird zunächst die Annotation beider Baumbanken beschrieben. Abschnitt 3 stellt die Vergleichsmethodik vor, die verwendet wird. In Abschnitt 4 wird der Versuchsaufbau vorgestellt, die Ergebnisse werden bewertet. Ein Schlusswort in Abschnitt 5 schließt die Arbeit ab.

## 2 Die Baumbanken

Mehrere Baumbanken in deutscher Sprache sind verfügbar, darunter zwei vergleichbare, NeGra (Skut et al., 1997), entstanden an der Universität des Saarlandes in Saarbrücken und TüBa-D/Z (Telljohann et al., 2003), entstanden an der Universität Tübingen. NeGra enthält ca. 20.000 Sätze, TüBa-D/Z<sup>1</sup> ca. 15.000, beide sind annotierte Zeitungskorpora. In beiden Baumbanken ist unter der Prämisse der Theorienneutralität die Prädikat-Argument-Struktur annotiert. Terminalknoten sind mit Part-of-Speech- und mit morphologischer Information versehen, nicht-terminale Knoten mit Phrasenetiketten. Alle Kanten sind mit grammatischen Funktionen etikettiert. Die Annotation wurde bei beiden Baumbanken halbautomatisch mit denselben Softwarewerkzeugen durchgeführt.

Der Hauptunterschied zwischen beiden Baumbanken ist in der Behandlung der freien Wortstellung des Deutschen begründet. Die Stellung von Komplementen und Adjunkten ist im Deutschen sehr variabel, was zu einer großen Zahl diskontinuierlicher Konstituenten führt, auch in kurzen Sätzen. Ein Annotationsschema für das Deutsche muss damit umgehen können. NeGra erlaubt

---

<sup>1</sup> in der vorliegenden Version

daher kreuzende Kanten, der kontextfreie Kern der Annotation wird aufgegeben. Da mit kreuzenden Kanten alle Tochterknoten aller Konstituenten, ob diskontinuierlich oder nicht, immer unter demselben Knoten zusammengefasst werden können, stellen diskontinuierliche Konstituenten keine Schwierigkeit mehr dar. Der Nachteil der kreuzenden Kanten ist, dass sie aufgelöst werden müssen, bevor die Baumbank mit einem PCFG-Parser benutzt werden kann. Das kann jedoch leicht durch das Umhängen von Tochterknoten diskontinuierlicher Konstituenten an höhere Knoten erreicht werden.

In TüBa-D/Z wird die freie Wortstellung des Deutschen anders behandelt. Anstatt diskontinuierliche Konstituenten zuzulassen, wird das in der deskriptiven Syntax weit verbreitete und sehr gut ausgearbeitete Konzept der *topologischen Felder* als zentrales Ordnungsprinzip im Satz eingeführt (vgl. Drach, 1937, Höhle, 1986). Die topologischen Felder bilden eine zusätzliche Gliederungsebene zwischen Satz- und Phrasenebene. Die empirische Motivation für die topologischen Felder liefern die drei Satztypen des Deutschen, die sich jeweils durch die Stellung des finiten Verbs unterscheiden. Die Felderannotation in TüBa-D/Z basiert auf Höhle (1986), der die üblichen fünf Felder (Vorfeld, linke Satzklammer, Mittelfeld, rechte Satzklammer, Nachfeld) um vier weitere ergänzt, für Linkversetzungen, Komplementierer sowie koordinierende und nicht-koordinierende Partikel. Da Wortstellungsvariationen sich meist auf den Bereich eines einzelnen topologischen Feldes beschränken, kann die freie Wortstellung des Deutschen auf natürliche Art beschrieben werden.

Auch auf der Ebene der Phrasenannotation weisen die beiden Baumbanken große Unterschiede auf. NeGra lässt keine Zwischenprojektionen („*bar*“-Projektionen) von Phrasen zu. Des Weiteren werden keine unären Produktionen zugelassen. Dies führt zu einer sehr flachen Phrasenstruktur. Prä- und Postmodifizierer werden direkt an den Phrasenknoten gehängt, Subjekte werden direkt an den Satz gehängt, nominales Material innerhalb von PPs projiziert nicht zu NPs, komplexe (nicht-koordinierte) NPs bleiben flach. Im Gegensatz dazu wird in TüBa-D/Z „tiefe“ Annotation erlaubt. Zwischenprojektionen und unäre Produktionen sind erlaubt und werden ausgiebig benutzt.

