

Spontanisierung von Textkorpora für die Sprachmodellierung

Universität Karlsruhe (TH)

Diplomarbeit

Borislava Mimer

Betreuer:

Dipl.-Inform. Sebastian Stüker

Prof. Dr. Alex Waibel

August 2005

Hiermit erkläre ich, dass ich diese Diplomarbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe.

Karlsruhe, den 31.08.2005

Borislava Mimer

Danksagung

Ich möchte mich herzlich bei meinem Betreuer Sebastian Stüker

Zusammenfassung

In modernen Spracherkennungssystemen mit großem Vokabular ist das Aussprachewörterbuch eine der teuersten...

Aus diesem Grund wurde in letzter Zeit

Inhaltsverzeichnis

1	Einleitung	1
1.1	Motivation	1
1.2	Spontane Sprache	1
1.3	TC-STAR Projekt	2
1.4	Maschinelle Übersetzung	3
1.5	Automatische Spracherkennung	3
1.5.1	Sprachmodelle	3
1.6	Ziel der Arbeit	4
2	Verwandte Arbeiten	5
2.1	Modeling von Disfluencies	5
2.2	Sprachmodelladaption	5
2.3	Text Transformation Method	5
2.4	Multilinguale Sprachmodellierung	5
3	EPPS-Korpus	6
3.1	Das Europäische Parlament	6
3.2	Trainings- und Testdaten	8
3.3	Textauswahl und Datenaufnahme	8
4	Experimente	11
4.1	Basiserkenner	11
4.2	Ansatz mit Filled Pauses	11
4.3	Ansatz von D. Van Uytsel	11
4.4	Ansatz von Umut	11
4.5	Maschinelle Übersetzung	11
5	Ergebnisse: Vergleich und Analyse	12
6	Zusammenfassung	13
	Literatur	17

Tabellenverzeichnis

3.1	Statistik über die Äußerungen des GlobalPhone-Korpus	9
3.2	Größe des GlobalPhone Wörterbuchs	9
6.1	Wortakkuratheit in % für phonembasierte Basiserkenner	15
6.2	Wortakkuratheit in % für graphembasierte Basiserkenner	15
6.3	Wortakkuratheit in % für vollkontinuierlichen graphembasierte Spracherkenner in Deutsch bei Anwendung des flexiblen Ballungsverfahren	15

Kapitel 1

Einleitung

In der moderne Welt ziehen auf sich immer größere Aufmerksamkeit Technologien, die den Menschen bequeme Kommunikation mit den elektronischen Geräten anbieten. Benutzerfreundliche Schnittstellen werden in der Mensch-Maschine Dialog somit von zentrale Bedeutung. Leute müssen in der Lage sein, mit der Spracherkennungssysteme in natürliche Sprache zu kommunizieren. Automatische Erkennung der spontane Sprache ist in dem Fall eine der wichtigste Forschungsgebieten in der Wissenschaft der Spracherkennung. Ob die sprachgesteuerte Telefondienstleistungen oder die Steuerung von Handheld (?) Geräten - das ist nur ein Paar Möglichkeiten für die praktische Anwendungen von Systemen, die spontane Sprache verstehen.

Allerdings, um solche Systeme zu trainieren, braucht man eine Korpora zu bauen, die auf spontane Sprachdaten basiert ist. In der Regel aber, ist die Anzahl von Transkriptionen der spontan gesprochene Texten relativ klein und nicht immer genug um eine gute Korpora zu kreieren. Der Grund dafür ist, dass solche Transkriptionen zu erzeugen, teuer und aufwendig ist. Deswegen wird vermehrt nach die Möglichkeiten gesucht, die große, nicht auf spontane Sprachdaten basierte, Textkorpora zu spontanisieren erlauben.

1.1 Motivation

...

1.2 Spontane Sprache

Es existieren verschiedene Typen von Korpora über die kann man sagen, dass sie spontan sind. Das liegt daran, dass man viele Spontanitätsniveau unterscheidet. Welche von diesen Niveau für eine bestimmte Spracherkennungssystem benutzt wird, hängt von der Anforderungen für diese System ab. Bei den

Faktoren, die Spontanitätsniveau beeinflussen können, handelt es sich, z.B. um erforderliche Größe von Vokabular oder um verwendete Mikrofone.

1.3 TC-STAR Projekt

TC-STAR Projekt (Technology and Corpora for Speech to Speech Translation) wurde am 01.04.2004 ins Leben gerufen. Die Projektlaufzeit beträgt 36 Monaten. TC-STAR beschäftigt sich mit alle Kern Technologien für die Sprache-ins-Sprache Übersetzung (SST) Systeme: automatische Spracherkennung (ASR), maschinelle Übersetzung und Sprachsynthese. Mit diesem Projekt will man einen Durchbruch in SST erreichen, das zu einer bedeutende Reduktion von Unterschied zwischen menschlichen und maschinellen Übersetzungsleistungen verhilft. Bei der Bewertung von der Projekt's Erfolgsgrad werden Fortschritte beobachtet, die in jeder Bereich der SSL Technologie gemacht worden waren.

