Einbettung von Grammatikregeln in Ngramm-Sprachmodelle

Studienarbeit am Institut für Logik, Komplexität und Deduktionssysteme
Prof. Dr. Alex Waibel
Fakultät für Informatik
Universität Karlsruhe (TH)

von

Falk Fleischer

Betreuer:
Prof. Dr. Alex Waibel
Dipl.-Inform. Christian Fügen

Tag der Anmeldung: 18. März 2004
Tag der Abgabe: 18. Juni 2004

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe.

Karlsruhe, den 17.06.2004

Falk Fleischer

Zusammenfassung

In dieser Arbeit werden Regeln einer kontext-freien Grammatik in ein Ngramm-Sprachmodell eingebunden. Ziel ist es, die Erkennungsrate des daraus resultierenden kombinierten Sprachmodells über die Leistung der beiden Ausgangsmodelle hinaus zu verbessern. Dabei sollen die Funktionsweisen der vorhandenen Modelle in möglichst großem Umfang genutzt werden. Nach einem Überblick über Ansätze anderer Autoren, werden detailliert die Grundlagen und Vorgehensweisen zur Erstellung des kombinierten Sprachmodells erläutert. Die Verbindung der Historien der beiden beteiligten Arten von Basismodellen erfolgt dabei über eine zustandsbasierte Historie. Im Anschluss werden unterschiedliche Varianten des kombinierten Sprachmodells und erste Resultate vorgestellt. Diese Ergebnisse dienen als Ausgangspunkt für weitere Änderungen am Modell. Dabei wird eine eindeutige Trennung des Vokabulars in Ngramm- und Grammatikvokabular, eine veränderte Offsetberechnung sowie eine Gewichtung der Basismodelle untersucht. Im besten Fall konnte mit dem hier erarbeiteten Ansatz eine Verbesserung der Wortfehlerrate um 7,11% relativ zum besten Ausgangsmodell mit 22,91% Wortfehlerrate erziehlt werden.

Inhaltsverzeichnis

1 Einleitung

Das Ziel der Spracherkennungsforschung ist das automatische Erkennen des wahrscheinlichsten Satzes aus einem Signal natürlicher, gesprochener Sprache. Die verschiedenen dazu verwendeten Ansätze zeigen jeweils spezifische Vor- und Nachteile in ihrer Anwendbarkeit. Die Frage die sich daher stellt ist, ob durch Kombination unterschiedlicher Ansätze ein Mehrwert an Leistung erreicht werden kann.

Die hier vorliegende Arbeit untersucht eben jene Frage auf dem Gebiet der Sprachmodelle. Die Aufgabe von Sprachmodellen im Bereich der automatischen Spracherkennung ist es, gegeben einer Historie erkannter Wörter eine Schätzung der apriori-Wahrscheinlichkeit der möglichen Folgewörter zu geben. Die beiden in dieser Arbeit verwendeten Ansätze sind zum einen ein statistisches Ngramm-Sprachmodell, im folgenden als 'Ngramm-Modell' bezeichnet, und zum anderen eine kontextfreie Grammatik mit gleichwahrscheinlichen Wortübergängen, im weiteren mit 'CFG' abgekürzt.

Der statistische Ansatz des Ngramm-Modell hat sich in der Vergangenheit als sehr erfolgreich bei domänenübergreifenden Erkennungsaufgaben bewährt [Ros00]. Problematisch ist, dass zur robusten Schätzung der Ngramm-Wahrscheinlichkeiten große Trainingsmengen nötig sind und selbst in diesen viele Ngramme oft nicht oder nicht ausreichend häufig vorkommen [Ros00]. Dies macht sich insbesondere bei domänenspezifischen Erkennungsaufgaben bemerkbar. Für spezielle Domänen liegen oft unzureichende Datenmengen für das Training vor. Während sich das Ngramm-Modell bei allgemeinen domänenübergreifenden Aufgaben auszeichnet, fällt dessen Erkennungsleistung aufgrund der nicht ausreichend trainierbaren Wortübergänge in spezifischen Domänen zurück.

Die CFG hingegen lässt sich sehr genau in Hinblick auf spezifische Domänen erstellen. Ihre Stärke liegt in der Modellierung von Sätzen, wie sie in einer speziellen Domäne, z.B. in einem Dialog, häufig vorkommen. Die CFG bietet dabei ein gutes Modell für entfernte Kontextabhängigkeiten zwischen Wörtern[Ben00]. Aufgrund der Historie bereits erkannter Wörter werden hierbei gezielt Wörter als alleinige mögliche Fortsetzung aufgeführt.
Eine CFG wird häufig per Hand erstellt. Diese Tatsache schränkt die Wortabdeckung, die Güte und somit auch die Erkennungsmöglichkeiten entscheidend ein. Durch den starren Aufbau der Satzstruktur ist die Flexibilität der Erkennung zusätzlich beschränkt. Da es unmöglich ist beim Bau einer Grammatik alle möglichen Sätze eines potentiellen Sprechers zu beachten, können die meisten Sätze außerhalb der spezifischen Domäne sowie spontane Äußerungen nicht erkannt werden.

Das Ziel, dieser Arbeit war es, die Kontextinformation der CFG mit der Flexibilität und Generalität des Ngramm-Modells zu verbinden und somit eine Erkennungsleistung zu erreichen, die über denen der einzelnen Ansätze liegt. Als Rahmen wurde der JANUS-Spracherkenner des ILKD Karlsruhe, Prof. Waibel, in Verbindung mit dem Ibis-Decoder [Sol01] genutzt. Im Folgenden werden zunächst einige Ideen für Kombinationsmöglichkeiten aus der Literatur vorgestellt. In den Abschnitten 3 und 4 wird ausführlich der hier genutzte Ansatz und seine Umsetzung im Ibis-Decoder erklärt. Die daraus folgenden experimentellen Resultate werden in Abschnitt 5 vorgestellt, bevor in Abschnitt 6 darauf aufbauend einige Veränderungen am ersten Ansatz und deren Auswirkungen erläutert werden.

2 Ansätze aus der Literatur

Die Kombination statistischer und linguistischer Modelle ist in den vergangenen Jahren in verschiedener Form untersucht worden. In diesem Abschnitt sollen einige ausgewählte Ansätze kurz vorgestellt werden.

2.1 Verbindung linguistischer Analyse mit statistischem Modell

Der in [Moo95] vorgestellte Ansatz basiert auf einer Neubewertung der N-best Ausgabe des DECIPHER ATIS Spracherkennungssystems. Die Autoren von [Moo95] nutzen dazu ein statistisches Sprachmodell mit 3 Ebenen:

1. Trigramm-Wahrscheinlichkeiten einer Folge von Fragmenten einer Äußerung
   Ein Beispiel für eine Folge von Fragmenten ist [begin-utterance][filler][sentence][end-utterance]. Die Übergänge zwischen den einzelnen Fragmenten werden durch ein Trigramm-Modell berechnet.

2. 4-Gramm-Modell von Wörtern und Wortklassen innerhalb jedes Fragmentes
   Die Wortübergangswahrscheinlichkeiten basieren auf einem 4-Gramm-Modell aller Übergänge von Wörtern und Wortklassen. Innerhalb der Wortklassen sind Substantive mit den gleichen syntaktischen und semantischen Merkmalen zusammengefasst. Zusätzlich wurden für jedes Fragment spezielle Anfangssymbole genutzt, um die Wahrscheinlichkeit der ersten Worte einer Aussprachesequenz an ihre Auftretenshäufigkeit in den einzelnen Fragmenten anzupassen.

3. Wortklassen
   Die Wahrscheinlichkeitsverteilung der Wörter innerhalb der einzelnen Klassen ist abhängig vom Nutzen ihrer Klasse in unterschiedlichen Domänen entweder auf dem Trainingskorpus geschätzt oder gleichverteilt.

Zunächst wird jede der N-best Hypothesen des Spracherkenners durch Gemini, ein Verarbeitungssystem für natürliche Sprache [Dow93], analysiert. Die Ausgabe der Analyse ist eine Folge semantisch bedeutungsvoller Satzfragmente, wie im Beispiel von 1. gezeigt. Folgende Fragmente sind dabei laut [Moo95] möglich: "sentence, nominal phrase, modifier phrase, filler, skipped". Gemini bietet hierbei neben Regeln zur syntaktischen sowie semantischen Analyse englischsprachiger Hypothesen auch Strategien zur Fehlerkorrektur.
Die Wahrscheinlichkeit der N-best Hypothesen des initialen Spracherkenners werden anschließend als Folge von Fragmenten durch das statistische Modell laut 1. bis 3. neu bewertet.

Nach [Moo95] wurde mit diesem Ansatz eine Verbesserung der Erkennungsrate um bis zu 15,3% relativ zum Basissystem mit statistischem Sprachmodell und ohne linguistische Analyse erreicht.