Die Abbildung 1 zeigt jeweils einen annotierten Satz aus TüBa-D/Z und NeGra, zur Illustration der verschiedenen Annotationsprinzipien beider Baumbanken. Im flach annotierten NeGra-Baum sehen wir eine kreuzende Kante, während im TüBa-D/Z-Baum die Felderannotation und eine tiefe Annotation auffallen.

### 3 Parserbasierter Baumbankvergleich

Wie können wir die Eigenschaften der beiden Baumbanken bzw. ihrer Annotationsschemata und deren Komponenten bei der Verwendung mit Parsern überhaupt jenseits eines simplen Vergleichs der Parsingergebnisse vergleichend bewerten? Die Parserleistung könnte im Kontext einer konkreten Anwendung ausgewertet werden, die die Parserausgabe verwendet. Verschiedene Anwendungen bieten sich an, so z.B. die automatische Auflösung von Anaphern (*Computational Anaphora Resolution*). Ein Vergleich dieser Art hätte zwar Aussagekraft,

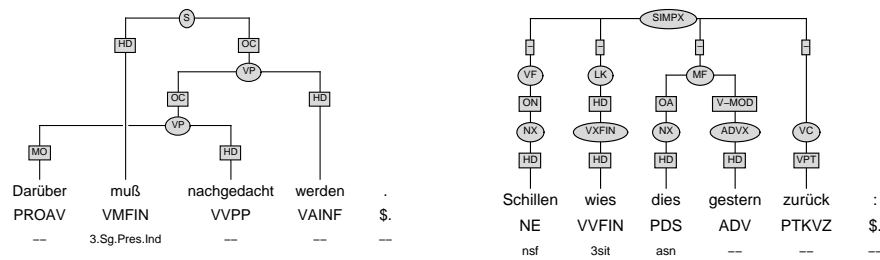


Abbildung 1. Sätze aus NeGra (links) und TüBa-D/Z (rechts)

aber nur im Kontext einer entsprechenden Anwendung. Eine darauf basierende Generalisierung wäre schwierig. Wir könnten außerdem nur Aussagen über ein Annotationsschema als Ganzes treffen, Aussagen über einzelne Annotationskomponenten wären nicht möglich.

Eine andere, naheliegende Idee ist, Parsingergebnisse für einzelne Konstituententypen direkt zu vergleichen. Dies würde jedoch aufgrund der völlig unterschiedlichen Strukturierung der Annotationsschemata keine sinnvollen Ergebnisse erbringen, da in beiden Baumbanken verschiedene Phrasenetiketten mit vermeintlich gleicher Funktionalität unterschiedliche Mengen von Konstituenten abdecken. Ein Beispiel hierfür sind Nominalphrasen. Des Weiteren sind in beiden Annotationsschemata Etiketten vorhanden, die im jeweils anderen Annotationsschema keine Entsprechung haben. NeGra enthält keine Etiketten, die TüBa-D/Zs Felderetiketten entsprechen, TüBa-D/Z wiederum enthält keine Etiketten, die den NeGra-Etiketten für koordinierte Phrasen entsprechen.

Gerade an der unterschiedlichen Strukturierung der Annotationsschemata muss aber ein Vergleich ansetzen. Wir orientieren uns an der Methode von Kübler (2005) und führen Experimente mit modifizierten Versionen der Baumbanken durch, in denen die Annotation durch Hinzufügen bzw. Entfernen von Annotationskomponenten an die jeweils andere Baumbank angenähert wird. Die veränderten Versionen sollen darüber Aufschluss geben, welche Elemente der Annotationsschemata welche Effekte auf die Parsing-Ergebnisse haben und eine Grundlage schaffen, auf der NeGra und TüBa-D/Z verglichen werden können.