Moderne state-of-the-art (in Deutsch???) Forschungssysteme können keine akkurate Übersetzung von uneingeschränkte Sprache zustandebringen. Deswegen, sind die Entwicklungen notwendig, um die state-of-the-art Technologien für die Spracherkennung und Sprachübersetzung zu verbessern.

In Rahmen von Projekt wurden natürlichgesprächige Sprachdomäne (unconstrained conversational speech domains???) ausgewählt, d.h. politische Gespräche, Diskussion's Forume, die Rundfunk Nachrichten. Für drei Sprachen - europäisches Englisch, europäisches Spanisch und Chinesisch (Mandarin) - wurde entschieden, weil sie relevant für die europäische Gesellschaft und Ökonomie sind.

Während den Projekt Ablauf sollten technische Veränderungen vorgenommen und bestimmte Ziele erreicht worden: die Entwicklung von neuen Algorithmen und Methoden mit Integration von relevanten menschlichen Kenntnisse, die bei Übersetzung verfügbar sind. Als Beispiele dafür können folgende neue Ansätze dienen: die statistische Aussprachemodellierung von natürlichgesprächige Sprache (unconstrained conversational speech???) für automatische Spracherkennung, neue akustische und prosodische Modelle für Generierung von ausdrucksvolle Sprache in der Sprachsynthese und andere.

Eine von Projekt's Zielen ist die Implementierung der Evaluation's Infrastruktur, die auf Wettbewerb eingestellte Evaluation basiert, um der gewünschte Durchbruch in Sprache-ins-Sprache Übersetzung zu erreichen. Für die Entwicklung einer solche Evaluation's Infrastruktur ist the European Parliament Plenary Session (EPPS) ein interessante Domain. Zwei tasks (?) wurden während der TC-STAR'05 Evaluation unterstützt: the EPPS task für Englisch und Spanisch und the Broadcast News task (BN) für Mandarin. BN Englisch und Spanisch sind zurzeit nicht berücksichtigt, können aber in der künftige Evaluationen hinzugefügt werden.

1.4 Maschinelle Übersetzung

1.5 Automatische Spracherkennung

Ziel der automatische Spracherkennung ist es, eine Wortsequenz zu finden, für die die bedingte Wahrscheinlichkeit $P(W|X)$ gemäß des erkennerinternen Modells und unter Voraussetzung, dass das akustisches Signal beobachtet wird, am größten ist. Sprachsignal wird meistens in eine Form von eine zeitliche Folge von Merkmalsvektoren X verwendet.

Die Gleichung (1.1) präsentiert die *Fundamentalformel der Spracherkennung*, die durch die Anwendung der bekannten Bayes-Regel bekommen werden kann:

$$P(W|X) = \frac{p(X|W) \cdot P(W)}{p(X)} \quad (1.1)$$

$p(X)$ ist die a-priori Wahrscheinlichkeit dafür, dass die Folge von Merkmalsvektoren X beobachtet werden kann. $p(X|W)$ ist die bedingte Wahrscheinlichkeit, mit der das Signal X beobachtet wird, wenn die Wortfolge W gesprochen wurde. Diese Teil der Gleichung heißt im allgemein *Akustische Modell*. $P(W)$ definiert man als die a-priori Wahrscheinlichkeit dafür, dass die Wortsequenz W gesprochen wird und man spricht dabei von *Sprachmodell*.

Bei der Dekodierung wird versucht für eine gesprochene Äußerung X eine möglichst gute Hypothese \bar{W} der zugrundeliegenden Wortfolge zu finden.

$$\begin{aligned} \bar{W} &= \underset{W}{\operatorname{argmax}} P(W|X) \\ &= \underset{W}{\operatorname{argmax}} \frac{p(X|W) \cdot P(W)}{p(X)} \\ &= \underset{W}{\operatorname{argmax}} p(X|W) \cdot P(W) \end{aligned} \quad (1.2)$$

Wie man aus der Umformung der Gleichung (1.2) sieht, spielt Sprachmodell $P(W)$ eine wesentliche Rolle in der Ermittlung der besten Hypothese \bar{W} .