Dabei ist anzumerken, dass dieser Ansatz allein durch den Einsatz der Fehlererkennungsfunktion von Gemini während der Analyse der Hypothesen einen wichtigen Vorteil gegenüber anderen Ansätzen besitzt. Laut [Moo95] ergab die Ersetzung des originalen Sprachmodells durch den Ansatz in 1. bis 3. eine Verschlechterung der Erkennungsleistung. Erst die beschriebene dreistufige Vorgehensweise aus Erstellung der N-best Hypothesen, nachfolgender Analyse und anschließender Neubewertung ergab eine Verbesserung.

2.2 Einsatz von Grammatiken innerhalb eines Ngramm-Modells

In [Gil98] wurde eine Kombination linguistischer und statistischer Modelle vorgeschlagen, die im Vergleich zu [Moo95] ohne ein linguistisches Analysewerkzeug arbeitet. Die Idee dabei ist, auf Grundlage eines gegebenen Textkorpus, Klassen von kurzen, spezifischen Ausdrücken zu erstellen und diese in ein Ngramm-Modell über dem gleichen Textkorpus einzubinden. Folgendes aus [Gil98] entnommenes Beispiel für die Klasse [Date] soll dies verdeutlichen:

[Date] → {month} {day}
{month} → 'january'
...
→ 'december'
{day} → {digit}
→ {digit} {digit}
{digit} → '0'
...
→ '9'

Jede dieser Klassen wird dabei durch eine stochastische kontextfreie Grammatik (SCFG) modelliert. Zusätzlich bildet jedes Wort des Vokabulars eine eigene primitive Klasse, z.B.:

[january] → 'january'

Über diesen Klassen wird nun das Ngramm-Modell trainiert.
Die Decodierung einer Äusserung zu einer Folge von Klassen ist auf mehrere Arten möglich. Die Wahrscheinlichkeit einer Äusserung u der Länge m ist, wie in [Gil98] dargestellt, die Summe aus allen möglichen Decodierpfaden p:

$$P(u) = \sum_{p} P(u,p)$$

Die Klassenübergangswahrscheinlichkeiten $P(C_i|C_{i-1}, C_{i-2})$ werden durch ein Trigramm-Modell geschätzt, die Wahrscheinlichkeit $P(\bar{w}_{C_i}|C_i)$ einer Wortfolge $\bar{w}_{C_i} = [w_{C_i1} \dots w_{C_ik}]$ innerhalb der Klasse C_i mittels SCFG.
In [Wan00] wird ein dem Ansatz in [Gil98] folgender Vorschlag für die Berechnung von $P(u,p)$ genutzt:

$$P(u,p) = \prod_{i=1}^{m} P(C_i|C_{i-1}, C_{i-2}) \prod_{i=1}^{m} P(\bar{w}_{C_i}|C_i)$$

Für die Berechnung der Wahrscheinlichkeit $P(\bar{w}_{C_i}|C_i)$ wurden in [Gil98] und [Wan00] verschiedene Vorschläge gemacht. In [Gil98] wird $P(\bar{w}_{C_i}|C_i$ über die relative Häufigkeit geschätzt, mit der die Klasse C_i im Textkorpus genutzt wird, um die Wortfolge $\bar{w}_{C_i}$ zu erzeugen. In [Wan00] werden folgende Ansätze benutzt:

1. Gleichverteilung über die Wörter der Klasse, die der aktuellen Historie folgen können
2. gleiche Trigramm-Wahrscheinlichkeiten aus Ngramm-Modell für alle Klassen übernommen
3. spezielle Trigramm-Wahrscheinlichkeiten für die einzelnen Klassen

Im Fall 2 und 3 wird zusätzlich über die Anzahl der möglichen Folgewörter einer aktuellen Historie normalisiert.

Die Experimente in [Gil98] ergaben eine Reduktion der Perplexität relativ zu einem einfachen

Trigramm-Modell um bis zu 19% in Abhängigkeit von der Erkennungsaufgabe. Die Wortfehlerrate konnte dabei nicht signifikant verbessert werden. Das Ergebnis wurde zum Teil sogar schlechter.
In [Wan00] wurde der Ansatz durch Interpolation des kombinierten Modells mit domänenspezifischen Trigramm-Modellen verbessert, um somit zu geringer Domänenspezifität und unzureichender Trainingsmenge entgegenzuwirken. Die Autoren berichten von einer signifikanten Perplexitätsreduktion im Vergleich zu Basis-Trigrammen um bis zu 67%. Angaben über eine Verbesserung der Erkennungsleistung wurden nicht gemacht.

2.3 Interpolation von statistischer und linguistischer Information

Ähnlich zu [Moo95] wurde von den Autoren in [Ben00] ein Modell vorgeschlagen, dass explizit syntaktische Strukturen in ein Sprachmodell einbindet. Die Idee dahinter ist, zwei Modelle mit zum einen "starken lokalen Relationen" und zum anderen "globalen Zusammenhängen zwischen syntaktischen Strukturen"[Ben00] in einem Modell zu vereinen. In [Ben00] wird dafür eine lineare Kombination eines Ngramm-Modells über den Wörtern der Trainingsmenge und eines stochastischen Grammatikmodells über dem selben Trainingskorpus genutzt.

Das Grammatikmodell besteht dabei aus einer kategorien-basierten SCFG und einem probabilistischen Modell der Wortverteilungen innerhalb der einzelnen Kategorien. Die Kategorien der SCFG entsprechen Part-Of-Speech Tags (POStag), also Konzepten wie Substantiv oder Verb. Ein Pfad in der SCFG ist also z.B.: [Substantiv][Verb][Substantiv]. Zum Trainieren des Grammatikmodells wird jedem Wort pro Trainingssatz sein POStag zugewiesen. Diese Trainingsmenge wird daraufhin in zweierlei Hinsicht verwendet. Zum einen werden die POStags zur Schätzung der Übergangswahrscheinlichkeiten zwischen den Kategorien der SCFG genutzt. Zum anderen wird die Wortverteilung innerhalb der einzelnen Kategorien berechnet. Die Wahrscheinlichkeit eines Wortes innerhalb einer Kategorie ist dabei die relative Häufigkeit mit der es der Kategorie zugeordnet wurde im Vergleich zu anderen Wörtern der Kategorie. Ein Wort kann dabei in mehreren Kategorien vorkommen.

Im Vergleich verschiedener Schätzalgorithmen für die SCFG wurde in [Ben00] im besten Fall eine Verringerung der Perplexität um 30.9% relativ zu einem entsprechenden Trigramm-Modell erziehlt.

3 Der eigene Ansatz - Grundlegende Herangehensweise

Die grundlegende Idee dieser Arbeit ist es, die Vorteile der kontextfreien Grammatik, insbesondere die Domänenspezifität und die syntaktischen Abhängigkeiten zwischen Wörtern, mit der Flexibilität eines Ngramm-Modells zu kombinieren. Die Kombination ist dabei ähnlich zu den Modellen in [Gil98] und [Wan00] umgesetzt. Im vorliegenden Ansatz soll versucht werden, einen Großteil der spezifischen Funktionalität bereits gegebener Modelle zu übernehmen und den Anpassungsaufwand der Modelle gering zu halten. Das kombinierte Sprachmodell soll entsprechend auf dem selben Textkorpus wie die zur Kombination verwendeten Basismodelle aufgebaut werden. Das Ziel dieser Arbeit ist es, ein Sprachmodell zu erstellen, welches eine geringere Wortfehlerrate als die bisher genutzten Ausgangsmodelle auf einer gegebenen Testmenge erreicht.

3.1 Die verwendeten Basismodelle

Zur Kombination wurden zwei Basismodelle genutzt:

1. Ngramm-Sprachmodell
   Das verwendete Modell ist auf domänenspezifischen Daten aufgebaut. Zur Erhöhung der Generalität kommen neben einfachen Wörtern auch Klassen von Wörtern vor. Diese Modell wird im weiteren Verlauf als Ngramm-Modell bezeichnet.

2. Kontext-freie Grammatik
   Das Grammatik-Basismodell bietet im Gegensatz zu den Ansätzen in Abschnitt 2 feste Übergänge ohne stochastische Verteilung über den Wörtern. Die hier verwendete Grammatik ist spezifisch für die gleiche Domäne wie das Ngramm-Modell. Theoretisch ist eine beliebige kontext-freie Grammatik einsetzbar. Im Folgenden wird auf die Grammatik mit CFG verwiesen.

3.2 Schnittstellen zum Ibis-Decoder

In dieser Arbeit werden die folgenden Schnittstellen zwischen dem Ibis-Decoder und dem kombinierten Sprachmodell zu Verfügung gestellt:

1. Erstellen einer initialen Historie
   Zu Beginn der Decodierung jeder Äusserung ist eine initiale Historie bestehend aus dem speziellen "begin-of-sentence" Wort (hier <s>) nötig. Eine entsprechende Historie wird hier erstellt.

2. Erweiterung einer gegebenen Historie mit einem Wort
   Hiermit können die im Laufe des Decodiervorgangs benötigten Erweiterungen der Historie erstellt werden. Dazu wird eine Historie und das zu erweiternde Wort vom Ibis-Decoder übergeben und die erweiterte Historie zurückliefert.