## 4 Die Experimente

Wir benutzen für unsere Experimente `lpar` (Schmid, 2000), einen Standard-PCFG-Parser. Grammatik, Lexikon und die Anzahl des Auftretens deren Elemente werden direkt von den Bäumen gelesen. Dem Parser werden in der Eingabe Gold-POS-Tags mitgegeben, um Parsing-Fehler zu vermeiden, die durch falsche POS-Tags verursacht werden. Bedingt durch Speichermangel können nur Sätze kürzer als 41 Wörter berücksichtigt werden.

Für die Experimente mit `lopar` sollen jeweils 90% der Baumbanken und ihrer verschiedenen Versionen als Trainingsset und 10% als Testset verwendet werden. Um eine möglichst homogene Verteilung linguistischer Phänomene auf Trainings- und Testset zu erreichen, wird jeder 10. Satz der Baumbanken in das Testset geschoben und die restlichen Sätze zum Trainingsset deklariert.

#### 4.1 Vorbereitung der Baumbanken

Da wir die Grammatiken für unseren Parser direkt von den Bäumen der Baumbanken lesen wollen, müssen wir die Baumbanken vorverarbeiten, da die Originalannotation aus mehreren Gründen nicht vollständig kontextfrei ist. In beiden Baumbanken ist die Zeichensetzung nicht in die Analyseebenen eingebunden, des Weiteren orientiert sich die Satzannotation nicht immer am linguistischen Begriff des Satzes. Dies führt in TüBa-D/Z wie in NeGra zu Sätzen, die aus mehreren nicht miteinander verbundenen Bäumen bestehen. Alle Knoten in einem Satz, d.h. Wurzelknoten und die Zeichensetzung, werden durch einen virtuellen Wurzelknoten gebunden, der diskontinuierlich sein kann. Des Weiteren erlaubt NeGras Annotationsschema aus linguistischen Gründen kreuzende Kanten, wie in Abschnitt 2 beschrieben.

Alle kreuzenden Kanten müssen vor Training bzw. Parsing entfernt werden. Die kreuzenden Kanten des NeGra-Annotationsschemas werden mit einem Programm von Thorsten Brants aufgelöst, das Töchter von diskontinuierlichen Knoten an höhere Knoten im Baum hängt. Unabhängig davon wird eine evtl. vorhandene Diskontinuität des virtuellen Wurzelknotens dadurch aufgelöst, dass alle Zeichensetzung an die höchstmögliche Stelle im Baum gehängt wird. Zusammengehörige Zeichen (Klammern und Anführungszeichen) werden bevorzugt an denselben Knoten gehängt, um niederfrequente Produktionen in der Grammatik zu vermeiden, die sich nur durch die unterschiedliche Position von Klammern bzw. Anführungszeichen auf ihrer rechten Seite unterscheiden.

#### 4.2 Ergebnisse des Baumbankvergleichs

Durch die an den Baumbanken durchzuführenden Modifikationen verändern sich deren Eigenschaften, die Anzahl der Knoten wird verändert, was einen Einfluss auf Breite und Tiefe der Bäume hat. Dies sollen drei Maße erfassen, die in der Tabelle 1 für beide Baumbanken und alle Modifikationen aufgeführt sind: die durchschnittliche Anzahl von Töchtern nichtterminaler Knoten ( $\mu T/K$ ), die durchschnittliche Baumhöhe<sup>2</sup> ( $\mu H(T)$ ) und das Knoten/Token-Verhältnis ( $K/W$ ), d.h. das Verhältnis nichtterminaler und terminaler Knoten.

Die Parserausgabe wird mit der üblichen *parseval*-Metrik ausgewertet. Diese misst den Prozentsatz der korrekt geparsten Konstituenten mit den Maßen Precision, Recall und F-Measure. Die Parser-Ausgabe jeder veränderten Baumbankversion wird gegen eine entsprechend modifizierte Version des Testsets ausgewertet. Ungeparste Sätze werden in die Auswertung eingeschlossen.