3. Bewertung aller Worte gegeben einer Historie
   Die eigentliche Aufgabe eines Sprachmodells, die Berechnung der Wortübergangswahrscheinlichkeiten, wird hiermit durchgeführt. Diese Berechnung geschieht auf Basis einer vom Decoder übernommenen Historie. Es wird die Wahrscheinlichkeit aller Worte gegeben dieser Historie zurückgegeben.

4. Bewertung eines einzelnen Wortes gegeben einer Historie
   Die zum Beispiel für die Neubewertung bereits dekodierter Äusserungen nützliche Bewertung eines einzelnen Wortes ist explizit möglich. Die Berechnung folgt dabei den Vorgängen in iii).

5. Vereinigung des Vokabulars der Basismodelle
   Das kombinierte Sprachmodell arbeitet über den Sprachmodellvokabularien der Basismodelle. Diese müssen dabei nicht identisch sein. Es ist deshalb nötig, ein gemeinsames Vokabular zu erstellen. Zusätzlich werden hierbei weitere Steuerungsinformationen zur Nutzung in i) bis iv) erstellt. Im Unterschied zu den vorangegangenen Funktionen wird dies vorbereitend vor dem eigentlichen Decodiervorgang ausgeführt.

Die Funktionsweise von 1. bis 3. wird bei der Beschreibung des Decodiervorgangs in Abschnitt 4 detailliert ausgeführt. Die Schnittstelle in 5. wurde speziell für das in dieser Arbeit vorgestellte Sprachmodell konzipiert.

3.3 Der Kombinationsvorgang - Die grundlegende Idee

Im folgenden soll gezeigt werden, wie die Vereinigung der beiden Sprachmodelle zu einem einzigen Modell erfolgt. Der Kombinationsvorgang folgt im Ansatz der Darstellung in Abschnitt 2.2. Betrachten wir als Beispiel zunächst folgende für den Textkorpus typische Aussagen:

Show me the way to the cinema.
What is the way to the market place?

Die Idee ist es nun, die in Sätzen dieser Art vorkommenden Ähnlichkeiten in der Satzstruktur zu erkennen und als Regeln einer CFG zu modellieren. Im obigen Beispiel ist eine solche Struktur der Verweis auf Lokalitäten. Eine mögliche Regel der Grammatik in JSGF-Format [Sun98] könnte folgendermassen lauten:

public <to_place> = *to the cinema* |
to the market place ;

Die aus der Regel ableitbaren Wortgruppen werden im Textkorpus entsprechend ersetzt. Dabei dienen ausschließlich die obersten (public) Nichtterminale als mögliche Einstiegspunkte in die Grammatik:

Show me the way [to_place].
What is the way [to_place]?

In den Textkorpus können auf diese Art verschiedene Regeln einer CFG eingebaut werden. Über dem so veränderten Textkorpus wird ein neues Trigramm-Modell erstellt.

3.4 Anpassung der Basismodelle

Die Kombination zweier so unterschiedlicher Modelle wie das Ngramm-Modell und die CFG erfordert gezielte Veränderungen im Aufbau beider. In dieser Arbeit soll keine eigenständige Grammatik erstellt werden, sondern eine bereits vorhandene Grammatik angepasst und genutzt

werden. Eine grundlegende Frage dabei ist, welche Regel der bestehenden Grammatik in welchem Umfang genutzt werden kann. Das hier genutzte Ausgangsmodell der Grammatik besteht aus öffentlichen (public) Regeln, mit denen stets vollständige Sätze erkannt werden können. Folgendes typisches Beispiel soll dies illustrieren:

public <request_way_description> = <show/tell> *the way* <to_place> [*please*] ;

<show/tell> = *show* [*me*] |
tell me ;

<to_place> = *to the cinema* |
to the market place ;

Somit werden bei der Ersetzung von Worten des Textkorpus entsprechend ganze Sätze bzw. Nebensätze durch ein öffentliches Nichtterminal der Grammatik repräsentiert. Durch gezielte Auswahl einzelner Regeln der CFG können Unterformen der Ausgangsgrammatik erstellt werden, die nur kleinere Wortfolgen abdecken. Eine solche Regel für kürzere Worteinheiten ist <to_place>, deren Einbettung beispielhaft in Abschnitt 3.3 dargestellt wurde.

Ein weiterer zu beachtender Punkt ist der Umgang mit Klassen von Wörten, z.B. Straßennamen, wie sie standardmäßig im Ngramm-Modell vorkommen. Durch eine 1:1 Abbildung der Klasseninhalte des Ngramm-Modells als Regeln innerhalb der Grammatik wird im kombinierten Sprachmodell die Verwendung der Klassen vereinheitlicht.
Im folgenden wird die typische Vorgehensweise bei der Erstellung angepasster Basismodelle aufgezeigt:

1. CFG anpassen
   Zuerst werden die zu verwendenden Regeln der CFG ausgewählt. Zusätzlich werden die Klassen des Ngramm-Modells in die CFG als Regeln integriert und angepasst.
2. Einbettung der Grammatikregeln in den Textkorpus
   Mittels eines auf den Regeln, der laut 1. erstellten Grammatik, basierenden Parsers werden alle Sätze des Textkorpus geparst und vollständig parsebare Wortfolgen bzw. Sätze durch die entsprechenden öffentlichen Nichtterminale der Grammatik ersetzt. Ein Beispiel für die Ersetzung wurde in Abschnitt 3.3 vorgestellt.
3. Ngramm-Modell trainieren
   Auf dem veränderten Textkorpus wird nun das Ngramm-Modell trainiert.

4 Das Anfangsmodell

In diesem Abschnitt wird in ausführlicher Form die grundlegende Funktionsweise des Sprachmodells während des Decodiervorgangs dargestellt.

4.1 Das Zustandsmodell

Grundlage für die Berechnung der apriori-Wahrscheinlichkeiten des Sprachmodells ist die vom Ibis-Decoder übergebene aktuelle Historie. Durch die Einbettung der Grammatikregeln in das Ngramm-Modell stellt sich nun die Frage, wie sich eine Historie zusammensetzt.
Auf der einen Seite bilden die Wortübergänge des Ngramm-Modells eine Historie und auf der anderen wird bei der Verzweigung in eine Regel innerhalb der Grammatik eine eigene Historie genutzt. Die Wahrscheinlichkeit $P(T)$ einer Wortfolge $T = [t_1 \dots t_m]$ ist dabei laut Ngramm-Modell:

$$P(T) = \prod_{i=1}^{m} P(t_i | t_{i-n+1} \dots t_{i-1})$$

Innerhalb der CFG gilt für die Wahrscheinlichkeit $P(W)$ der Wortfolge $W = [w_1 \dots w_k]$:

$$P(W) = \prod_{j=1}^{k} P(w_j | w_1 \dots w_{j-1})$$

Diese beiden Wege der Berechnung werden einzeln betrachtet und zu einem gemeinsamen Pfad zusammengefasst. Dabei gilt, dass die Übergangswahrscheinlichkeiten grundsätzlich durch das Ngramm-Modell berechnet werden. Verzweigt die Berechnung aber in ein Nichtterminal der Grammatik, werden die folgenden Wortübergangswahrscheinlichkeiten durch die Grammatik berechnet bis diese an einem Ende einer Regel angelangt ist. Nachdem daraufhin die Ngramm-Historie durch das verwendete Nichtterminal erweitert wurde, wird die Berechnung durch das Ngramm-Modell fortgeführt. Hierzu sei folgender Beispielsatz gegeben:

Show me the way [to_place] *please.*

Bei der Decodierung des Beipielsatzes wird mit einer Trigramm-Wahrscheinlichkeit $P([to\_place]|the\ way)$ in die Grammatikregel <to_place> verzweigt. Innerhalb der Grammatik werden die Wahrscheinlichkeiten möglicher Wortfolgen, z.B. der Folge *to the cinema* - vgl. Abschnitt 3.3, durch die Grammatik selbst berechnet und an den Decoder zurückgeliefert. Am Ende der Grammatikregel wird die bis auf weiteres nicht veränderte Historie des Ngramm-Modells mit dem aktuellen Nichtterminal [to_place] erweitert. Die Bewertung des nächsten Wortes geschieht dann wieder auf Basis der Trigramm-Wahrscheinlichkeit, also $P(please|way\ [to\_place])$.
Auf die Einzelheiten des Berechnungsvorgangs wird in Abschnitt 4.2.3 eingegangen werden. Zunächst aber soll eine Konsequenz aus obigen Vorgehen erläutert werden.

Da die Art der Berechnung abhängig davon ist, ob ein Wort durch die Grammatik berechnet werden soll oder nicht, werden drei verschiedene Zustände der Historie unterschieden:

1. Grammatik inaktiv
   Die Historie befindet sich in diesem Zustand, falls sie im vorangegangenen Schritt mit

einem Wort erweitert wurde, dass nur im Ngramm-Vokabular vorkommt, oder falls die Grammatik am Ende einer Regel angelangt war.

2. Grammatik aktiv
   Dieser Zustand liegt entsprechend dann vor, wenn im vorhergehendene Schritt in die Grammatik verzweigt wurde bzw. die Grammatik nach der Erweiterung nicht am Ende einer Regel angelangt war.

3. Grammatik optional aktiv
   Dieser Sonderfall wird aufgrund der Eigenschaft der Grammatik, optionale Fortsetzungen zu erlauben, eingeführt. Betrachten wir dazu folgendes Beispiel von oben in leichter Abwandlung:

   Show me the way [to_place].

   mit *please* als optionalem Grammatikwort:

   public <to_place> = *to the cinema* [*please*] |
   to the market place ;

   Ist die Regelanwendung in diesem Fall bei *cinema* angelangt, besteht für sie die Option das Wort *please* zu erweitern oder aber am Ende der Regel angelangt zu sein. In diesem Zustand des optionalen Regelendes soll daher die mögliche Erweiterung durch das Ngramm-Modell ebenfalls in Betracht gezogen werden. Während diese Möglichkeit ursprünglich nur am Ende der CFG vorgesehen war, wurde bei der Implementierung die Ngramm-Erweiterung an allen Stellen der Grammatikregeln eingeführt, an denen diese eine optionale Fortsetzung erlauben. Dadurch ist ein größeres Maß an Flexibilität beim Austritt aus der Grammatik möglich, welches sich positiv auf die Testresultate auswirkte.

4.2 Anpassung der Schnittstellen

4.2.1 Erstellen einer initialen Historie

Zu Beginn der Decodierung jeder Äusserung wird eine initiale Historie erstellt, die nur aus dem Startwort <s> besteht. Die Anfrage des Ibis-Decoders wird an die beiden Basismodelle weitergeleitet. Beide daraus resultierenden initialen Historien werden parallel gespeichert und als eine gemeinsame Historie dem Decoder übermittelt. Der Decoder ist dabei für die Verwaltung aller erstellten Historien verantwortlich. Die gemeinsame initiale Historie ist immer im CFG-inaktivem Zustand.

4.2.2 Erweitern der Historie mit einem Wort

Auf Anfrage des Ibis-Decoders wird die Historie mit einem Wort erweitert. Dafür übergibt der Decoder die zu erweiternde Historie sowie das entsprechende Wort. Von dem aktuellen Zustand der übernommenen Historie und von den Eigenschaften des Wortes innerhalb des Sprachmodells ist es nun abhängig, ob für die neue Historie die zugrundeliegende Ngramm-Historie oder die CFG-Historie erweitert wird. Die unterschiedlichen Fälle abhängig vom Zustand der aktuellen

Historie seien im folgenden betrachtet:

1. Zustand CFG-inaktiv
   Kommt das übergebene Wort nur im Ngramm-Modell vor, dann wird auch nur die Ngramm-Historie erweitert. Ist das Wort aber aus dem gemeinsamen Vokabular von Ngramm-Modell und CFG, dann muss überprüft werden, ob das Wort zu Beginn mindestens einer Regel der CFG vorkommt. Wenn dies der Fall ist, wird in die Grammatik verzweigt, indem die sich im initialen Zustand befindliche CFG-Historie mit dem Wort erweitert wird. Die Ngramm-Historie bleibt dabei unverändert.

2. Zustand CFG-aktiv
   In diesem Fall kann ausschließlich die Historie der CFG erweitert werden. Dass das Wort, welches durch den Decoder übergeben wurde, eines ist, welches nur im Ngramm-Modell vorkommt, sollte dabei ausgeschlossen sein. Wie in Abschnitt 4.2.3 gezeigt werden wird, wird solchen Wörtern im Zustand CFG-aktiv immer eine maximal schlechte Bewertung gegeben.

3. Zustand CFG-optional-aktiv
   Die Erweiterung hier ist wieder abhängig von der Zugehörigkeit des Wortes zu den einzelnen Basisvokabularien. Ein Wort, welches nur im Ngramm-Modell vorkommt, wird auch die Ngramm-Historie erweitern. Dazu ist es nötig, diese vorher mit der zuletzt besten aktiven Regel der CFG zu erweitern. Danach kann die CFG-Historie auf den Initialzustand gesetzt werden.
   Kommt das Wort aber im Vokabular der Grammatik vor, dann wird ausschließlich die CFG-Historie erweitert.

Bei der Erweiterung sind im weiteren einige Besonderheiten zu beachten:
Wichtigster Punkt dabei ist es, den Zustand der erweiterten Historie zu aktualisieren. Nach jeder Erweiterung einer vom Ibis-Decoder übergebenen Historie mit einem Wort ist es möglich, dass die daraus resultierende neue Historie in einem anderen Zustand ist. Ein einfaches Beispiel dafür ist die Anfrage nach Erweiterung einer Historie im CFG-aktiven Zustand mit einem Wort, welches nur im Ngramm-Vokabular vorkommt. Da dies einem Austritt aus der CFG entspricht, muss der Zustand der erweiterten Historie zu CFG-inaktiv aktualisiert werden.

Ist die Grammatik an einem Ende angelangt, ist es nötig die Ngramm-Historie mit dem Anfangsnichtterminal der benutzen Regel zu erweitern. Dabei ist beim Einstieg in die Grammatik durchaus nicht eindeutig, welche Regel das gewesen ist. Alle Regeln sind in einer einzigen CFG enthalten und nutzen somit auch das gleiche Vokabular. Nun werden zwei gleiche Einstiegswörter in unterschiedliche Regeln durch das Ngramm-Modell aber unterschiedlich bewertet. In einem solchen Fall wird in die wahrscheinlichere Regel erweitert. Im Laufe der Erweiterung der CFG-Historie kann es in der CFG aber dazu kommen, dass in eine andere Regel übergegangen wird. Dies wird von der CFG nach außen aber nicht kenntlich gemacht, sondern intern behandelt. Bei der Einbettung der Regeln in das Ngramm-Modell hat diese Änderung aber durchaus globale Folgen! Hierzu ein Beispiel zweier CFG-Regeln:

public <way> = *show me the way to my hotel* ;
public <all> = *show me all bars of the city* ;

Die Bewertung der Verzweigung in beide Regeln wird mittels des Ngramm-Modells vorgenommen. Dabei gilt im Allgemeinen $P([way]|XY) \neq P([all]|XY)$. Nehmen wir an das $P([way]|XY) > P([all]|XY)$. In Vorwegnahme der Bewertungsfunktion aus Abschnitt 4.2.3 wird die Produktion von *show* durch die Regel <way> als wahrscheinlicher bewertet. Führen wir das Beispiel nun fort, indem die CFG-Historie mit *show* und danach mit *me* erweitert wird. Folgt darauf eine Anfrage des Decoders nach Erweiterung der Historie mit *all*, dann nimmt die CFG intern <all> als aktive Regel an. Diese Änderung der aktiven Regel bedeutet aber, dass beim Einstieg in die CFG *show* eine falsche Bewertung zugewiesen bekommen hat. Der Unterschied in der Bewertung muss deshalb ausgeglichen werden. Dies geschieht dadurch, dass beim Austritt aus der CFG überprüft wird, durch welche Regel die CFG-Historie erstellt wurde und die mögliche Differenz zur Einstiegsregel

$$Offset = \frac{P([all]|XY)}{P([way]|XY)}$$

den nächsten zu bewertenden Wörtern aufgeschlagen wird.

4.2.3 Die Bewertungsfunktion

Die entscheidenden Berechnungen der apriori-Wahrscheinlichkeiten von Wörtern wird durch die Bewertungsfunktion gewährleistet. Dazu wird vom Ibis-Decoder eine Historie übergeben, auf deren Basis die Bewertung aller Wörter des kombinierten Sprachmodellvokabulars stattfindet. Grundsätzlich gilt für die kombinierte Wahrscheinlichkeit $P(W)$ der Wahrscheinlichkeiten beider Basismodelle nach Abschnitt 4.1:

$$P(W) = \prod_{i=1}^{m}((Offset * P(t_i|t_{i-n+1} \dots t_{i-1})) * P(\bar{w}_{t_i}|t_i))$$

wobei t_i einem Wort oder einem Nichtterminal aus dem Ngramm-Modell entspricht und $P(\bar{w}_{t_i}|t_i)$ die Wahrscheinlichkeit ist, die Wortfolge $\bar{w}_{t_i} = [w_{t_i1} \dots w_{t_ik}]$ aus der Regel t_i der CFG zu erzeugen. (Vgl. auch [Wan00])
Die Einzelheiten der Berechnung sind dabei wieder vom aktuellen Zustand der Historie abhängig:

1. Zustand CFG-inaktiv
   Ist die vom Decoder übergebene Historie im Zustand CFG-inaktiv, dann werden alle Wörter des Sprachmodellvokabulars durch das Ngramm-Modell bewertet, natürlich auch die Nichtterminale. Da aber das Suchvokabular ausschließlich über Wörtern arbeitet, wird die Bewertung der Nichtterminale auf die möglichen Anfangswörter der entsprechenden Grammatikregeln übertragen. Wie das Beispiel des vorangegangenen Abschnittes zeigt, können dabei gleiche Anfangswörter bei unterschiedlichen Regeln auftreten. In diesem Fall wird einem Anfangswort die Bewertung der Regel mit der höchsten Wahrscheinlichkeit, also der besten Bewertung, zugewiesen.
   Bei der Berechnung muss außerdem beachtet werden, dass bei der vorangegangenen Erweiterung der Historie eventuell die CFG verlassen wurde und somit möglicherweise ein Offset nach Abschnitt 4.2.2 aufgeschlagen werden muss. Für die Berechnung der Übergangswahrscheinlichkeit eines Wortes w_i in diesem Zustand gilt also:

   $$P(w_i) = Offset * P(w_i|w_{i-n+1} \dots w_{i-1})$$

   mit

   $$P(w_i) = P(r_i)$$

wenn $w_i = w_{r_i 1}$ Anfangswort der Regel r_i und r_i wahrscheinlichste Regel mit Anfangswort w_i ist.
Wörtern, die nur im Vokabular der CFG vorkommen, wird eine maximal schlechte Bewertung zugewiesen.