<sup>2</sup> Als Baumhöhe wird hier der maximale Abstand eines terminalen Knotens zur Wurzel (VROOT) bezeichnet

	NeGra	NE_f	NE_NP	TüBa	Tü_NF	Tü_NU	Tü_f	Tü_f_NU	Tü_f_NU_NF
K/W	0,41	0,70	0,50	1,21	0,89	0,54	1,00	0,42	0,35
$\mu$ T/K	2,92	2,22	2,59	1,61	1,89	2,53	1,83	2,93	3,35
$\mu$ H(T)	4,86	5,81	5,16	6,88	5,68	5,45	5,94	4,72	4,15

**Tabelle 1.** Eigenschaften der einzelnen Baumbankversionen<sup>3</sup>

	NeGra	NE_f	NE_NP	NeGra	NE_f	NE_NP	NE_trace
	<i>ohne gr. Funktionen</i>			<i>mit grammatischen Funktionen</i>			
crossing brackets	1,10	1,67	1,14	1,10	1,21	1,27	1,05
labeled precision	68,14%	74,96%	70,43%	52,67%	67,90%	59,77%	51,81%
labeled recall	69,98%	70,37%	72,81%	52,17%	65,18%	60,36%	49,19%
labeled F <sub>1</sub>	69,05	72,59	71,60	52,42	66,51	60,06	50,47%
not parsed	1,00%	0,10%	0,15%	12,90%	1,66%	9,88%	16,01%

**Tabelle 2.** Parsingergebnisse für NeGra

**NeGra** Neben der unveränderten Baumbank untersuchen wir zwei Modifikationen von NeGra, *NE\_f* und *NE\_NP*. In *NE\_f* wird eine **Annotation topologischer Felder**<sup>4</sup> hinzugefügt. Diese führt zu mehr kreuzenden Klammern, was auf die gestiegene Anzahl der zu erkennenden Konstituenten zurückzuführen ist: das Knoten/Token-Verhältnis steigt auf 0,70 gegenüber 0,41 in der Basisversion. Mit *NE\_f* steigt der Precision-Wert, beim Parsen mit grammatischen Funktionen steigen Precision und Recall gegenüber der Basisversion gleichermaßen an. Die Regeln, die durch die Ebene der topologischen Felder hinzukommen und Phrasen unterhalb der Satzebene zusammenfassen, sind also nützlich für das Parsing.

Ein Blick in die Auswertung der einzelnen Kategorien zeigt, dass alle neu eingeführten Felderkategorien hohe F-Score-Werte erreichen (z.B. 97,79 F-Score für LK), also offensichtlich einfach zu finden sind. Beinahe alle Kategorien zeigen einen erhöhten Precision-Wert. Auffallend ist aber v.a. ein Abfall von Precision und von Recall für Verbalphrasen (von 59,41 auf 26,70 F-Score). Dies liegt an der hinzugefügten Felder-Annotation: Nur VPs, die sich nicht über mehrere Felder erstrecken, sind annotiert.

In *NE\_NP* wird **PP-internes Nominalmaterial zu NPs zusammenfasst**. Auch diese Modifikation spiegelt sich deutlich in den Baumbankdimensionen wider (siehe Tabelle 1). Precision und Recall steigen leicht an (ca. 2-3%). Obwohl mehr Klammern erzeugt werden müssen, steigt die Anzahl der kreuzenden Klammern nur leicht. Dies liegt daran, dass der Raum unterhalb von PPs zu klein ist, um eine größere Zahl falscher Klammern entstehen zu lassen.

<sup>3</sup> Die einzelnen Baumbankversionen, bezeichnet durch die durch die in dieser Tabelle verwendeten Abkürzungen, werden im folgenden Text vorgestellt.

<sup>4</sup> Die Felderannotation für die NeGra-Baumbank wurde mir dankenswerterweise vom DFKI Saarbrücken zur Verfügung gestellt.

Abschließend parsen wir eine Version von NeGra, in der für jede Knotenbewegung während der Auflösung der kreuzenden Kanten **eine Spur** in der entsprechenden Kante erzeugt wurde (*NE\_trace*). Obwohl wir die Baumbank damit an das Format von TüBa-D/Z annähern, verschlechtern sich die Parsingergebnisse gegenüber der unveränderten Version. Die hohe Zahl ungeparster Sätze weist jedoch darauf hin, dass dieses Ergebnis wegen Datenknappheit nicht verlässlich ist.