2. Zustand CFG-aktiv
Befindet sich die aktuelle Historie in diesem Zustand, werden alle Wörter des Grammatikvokabulars durch die CFG bewertet. In diesem Fall wird den Wörtern, die nur im Ngramm-Vokabular vorkommen, eine maximal schlechte Bewertung zugewiesen. Es gilt also:
$$P(w_{t_i j}|t_i) = P(w_{t_i j}|w_{t_i 1} \ldots w_{t_i j-1})$$

3. Zustand CFG-optional-aktiv
In diesem Fall werden alle Wörter des CFG-Vokabulars durch die CFG entsprechend 2. bewertet. Um die Wörter zu bewerten, die nur aus dem Ngramm-Vokabular stammen, muss zuerst die bisher nicht erfolgte Erweiterung der Ngramm-Historie mit dem besten aktiven Nichtterminal durchgeführt werden. Da zum Zeitpunkt der Berechnung noch nicht feststeht welche Erweiterung der Historie der Ibis-Decoder als nächstes anfragen wird, kann noch keine feste Entscheidung getroffen werden mit welchem Nichtterminal die Ngramm-Historie erweitert werden soll, vgl. dazu auch Abschnitt 4.2.2. Aus diesem Grund wird die in diesem Zustand nötige Bewertung der Übergangswahrscheinlichkeit der Worte die nur im Ngramm-Vokabular vorkommen mit Hilfe einer Kopie der aktuellen Ngramm-Historie durchgeführt. Diese Kopie wird dabei mit dem aktuellen besten Nichtterminal erweitert und auf dieser Basis die Wortübergangswahrscheinlichkeiten in die Ngramm-Wörter nach 1. berechnet.

5 Experimentelle Varianten und erste Ergebnisse

In diesem Abschnitt sollen die Ergebnisse dargestellt werden, die mit dem bisherigen Modell erreicht wurden. Anfangs wird eine Übersicht über die Eigenschaften getesteter Varianten des Sprachmodells gegeben. Danach werden die erzielten Resulte aufgezeigt und anschliessend kritisch bewertet. Zu Beginn sollen die Eigenschaften der Ausgangsmodelle als Vergleichsgrundlage beschrieben werden.

Alle folgenden Modelle basieren auf dem gleichen domänenspezifischen Trainingskorpus. Dieser besteht aus 32317 Äußerungen und 293231 Wörtern bzw. Klassen der LingWear-Domäne. Diese Domäne ist spezifisch für den Bereich Touristik, insbesondere Navigation, Information und Hotelreservierung [Fue01].
In den Klassen sind dabei spezielle Wörter, wie z.B. Straßen- oder Platznamen, enthalten.

Getestet wurden die einzelnen Modelle auf den Daten von 9 unterschiedlichen Sprechern mit insgesamt 362 Äußerungen. Diese Testdaten sind ebenfalls aus der LingWear-Domäne.

5.1 Die Ausgangsmodelle

Die hier beschriebenen Ausgangsmodelle folgen den Ausführungen in Abschnitt 3.1. Sie dienen als Vergleich zur Einschätzung der Leistung des kombinierten Modells.

1. Trigramm-Sprachmodell
   Zur Schätzung der Trigramm-Wahrscheinlichkeiten wurde der gesamte Textkorpus genutzt. Die Einzelheiten des Modells sind in Tabelle 1 dargestellt.

Perplexität	4,411
Vokabulargröße	2015
Wortfehlerrate	22,91%

Tabelle 1: Daten des Trigramm-Ausgangsmodells

2. CFG-Sprachmodell
   Die hier verwendete Ausgangs-CFG ist speziell auf den Einsatz in der verwendeten Navigationssystem-Domäne abgestimmt. Die in der Ausgangsgrammatik enthaltenen Regeln decken vollständige Sätze ab. Die für diese Arbeit wichtigsten Eigenschaften der Grammatik sind in Tabelle 2 aufgeführt.

Anzahl Einstiegsregeln	24
Vokabulargröße	2015
Wortfehlerrate	26,23%

Tabelle 2: Daten des CFG-Ausgangsmodells

5.2 Varianten der Kombination

Die verschiedenen Varianten basieren alle auf unterschiedlichen Änderungen der Ausgangsgrammatik. Die allgemeine Vorgehensweise bei der Erstellung dieser Varianten wurde in Abschnitt 3.4 erklärt. Die Vokabularien der der Kombination jeweils zugrundeliegenden Modelle werden, wie in Abschnitt 3.2 angedeutet, vor dem eigentlichen Erkennungsdurchlauf vereinigt.

5.2.1 Variante A - Vollständige Übernahme der Ausgangs-CFG

In dieser Variante wurde die Ausgangs-CFG nicht verändert. Die einzige Anpassung an der Grammatik bildete nur das Einfügen aller Klassen aus dem Trigramm-Modell. Aus jeder Klasse resultiert dabei eine neue Einstiegsregel in die Grammatik. Vgl. hierzu auch Abschnitt 3.4. Der Vorteil dieser Vorgehensweise ist zum einen der relativ geringe Aufwand der Anpassung. Dadurch wird man der Forderung nach möglichst umfangreicher Verwendung der Ausgangsmodelle gerecht und nutzt die schon vorliegende Modellierungsleistung. Zum anderen bleibt somit die der CFG zugrunde liegende Satzstruktur erhalten. Eben diese Struktur stellt einen der entscheidenden Vorteile der Modellierung mit kontext-freien Grammatiken im Vergleich zu Ngramm-Modellen dar. Einen Überblick über die resultierenden Basismodelle gibt Tabelle 3

Perplexität Trigramm	4,749
Anzahl Einstiegsregeln der CFG	36
Anzahl Wörter im Gesamtvokabular	2015

Tabelle 3: Daten der Basismodelle von Variante A

Es wurden insgesamt 34727 Wörter des Textkorpus durch Nichtterminale der Grammatik ersetzt.

5.2.2 Variante B - Zerlegung der CFG-Regeln in kürzere Regeln

In zweiten Fall wurden einschneidende Veränderungen am Aufbau der Grammatik vorgenommen. Aus den Regeln der Ausgangs-CFG wurden dabei mehrere kurze Regeln ausgewählt. Dieses Vorgehen wird anhand des Beispiels in Abschnitt 3.4 verdeutlicht. Dabei zeigt sich auch, dass es zu einer geringeren Wortabdeckung der resultierenden Grammatik kommen kann.
Die Idee hinter dieser Vorgehensweise ist es, den Einsatz der Grammatik flexibler zu gestalten. Die Übergänge zwischen den kürzeren Satzfragmenten werden aus dem Trigramm-Sprachmodell berechnet. Dabei bleiben aber die lokalen Zusammenhänge und Einschränkungen der Wortübergänge durch die CFG bestehen. In Tabelle 4 sind die Eigenschaften der resultierenden Basismodelle dargestellt.

Perplexität Trigramm	4,531
Anzahl Einstiegsregeln der CFG	62
Anzahl Wörter im Gesamtvokabular	2015

Tabelle 4: Daten der Basismodelle von Variante B

Es wurden insgesamt 33685 Wörter des Textkorpus durch Nichtterminale der CFG ersetzt.

5.2.3 Variante C - Kombination aus kurzen und vollständigen Regeln

Die dritte betrachtete Variante ist eine Kombination aus den beiden obigen Darstellungen. Hier wurde nur ein kleiner Teil der Regeln der Ausgangs-CFG anhand des im letzten Unterabschnittes erläuterten Vorgehens zerlegt. Dadurch können die Unterschiede in den Auswirkungen der obigen Varianten besser kontrolliert werden. Die Eigenschaften dieses Modells sind in Tabelle 5 beschrieben.

Perplexität Trigramm	5,122
Anzahl Einstiegsregeln der CFG	47
Anzahl Wörter im Gesamtvokabular	2015

Tabelle 5: Daten der Basismodelle von Variante C

Es wurden 54597 Wörter des Textkorpus ersetzt.

5.3 Resultate

Die folgenden Resultate wurden auf Basis der Vorgehensweise in Abschnitt 4 erzielt. Die Testmenge wurde zu Beginn von Abschnitt 5 beschrieben. Die Wortfehlerrate (WER) der vorgestellten Varianten des kombinierten Sprachmodells und die relative Veränderung zu den Ausgangsmodellen sind in Tabelle 6 aufgeführt.

Die Resultate zeigen in allen drei Varianten eine Verschlechterung der Erkennungsleistung.

Varianten	WER	Vgl. Ausgangs-Trigramm	Vgl. Ausgangs-CFG
Ausgangstrigramm	22,91%		
Ausgangsgrammatik	26,23%		
Variante A	25,86%	+12,88%	- 1,41%
Variante B	37,30%	+62,81%	+42,20%
Variante C	28,05%	+22,44%	+ 6,94%

Tabelle 6: Wortfehlerrate der einzelnen Varianten

Die besten Resultate liefert dabei Variante A. Dramatisch ist der Einbruch der Erkennungsleistung bei Variante B. Diese Verschlechterung deutet auf grundlegende Probleme dieser Art der Einbettung hin. Wie zu erwarten war, liegt die Erkennungsleistung von Variante C zwischen denen von Variante A und Variante B. Im folgenden Abschnitt sollen die möglichen Gründe für die hier erzielten Ergebnisse analysiert werden.

5.4 Analyse der Ergebnisse

Nach den bisherigen Ergebnissen stellt sich die Frage nach den Gründen für das schlechtere Abschneiden der einzelnen Varianten im Vergleich zu den Ausgangsmodellen. Die Korrektheit des Ansatzes wurde durch ein Basisexperiment verifiziert, in welchem als einzige Regeln der CFG die aus dem Ngramm-Modell 1:1 übernommenen Klassen, vgl. Abschnitt 3.4, vorkamen. Dabei konnte kein Unterschied in der Erkennungsleistung im Vergleich zum Trigramm-Ausgangsmodell

nach Abschnitt 5.1 festgestellt werden.
Als Ausgangspunkt der Analyse soll zunächst ein Beispiel für Unterschiede in der Erkennungsleistung durch die einzelnen Varianten gegeben werden. Das Beipiel ist direkt dem Erkenner entnommen. Der verwendete Referenzausdruck wird dabei sowohl von der Ausgangs-CFG als vom Ausgangs-Trigramm korrekt erkannt. Falsch erkannte Wörter sind durch Großschreibung gekennzeichnet:

Referenz: *how do i get to christofstrasse*

Variante A: *how do i get to christofstrasse FROM*

Variante B: *HOTEL EDEN to christofstrasse OF*

Variante C: *how do i get to IT POSTOFFICE*

Dieses Beispiel zeigt einige der wichtigsten Fehler der kombinierten Modelle auf. Der Fall in Variante A tritt dann auf, wenn eine Regel der Grammatik beendet wurde. Im kombinierten Modell kann die mit dem entsprechenden Nichtterminal erweiterte Ngramm-Historie fortgeführt werden und somit auch zu falscher Erweiterung der Historie führen. Das Ausgangs-Trigramm hat sich an dieser Stelle als robuster trainiert herausgestellt.
Ein Grund für die schlechten Resultate von Variante B ist die mangelnde Kontextinformation innerhalb der kurzen Regeln. Dadurch kann der Vorteil der Grammatik nicht genutzt werden, während sich aber die fehlende stochastische Information innerhalb der CFG-Regel negativ auswirkt. Da die Wörter des CFG-Vokabulars im Durchschnitt eine bessere Bewertung haben als Wörter die nur im Ngramm-Vokabular vorkommen, werden die Pfade der Grammatik oft überbewertet, was sich besonders bei kurzen Regeln negativ auswirkt.

Einer der entscheidenden Punkte und gleichzeitig auch eines der größten Probleme ist die Frage nach der besten Behandlung des Ein- und Ausstiegs aus den Grammatikregeln. Durch eine flexible Handhabung dieser Situationen soll ja gerade versucht werden, die Unzulänglichkeiten der CFG zu umgehen. Hier zeigt sich in der Analyse ein entscheidender Nachteil des Modells aus Abschnitt 4, der alle hier vorgestellten Varianten betrifft. Bei der Modellierung der Erweiterungsfunktion der Historie in Abschnitt 4.2.2 wurde festgelegt, dass wenn das zu erweiternde Wort ein Anfangswort einer CFG-Regel ist, *immer* in diese Regel auch verzweigt wird. Kommt das Wort auch im Ngramm-Modell vor, dann kann die Ngramm-Historie damit nicht erweitert werden. Es wird sich hier also eindeutig für den Einstieg in die Grammatik entschieden. Das Problem, welches aus diesem Vorgehen resultiert, lässt sich leicht am Beispiel aus Variante C erläutern. Betrachten wir dazu die Regel aus Variante C, die zur Entstehung des obigen Beispieles beigetragen hat:

public <how-to-go-there> = *how do i get there* |
how do i get to it |
give me directions there ;

Dazu ist zu bemerken, dass obige Regel die einzige der CFG ist, die die Wortfolge '*how do i get to*' abdeckt.
Fragt der Ibis-Decoder die Erweiterung der initialen Historie mit dem Wort *how* an, wird immer in die beste Regel mit *how* als Anfangswort erweitert. Nehmen wir zur Vereinfachung an, die

beste Regel wäre die oben dargestellte. Wurde im weiteren Verlauf des Decodierprozesses die Historie *how do i get to* erstellt, dann ist daraufhin die Bewertung durch das CFG-Basismodell für das Wort *it* optimal, während der Strassenname *christofstrasse* durch diese Regel maximal schlecht bewertet werden würde. Die korrekte Erkennung der Referenzäusserung ist durch das Sprachmodell somit nicht möglich, wenn keine andere Regel den Referenzsatz modelliert. Damit werden die Möglichkeiten der Erkennung eingeschränkt und kaum ein Zuwachs an Flexibilität in Hinsicht auf die Ausgangs-CFG erreicht. Dieser Sachverhalt wird in Abschnitt 6 wieder aufgegriffen werden.

Im Zusammenhang mit der Bewertung der Grammatikregeln ist im weiteren die Festlegung der Offsetberechnung kritisch zu sehen. Wie in Abschnitt 4.2.2 erwähnt, wird der Offset am Ende jeder CFG-Regel dem nachfolgenden Wort aufgeschlagen. Solange die CFG nicht am Ende angelangt ist, ist der aktuelle Pfad gegebenenfalls überbewertet. Durch die Eigenschaft des Suchvorgangs weniger wahrscheinliche Pfade aus Effizienzgründen nicht weiter zu betrachten, kann es passieren, dass während der Bearbeitung der überbewerteten Regel Pfade nicht weiter verfolgt werden, die sich nach der Offsetbewertung als wahrscheinlicher herausgestellt hätten. Theoretisch ist es also nötig, den Offset so früh wie möglich in die Bewertung eingehen zu lassen. Diese Tatsache wird in Abschnitt 6 untersucht werden.

6 Veränderungen des Anfangsmodells

Die Analyse der Resultate in Abschnitt 5.4 ergab einige Probleme der bisherigen Vorgehensweise. In diesem Abschnitt sollen nun darauf aufbauend zwei Veränderungen am bisherigen Modell vorgestellt werden.

6.1 Trennung des Vokabulars

Die erste Änderung betrifft den Einstieg in die CFG-Regeln. Wie in Abschnitt 5.4 untersucht, wird bei der Erweiterung einer Historie mit einem Einstiegswort immer in die CFG verzweigt. Daraus resultiert, dass in keinem Pfad des Ngramm-Modells dieses Wort vorkommen kann. Diese Beschränkung wird durch eine Trennung des Vokabulars der dem kombinierten Modell zugrundeliegenden Basismodelle aufgelöst. Das kombinierte Sprachmodell arbeitet nunmehr auf der Vereinigung zweier disjunkter Vokabularien. Dadurch ist es möglich, Entscheidungssituationen, z.B. welche Historie erweitert werden soll, eindeutig aufzulösen, da ein Wort nur in einem Modell vorkommen kann. Der bisherige Bewertungsvorgang nach Abschnitt 4.2.3 wird beibehalten, insbesondere wird die Bewertung der besten Regel auf das Anfangswort der entsprechenden Regeln übertragen.

Wie sieht diese Trennung aus? Zum einen wäre es möglich, zwei grundsätzlich disjunkte Vokabularien zu erstellen. Dies bedeutet, dass Wörter, die z.B. in der CFG vorkommen, nicht im Ngramm-Modell erkannt werden können. Daraus würde aber folgen, dass insbesondere das Ngramm-Modell abhänig von der CFG stark eingeschränkt wäre. Damit wäre kein Vorteil hinsichtlich der Flexibilität des kombinierten Modells erreicht. Dies wird dann deutlich, wenn man den Inhalt einer typischen Grammatikregel bedenkt. Schon die Beispielregel [to_place] = *to the cinema* ... aus Abschnitt 3.3 zeigt die Verwendung kurzer, häufig vorkommender Worte wie *to* und *the* in der Grammatik. Stünden diese Wörter aber nicht im Ngramm-Modell zur Verfügung, wären dessen Möglichkeiten stark eingeschränkt.