	TüBa	Tü_NF	Tü_NU	Tü_f	Tü_f_NU	Tü_f_NU_NF
<i>ohne grammatische Funktionen</i>						
crossing brackets	2,21	1,82	1,67	1,04	0,80	1,03
labeled precision	87,39%	86,31%	79,97%	86,22%	75,18%	63,05%
labeled recall	83,57%	83,43%	78,52%	85,41%	76,11%	66,86%
labeled F <sub>1</sub>	85,44	84,85	79,24	85,81	75,64	64,90
not parsed	0,07%	0,07%	2,45%	0,07%	2,99%	6,87%
	TüBa	Tü_NF	Tü_NU	Tü_f	Tü_f_NU	Tü_f_NU_NF
<i>mit grammatischen Funktionen</i>						
crossing brackets	1,84	1,82	1,79	0,98	1,01	1,12
labeled precision	76,99%	68,55%	63,71%	76,93%	58,91%	45,15%
labeled recall	75,30%	68,40%	62,79%	77,21%	58,92%	44,76%
labeled F <sub>1</sub>	76,14	68,47	63,25	77,07	58,92	44,96
not parsed	0,07%	0,27%	4,49%	0,07%	7,21%	17,76%

**Tabelle 3.** Parsingergebnisse für TüBa-D/Z

**TüBa-D/Z** Neben der unveränderten Baumbank testen wir fünf veränderte Versionen von TüBa-D/Z. In jeder Version wird Material entfernt, um zu NeGra-ähnlichen Strukturen zu kommen. Zunächst löschen wir die Annotation der **topologischen Felder**, da diese auch in NeGra nicht vorhanden ist (*Tü\_NF*). Dies führt zu einem nur geringen Abfall der Parsingergebnisse. Ein genauere Blick auf die Ergebnisse der einzelnen Kategorien zeigt, dass vor allem die Kategorien der Satzebene verlieren. Strukturen innerhalb der topologischen Felder verschlechtern sich nicht. Wir können daraus schließen, dass die Annotation der topologischen Felder vor allem für die Satzebene wichtig ist.

In der zweiten veränderten Version von TüBa-D/Z, *Tü\_NU*, werden **unäre Produktionen** entfernt, um NeGra-ähnliche Strukturen zu erzeugen. Die Parsingergebnisse zeigen, dass unäre Knoten sehr nützlich sind, ohne sie fällt der F-Measure-Wert um ca. 6 Punkte. Die Anzahl der kreuzenden Klammern fällt jedoch auch, was mit der Tatsache zusammenhängt, dass die Bäume in dieser Baumbankversion insgesamt weniger Knoten enthalten (siehe Tabelle 1). Beim

Parsen mit grammatischen Funktionen hat das Entfernen der unären Produktionen einen besonders großen Effekt: Der F-Measure-Wert sinkt um ca. 13 Punkte. Eine plausible Erklärung hierfür könnte Datenknappheit sein. 32,78% der Regeln, die der Parser braucht um korrekte Bäume zu erzeugen, kommen nicht im Trainingsset vor.

Eine Auswertung der Ergebnisse für die einzelnen Kategorien zeigt, dass bei allen wichtigen Phrasenkategorien sowohl Precision als auch Recall sinkt. Knoten, die topologische Felder repräsentieren, sind nicht betroffen, da die meisten von ihnen unär sind und daher in dieser Version der Baumbank gelöscht werden. Die meisten Mittelfeldknoten bleiben jedoch bestehen, da das Mittelfeld i.d.R. mehr als ein Element enthält. Deren Recall-Wert fällt nichtsdestotrotz um ca. 10%, was daran liegt, dass es für den Parser schwieriger ist, das Mittelfeld „alleine“ zu finden, d.h. ohne die anderen Feldknoten.

Wir testen des Weiteren eine Version von TüBa-D/Z, in der wir **Phrasen abflachen**, *Tü\_flat*. Die Phrasen in *Tü\_flat* sind mit den flachen Phrasen aus NeGra vergleichbar. Mit dieser Baumbankversion werden Ergebnisse erreicht, die mit denen der unmodifizierten Baumbank (*TüBa*) vergleichbar sind. Die F-Measure-Werte sind geringfügig höher und der Parser erzeugt weniger kreuzende Klammern. Eine einzelne Kategorie, EN-ADD, profitiert am meisten von dieser Modifikation: Ihr F-Measure-Wert steigt um ca. 45 Punkte. EN-ADD wurde zur Markierung von *named entities* eingeführt, d.h. das Etikett hat keine syntaktische Funktion. In *Tü\_flat* werden viele der Nominalknoten unterhalb von EN-ADD entfernt, was EN-ADD näher an die lexikalische Ebene bringt. Somit bekommt EN-ADD mehr sinnvollen Kontext und kann vom Parser leichter gefunden werden.