Obige Problematik wird durch eine gesonderte Kennzeichnung der Wörter der einzelnen Basisvokabulare umgangen. Dies geschieht, indem innerhalb der CFG den Wörtern ein gesondertes Symbol angefügt wird. Die resultierenden erkannten Wortfolgen bestehen daher aus Wörtern des Ngramm-Modells und gekennzeichneten Wörtern aus der CFG. Die gesonderte Kennzeichnung kann nach der Erkennung wieder entfernt werden. Durch die Kennzeichnung ist es möglich, dass ein Wort in beiden Historien vorkommen kann, ohne Kennzeichen in der Ngramm-Historie und mit in der CFG-Historie. Somit wird neben der Vereinfachung von Entscheidungssituationen auch das gewünschte Maß an Flexibilität des kombinierten Sprachmodells erreicht.

6.1.1 Resultate

Die mit dieser Änderung des Ansatzes erzielten Resultate sind in Tabelle 7 dargestellt.

Dabei ist eine deutliche Verbesserung der Erkennungsrate im Vergleich zum Ausgangsmodell zu verzeichnen. Insbesondere Variante A liefert ein Resultat, welches das gewünschte Ziel, die Verbesserung der Erkennungsleistung über die beiden Ausgangsmodelle hinaus, erreicht. Bemerkenswert ist die Wortfehlerrate von Variante B, die sich relativ zu den Ergebnissen in Tabelle 6 um 36,94% verbessert hat. Variante B und Variante C bleiben trotz der Verbesserung in ihrer Leistung hinter dem Ausgangstrigramm-Modell zurück.

Varianten	WER	Vgl. Ausgangs-Trigramm	Vgl. Ausgangs-CFG
Ausgangstrigramm	22,91%		
Ausgangsgrammatik	26,23%		
Variante A	21,28%	-7,11%	-18,87%
Variante B	23,52%	+2,66%	-10,33%
Variante C	24,45%	+6,72%	- 6,79%

Tabelle 7: Wortfehlerrate mit getrenntem Vokabular

6.1.2 Analyse der Ergebnisse

Die Resultate aus Tabelle 7 zeigen den Erfolg der vorgestellten Veränderung am Anfangssmodell. Die Trennung des Vokabulars führte zu einer deutlichen Erhöhung der Flexibilität des Ansatzes bei gleichzeitiger Vereinfachung der Berechnung. Stellt sich die Frage, warum die Verbesserung gerade auf Variante B einen solch starken Einfluss hat. Eine Erklärung dafür bietet die erhöhte Abdeckung von Wörtern, die als Einstiegswörter in Regeln der CFG dienen. Dadurch wird häufiger als in den anderen Varianten in die Grammatik verzweigt. Die Erhöhung der Flexibilität durch Trennung des Vokabulars hat entsprechend größte Auswirkungen auf Variante B.
Dass die Ergebnisse von Variante B und Variante C weiterhin hinter denen des Trigramm-Ausgangsmodells zurückbleiben, liegt wie in Abschnitt 5.4 erwähnt, an der unzureichenden Kontextinformation innerhalb der kurzen Regeln bzw. der fehlenden stochastischen Information der CFG. Dies spricht für die Bedeutung einer möglichst genauen Modellierung der CFG-Regeln.

6.2 Zeitpunkt der Offset-Berechnung

Die zweite am Anfangsmodell vorgenommene Veränderung betrifft den in den Abschnitten 4.2.2 und 5.4 diskutierten Zeitpunkt der Offset-Berechnung. Der bisherige Zeitpunkt, nach Beendigung einer CFG-Regel, kann zu einer Überbewertung des aktuell betrachteten Suchpfades führen. Aus diesem Grund wird dieser Ansatz so verändert, dass der Offset eines Regelübergangs innerhalb der CFG sofort in die Berechnung der Wortwahrscheinlichkeiten einfließt. Dadurch findet die Offsetberechnung nun ausschließlich dann statt, wenn die Historie im Zustand CFG-aktiv oder CFG-optional-aktiv ist. Für die Berechnung nach Abschnitt 4.2.3 gilt entsprechend:

$$P(w_{t_i j}|t_i) = Offset * P(w_{t_i j}|w_{t_i 1} \ldots w_{t_i j-1})$$

Dafür muss bei jedem Berechnungsvorgang innerhalb dieser Historienzustände herausgefunden werden, welcher Regel [rule_actual] ein Wort des CFG-Vokabulars am wahrscheinlichsten folgt. Deren Unterschied in der Einstiegswahrscheinlichkeit zur bisher angenommenen Regel [rule_previous] bildet den Offset.

$$Offset = \frac{P([rule\_actual]|X\ Y)}{P([rule\_previous]|X\ Y)}$$

6.2.1 Resultate

Die Ergebnisse der veränderten Offset-Berechnung sind in Tabelle 8 dargestellt. Dabei zeigt sich, dass die Veränderung der Offsetberechnung in allen Varianten eine Verschlechterung des Ergebnisses hervorgerufen hat. Insbesondere Variante C zeigt einen deutlichen Abfall der Erkennungsleistung um 31,56% relativ zur Leistung des Trigramm-Ausgangsmodells.

Varianten	WER	Vgl. Ausgangs-Trigramm	Vgl. Ausgangs-CFG
Ausgangstrigramm	22,91%		
Ausgangsgrammatik	26,23%		
Variante A	23,16%	+ 1,09%	-11,70%
Variante B	24,12%	+ 5,28%	- 8,04%
Variante C	30,14%	+31,56%	+14,91%

Tabelle 8: Wortfehlerrate mit veränderter Offset-Berechnung

6.2.2 Analyse der Ergebnisse

Die Gründe für die überraschende Verschlechterung der Erkennungsleistung sind nicht offensichtlich. Das Vorziehen der Offset-Berechnung bedeutet, dass vormals überbewertete Pfade nun weniger wahrscheinlich werden. Dadurch sollte die Erkennung theoretisch besser werden. Dies konnte trotz intensiver Überprüfung nicht bestätigt werden. Der Grund dafür scheint im Verlauf der Wortfehlerfunktion zu liegen. Wie sich in den Untersuchungen des folgenden Abschnitts 6.3 zeigte, ist die Wortfehlerfunktion durch viele lokale Minima gekennzeichnet. Daraus resultiert, dass eine Veränderung an der Berechnung, wie sie durch den Offset durchgeführt wurde, durchaus unerwartete Resultate liefern kann. Das insbesondere die Leistung von Variante C abfällt, scheint an der Wahl ihrer Regeln zu liegen, die sich in diesem Fall als besonders ungünstig herausstellte.

6.3 Gewichtung der Basismodelle

Gerade im letzten Abschnitt wurde deutlich, dass Veränderungen an der Berechnung der Pfadwahrscheinlichkeiten umfangreiche, teils unerwartete Ergebnisse liefern können. Um den Einfluss der beiden Basismodelle optimieren zu können, wird eine Gewichtung der einzelnen Basismodelle mittels *weight* vorgenommen. Die Berechnung der apriori-Wahrscheinlichkeit $P(W)$ einer Wortfolge W laut Abschnitt 4.2.3 ergibt sich damit zu:

$$P(W) = \prod_{i=1}^{m} (P(t_i|t_{m-n+1} \dots t_{m-1})^{weight} * P(\bar{w}_{t_i}|t_i)^{(2-weight)})$$

Somit wird jedes Wort abhängig aus welchem Basisvokabular es stammt entsprechend gewichtet. Der Wertebereich von $weight \epsilon [0,2]$ wurde aus Implementierungsgründen gewählt. Dabei gilt, dass je größer *weight* ist, die Basis-CFG umso mehr Einfluss hat. Der Gleichgewichtszustand zwischen den beiden Basismodellen ist entsprechend bei $weight = 1.0$.

6.3.1 Resultate

Die erreichte Wortfehlerrate der einzelnen Varianten anhand des Beispiels zweier unterschiedlicher Gewichtungen ist in Tabelle 9 angegeben. Dabei ist zu erkennen, dass die einzelnen Varianten von den selben Gewichten unterschiedlich profitieren können. Die Wortfehlerrate sank bei Variante C sogar unabhängig davon, ob dem Basis-Trigramm-Modell oder der Basis-CFG ein höheres Gewicht zugewiesen wurde.
Es wurde desweiteren untersucht, welche Auswirkungen die Gewichte auf den Ansatz der vorgezogenen Offset-Berechnung aus Abschnitt 6.2 haben. Die dabei erzielten Resultate sind in Tabelle 10 dargestellt. Auch hier zeigt sich, dass die Auswirkungen auf die einzelnen Varianten

Varianten	Gewicht	WER	Vgl. Ergebnisse Tabelle 7
Variante A Tab. 7		21,28%	
Variante A	0.9	22,24%	+4,51%
	1.1	22,47%	+5,59%
Variante B Tab. 7		23,52%	
Variante B	0.9	23,98%	+1,96%
	1.1	24,67%	+4,89%
Variante C Tab. 7		24,45%	
Variante C	0.9	23,64%	-3,31%
	1.1	23,50%	-3.89%

Tabelle 9: Wortfehlerrate mit unterschiedlichem Gewicht

Varianten	Gewicht	WER	Vgl. Ergebnisse Tabelle 8
Variante A Tab. 8		23,16%	
Variante A	0.9	22,38%	-3,37%
	1.1	24,35%	+5,14%
Variante B Tab. 8		24,12%	
Variante B	0.9	24,62%	+2,07%
	1.1	25,54%	+5,89%
Variante C Tab. 8		30,14%	
Variante C	0.9	28,05%	-6,93%
	1.1	30,18%	+0.13%

Tabelle 10: Wortfehlerrate mit unterschiedlichem Gewicht

sehr unterschiedlich sind.

6.3.2 Analyse der Ergebnisse

Durch die Nutzung eines Gewichts ist es möglich eine Art 'Feinabstimmung' vorzunehmen. Die Resultate zeigen, dass die Verschiebung des Einflusses der einzelnen Basis-Modelle durchaus positive Wirkung auf die Erkennungsleistungen haben kann. Dabei war häufig zu beobachten, dass die Verstärkung des Einflusses des Trigramm-Modells bessere Ergebnisse nach sich zieht, als die erhöhte Gewichtung des CFG-Modells. Besonders deutlich zeigt sich dies in den Resultaten von Tabelle 10.
Als sehr problematisch hat sich bei verschiedenen Tests die Wahl eines geeigneten Gewichts erwiesen. Wie im vorangegangenen Abschnitt erwähnt, scheint der Verlauf der Wortfehlerrate durch viele lokale Minima geprägt zu sein, so dass die Auswirkungen einer Veränderung der Gewichtung schwer vorhersagbar sind. Somit ist die Suche nach einem möglichst optimalen Gewicht sehr aufwändig. Aus diesem Grund können die im vorangegangenen Abschnitt genutzten Gewichte auch nur als exemplarisch angesehen werden.

7 Fazit

Auf Basis der in dieser Arbeit gewonnenen Resultate lässt sich ein geteiltes Fazit ziehen. Einige Ergebnisse sind sehr vielversprechend für weitere Untersuchungen, andere hingegen konnten die in sie gelegten Hoffnungen nicht erfüllen. Als bester Ansatz hat sich Variante A aus Abschnitt 5.2.1 erwiesen. Die Einbettung von Regeln in das Ngramm-Modell, aus denen ganze Sätze abgeleitet werden können, bietet dabei die optimale Kombination von Flexibilität und Kontextinformation. Mit Variante A wurde somit das Ziel dieser Arbeit erreicht: die Verbesserung der Erkennungsleistung unter Ausnutzung der Funktionsweise bereits bestehender Sprachmodelle. Als generell wichtigster Punkt hat sich die Ausnutzung aller möglichen Pfade durch Trennung des Vokabulars erwiesen. In einigen Fällen ist durch die Wahl einer möglichst optimalen Gewichtung eine zusätzliche Verbesserung der Ergebnisse möglich. Das Auffinden eines solchen optimalen Gewichts ist aber sehr aufwändig. Die besten Ergebnisse aller Varianten sind nochmals zusammengefasst in Tabelle 11 aufgeführt.

Varianten	WER	Vgl. Ausgangs-Trigramm
Ausgangstrigramm	22,91%	
Variante A	21,28%	-7,11%
Variante B	23,52%	+2,66%
Variante C	23,50%	+2,58%

Tabelle 11: Beste WER aller Varianten

Nicht erreicht werden konnte ein zufriedenstellendes Ergebnis für Variante B aus Abschnitt 5.2.2. Der Verlust an Kontextinformation durch Nutzung kurzer Regeln sowie die fehlende stochastische Information innerhalb der CFG-Regeln verhinderte in diesem Fall eine Verbesserung. Zusätzlich hat sich das stochastische Modell als sehr sensibel gegenüber Veränderungen erwiesen. Kleine Eingriffe resultierten teilsweise in großen Leistungsunterschieden. Darauf baut sich aber auch die Hoffnung, den flexiblen Ansatz der kurzen Regeln von Variante B effizienter einzusetzen. Besonders wichtig ist dabei eine genaue Modellierung passender Regeln und eine optimale Gewichtung der Basismodelle.

Erwähnt werden muss auch die generelle Anwendbarkeit des vorgestellten Ansatzes bzw. dessen Grenzen. Ein zentraler Punkt dabei ist, dass sowohl die einzelnen Modelle als auch deren Ergebnisse auf Basis einer beschränkten Domäne erstellt wurden. Dies hat bedeutende Auswirkungen auf den Einbettungsprozess, wie er in den Abschnitten 3.3 und 3.4 beschrieben wurde. Würde man versuchen, ein Ngramm-Modell und eine CFG zu kombinieren, die aus unterschiedlichen Domänen stammen bzw. sehr verschieden spezifisch sind, ist die Anwendbarkeit des vorgestellten Ansatzes beschränkt. Dies kommt von der Tatsache, dass die Grammatikregeln nur Wortfolgen des Ngramm-Textkorpus ersetzen, die sie auch abdecken. Je unterschiedlicher die Domäne des Trainingskorpus und der CFG sind, um so seltener ist diese Ersetzung möglich und umso eingeschränkter ist die Kombination.
Ein gerade auch in der automatischen Spracherkennung wichtiger Punkt ist die Laufzeit des Systems. Durch die Verwaltung zweier Historien und umfangreiche Berechnungen im Bereich des Übergangs zwischen den CFG-Regeln und der Ngramm-Historie ist der Aufwand sehr hoch. Hierbei ist aber anzumerken, dass der vorgestellte Ansatz keine Laufzeitoptimierung erfahren hat.

Neben der erwähnten Laufzeitoptimierung scheint insbesondere die Interpolation mit entsprechenden Ngramm-Modellen, wie sie in [Wan00] vorgestellt wurde, ein vielversprechender Weg zur weiteren Verbesserung des Ansatzes zu sein. Dabei kann die Interpolation zum einen mit Ngramm-Modellen der selben Domäne erfolgen, um somit ein robusteres Gesamtmodell zu erstellen. Zum anderen kann die Interpolation auch mit einem domänenunspezifischen Ngramm-Modell erfolgen, um die Generalität des Sprachmodells zu erhöhen.
Eine andere Verbesserungsmöglichkeit ist die Nutzung stochastischer Information innerhalb der CFG entsprechend [Wan00] oder [Ben00]. Insbesondere Variante B sollte von dieser Verbesserung profitieren und könnte somit die Leistung des Ausgangstrigramm-Modells übertreffen.

Literatur

[Moo95] R. Moore, D. Appelt, J. Dowding, J.M. Gawron, D. Morgan. "Combining linguistic and statistical language sources in natural-language processing for atis". In *Spoken Language Systems Technology Workshop*, pages 261-264, Austin, Texas, February 1995. Morgan Kaufmann Publishers, Inc.

[Gil98] J. Gillett and W. Ward. "A Language Model Combining Trigrams and Stochastic Context-Free Grammars". In *ICSLP*. 1998. Sidney, Australia.

[Ben00] J. Benedi and J. Sanchez. "Combination of N-grams and stochastic context free grammars for language modeling"., in *Proc. Coling-2000*, Saarbrucken, Germany, 2000.

[Wan00] Y.-Y. Wang, M. Mahajan and X. Huang. "A Unified Context-Free Grammar and N-Gram Model for Spoken Language Processing". In *Proc. ICASSP*, 2000.

[Ros00] R. Rosenfeld. "Two decades of statistical language modeling: Where do we go from here?".In *Proceedings of the IEEE*, 88(8). 2000.

[Dow93] J. Dowding, J. M. Gawron, D. Appelt, J. Bear, L. Cherny, R. Moore, and D. Moran,"GEMINI: A natural language system for spoken-language understanding". In *Proceedings of the 31st Annual Meeting of the Association for Computational Linguistics*, pp. 54–61, 22–26. Ohio State University, Columbus. Ohio. June 1993.

[Sol01] H. Soltau, F. Metze, C. Fügen, and A. Waibel, "A One Pass-Decoder Based on Polymorphic Linguistic Context Assignment", in *Proceedings of the ASRU*, Madonna di Campiglio Trento, Italy, December 2001.

[Fue01] C. Fügen, M. Westphal, M. Schneider, T. Schultz, and A. Waibel. "LingWear: A Mobil Tourist Information System", *HLT*, San Diego, 2001.

[Sun98] Sun Microsystems. Grammar Format Specification. Version October 1998. [WWW document]. URL http://java.sun.com/products/java-media/speech/forDevelopers/JSGF/