Weiterhin testen wir Kombinationen der eben beschriebenen Modifikationen. Mit Ausnahme der durchschnittlichen Baumhöhe ähneln alle Werte von TüBa-D/Z mit **abgeflachten Phrasen und ohne unäre Produktionen** (*Tü\_f\_NU*) denen der unmodifizierten NeGra-Baumbank. Dies weist auf die Ähnlichkeit beider Baumbanken hin. Nichtsdestotrotz sind die Parsingergebnisse auf NeGra schlechter, woraus geschlossen werden kann, dass TüBa-D/Z noch von den übrigen topologischen Feldknoten profitiert. Die Anzahl der kreuzenden Klammern ist in dieser Baumbankversion am geringsten.

In der letzten Baumbankversion, die wir testen, werden **alle bisher durchgeführten Modifikationen kombiniert** (*TÜ\_f\_NU\_NF*). Hier sinken, wie erwartet, alle Werte dramatisch ab. Der F-Measure-Wert ist in dieser Version nichtsdestotrotz nur 5 Punkte schlechter als der F-Measure-Wert der unmodifizierten NeGra-Baumbank.

## 5 Schlussbetrachtung

Wir haben die beiden Baumbanken NeGra und TüBa-D/Z miteinander verglichen, indem wir die Rolle einzelner Annotationsbestandteile im Zusammenhang mit PCFG-Parsing untersucht haben. Zu diesem Zwecke haben wir die Annotation der Baumbanken an die Annotation der jeweils anderen Baumbank an-



genähert und das Verhalten der veränderten Versionen mit einem PCFG-Parser untersucht. Die vier folgenden Aspekte fassen die Ergebnisse der Arbeit zusammen.

- *Dimensionen der Baumbanken und ihrer modifizierten Versionen*: Wenn wir Struktur aus den Baumbanken entfernen oder hinzufügen, so verändert sich die Höhe der Baumbank, die durchschnittliche Anzahl der Töchter nicht-terminaler Knoten, sowie das Knoten/Token-Verhältnis. Durch Modifikationen der Annotation, die die Baumbanken aneinander annähern, nähern sich auch die drei Dimensionen aneinander an. Nochmals sei auf die Ähnlichkeit der Werte von *TÜ\_NU\_NF* und der NeGra-Basisversion hingewiesen.
- *Precision/Recall und F-Score*: Was hilft dem Parser? Wenn dem Parser beim Training viele tiefe Strukturen zur Verfügung stehen, sind die daraus resultierenden Ergebnisse i.d.R. besser als mit flachen Strukturen. Sichtbar wird dies in den Ergebnissen für TüBa-D/Z (Parsing mit Gold-POS-Tagging): Während die Originalversion zu 85,44 F-Score führt, ist der Wert bei abgeflachten Phrasen, ohne unäre Produktionen und ohne topologische Felderebene, um mehr als 20 Punkte schlechter (64,90). Vor allem die unären Produktionen sind dem Parser eine große Hilfe. Bei NeGra zeigt sich ein ähnliches Bild. Während mit der Basisversion 69,05 F-Score erreicht wird, ergeben sich bei unter PPs eingefügten NPs (*NE\_NP*) 71,60, beim Parsen mit der Felderebene 72,59 F-Score.
- *Kreuzende Klammern*: Kreuzende Klammern verhalten sich umgekehrt zu den Precision/Recall- und F-Score-Werten. Je weniger Struktur dem Parser zum Trainieren übergeben wird, desto weniger kreuzende Kanten entstehen. Auch dies ist sowohl an TüBa-D/Z, als auch an NeGra sichtbar. Bei den TüBa-D/Z-Varianten nimmt die Anzahl der kreuzenden Kanten proportional zur Menge der entfernten Struktur ab, bei NeGra nimmt die Anzahl beim Hinzufügen von Struktur zu. Es muss hier aber gefragt werden, ob die niedrige Zahl kreuzender Klammern in NeGra nur dadurch entsteht, dass die Bäume wenige Konstituenten enthalten.

Einen Ansatz für weitere Forschung im Sinne der vorliegenden Arbeit bietet die Untersuchung der Baumbanken unter Verwendung anderer Parsingmodelle als des verwendeten PCFG-Modells von *lopar*. Verschiedene mit Blick auf das Englische erweiterte PCFG-Modelle könnten untersucht werden. Obwohl Lexikalisierung für das Englische gute Erfolge erzielt (siehe z.B. Collins (1997) und Klein and Manning (2002)), zeigen Dubey and Keller (2003) mit der NeGra-Baumbank, dass dies für das Deutsche nicht gilt. Vergleichende Experimente mit den für das Englische entwickelten lexikalisierten Modellen und dem Schwester-Kopf-Modell von Dubey and Keller (2003) wären von Interesse, aber auch eine Prüfung ihrer Thesen über Lexikalisierung. Hier bietet sich vor allem die von Klein and Manning (2002) verwendete Lexikalisierung an, die Kombination eines unlexikalisierten PCFG-Modells und eines Abhängigkeitsparsingmodells, die bereits in Kübler et al. (2006) erfolgreich verwendet wurde.

## Literaturverzeichnis

- Collins, M. (1997). Three generative, lexicalised models for statistical parsing. In *Proceedings of 35th Annual Meeting of the Association for Computational Linguistics and 8th Conference of the European Chapter of the Association for Computational Linguistics (ACL/EACL 1997)*, pages 16–23.
- Collins, M. (1999). *Head-Driven Statistical Models for Natural Language Parsing*. PhD thesis, University of Pennsylvania.
- Corazza, A., Lavelli, A., Satta, G., and Zanolini, R. (2004). Analyzing an Italian treebank with state-of-the-art statistical parsers. In *Proceedings of the 3<sup>rd</sup> Workshop on Treebanks and Linguistic Theories (TLT 2004)*, pages 39–50.
- Drach, E. (1937). *Grundgedanken der deutschen Satzlehre*. Diesterweg, Frankfurt/Main.
- Dubey, A. and Keller, F. (2003). Probabilistic parsing for German using sister-head dependencies. In *Proceedings of the 41th Annual Meeting of the Association for Computational Linguistics*, pages 96–103.
- Gildea, D. (2001). Corpus variation and parser performance. In *Proceedings of the 2001 Conference on Empirical Methods in Natural Language Processing*, pages 167–202.
- Höhle, T. (1986). Der Begriff „Mittelfeld“, Anmerkungen über die Theorie der topologischen Felder. In *Kontroversen alte und neue. Akten des Siebten Internationalen Germanistenkongresses 1985*, pages 329–340, Göttingen.
- Klein, D. and Manning, C. D. (2002). Fast exact inference with a factored model for natural language parsing. In *Advances in Neural Information Processing Systems 15 (NIPS 2002)*, pages 3–10.
- Klein, D. and Manning, C. D. (2003). Accurate unlexicalized parsing. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, pages 423–430.
- Kübler, S. (2005). How do treebank annotation schemes influence parsing results? Or how not to compare apples and oranges. In *Proceedings of the 5th International Conference on Recent Advances in Natural Language Processing (RANLP 2005)*, pages 293–300.
- Kübler, S., Hinrichs, E. W., and Maier, W. (2006). Is it really that difficult to parse German? In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, pages 111–119.
- Schmid, H. (2000). LoPar: Design and implementation. Technical report, Universität Stuttgart, Germany.
- Skut, W., Krenn, B., Brants, T., and Uszkoreit, H. (1997). An annotation scheme for free word order languages. In *ANLP 1997, 5th Applied Natural Language Processing Conference*, pages 88–95.
- Telljohann, H., Hinrichs, E. W., and Kübler, S. (2003). *Stylebook for the Tübingen Treebank of Written German (TüBa-D/Z)*. Seminar für Sprachwissenschaft, Universität Tübingen.