1.5.1 Sprachmodelle

Interpolation WER PPL POS

Das Sprachmodell dient zur Schätzung der Wahrscheinlichkeit $P(W)$ für eine gegebene Folge von n Wörtern $W=(w_1, w_2, \dots, w_n)$, unabhängig davon, wie der akustische Signal X aussieht. Zur Berechnung von $P(W)$ existieren zwei Verfahren: die statistisch und die linguistisch basierte Modellierung. Beim erste Verfahren

werden Wahrscheinlichkeiten für Wortübergänge auf großen Textkorpora berechnet, während bei linguistisch motivierte Modellierung die syntaktische (und semantische) Struktur von Sprache analysiert wird und aufgrund diese Kenntnisse die Wahrscheinlichkeit von Wortsequenz abgeleitet wird. Der statistische Verfahren hat sich in der automatische Spracherkennung gegenüber dem linguistischen durchgesetzt. Bei der Beschreibung wird deswegen nur die statistische Modellierung betrachtet.

1.6 Ziel der Arbeit

Die Aufgabe der Studienarbeit bestand darin, das flexibles Ballungsverfahren aus [St'03],[Pak99] auf graphembasierte Spracherkennungssysteme umzusetzen. Es mussten monolinguale Spracherkenner für zwei Sprachen - Deutsch und Englisch - trainiert werden. Bei der Anwendung des flexiblen Ballungsverfahrens sollte die Anzahl der Teilbäume und Gaussiens in zahlreichen Experimenten variiert werden. Drei Fragenmengen für den Entscheidungsbaum musste ich untersuchen.

Die Ergebnisse der Experimente machen es möglich, die Wortakkuratheit bei Anwendung des flexiblen Ballungsverfahrens mit der Akkuratheit bei der Verwendung des bisherigen Ballungsverfahrens zu vergleichen.

Kapitel 2

Verwandte Arbeiten

2.1 Modeling von Disfluencies

....

2.2 Sprachmodellenadaptation

....

2.3 Text Transformation Method

....

2.4 Multilinguale Sprachmodellierung

...

Kapitel 3

EPPS-Korpus

3.1 Das Europäische Parlament

Das Europäische Parlament ist eines der Organe der Europäischen Union. Seine Hauptaufgaben sind:

- Gesetzgebungsbefugnis
- Haushaltsbefugnisse
- die politische Kontrolle der europäischen Organe, insbesondere der Kommission.

Seit 1979 werden die Europaabgeordneten in allgemeinen und direkten Wahlen alle fünf Jahre gewählt. Repräsentanten aus 25 Nationen vertreten derzeit in dem Europäischen Parlament 455 Millionen Bürgerinnen und Bürger. Bis vor kurzem wurden dort 20 offizielle Sprachen gesprochen, am 13. Juni 2005 ist eine weitere Sprache als Amtssprache der Europäische Union anerkannt worden. Die Regelung tritt aber erst am 1. Januar 2007 in Kraft [Wik]. Nach den Wahlen von 2004 hat das Europäische Parlament 732 Abgeordnete, die sich nach der Bevölkerungszahl der Mitgliedstaaten verteilen. Die Zahl der Abgeordnete darf nach der Europäischen Verfassung auf höchstens 750 ansteigen.

Das Europäische Parlament hat seinen Sitz in Straßburg, wo die Europaabgeordneten in der Regel eine Woche im Monat tagen (Plenartagung). In Brüssel finden zusätzliche zweitagige Plenartagungen statt. Die Fraktionen und Ausschüsse treten in Brüssel und anderswo in Europa zusammen. Das Generalsekretariat des Europäischen Parlaments und dessen Dienststellen befinden sich in Luxemburg.

Den Vorsitz bei den Sitzungen führt President des Europäischen Parlaments. Zu dessen Aufgaben gehören unter anderem: auf die Einhaltung der Geschäftsordnung zu achten, die Ordnung zu wahren, das Wort zu erteilen, die Aussprachen für geschlossen zu erklären. Wenn der President einem Mitglied des Parlaments das Recht zum Auftreten gewährt, wird Sprecher's Mikrofone aktiviert. Die Einwürfe

von Parlament sind deshalb in den Aufnahmen nicht vorhanden.

Alle Europaabgeordneten haben das Recht, im Parlament die Amtssprache ihrer Wahl zu sprechen. Die Dolmetschern verschaffen simultane Übersetzung der original Rede in alle offizielle Sprache der EU, sowie in jede weitere Sprache, die das Präsidium für erforderlich erachtet.

Es ist möglich alle Rednern folgendermaßen zu kategorisieren:

- (native?) solche Rednern, die auf seine Muttersprache sprechen, und (non-native?) diejenige, die auf nicht Muttersprache sprechen (sprechen mehr oder wenige mit Akzent)
- Original Rednern und Dolmetschern.

Obwohl die Reden im Allgemein geplant sind, kann man bei fast allen Sprecher solche aus der spontane Sprache bekannte Effekte, wie Unschlüssigkeiten, Fehlerstarten usw., beobachten. Der Sprechstil von Dolmetschern kann man als etwas kabbelig bezeichnen: Intervallen, wo es ziemlich dicht gesprochen wird, wechseln sich mit Pausen ab, während dessen der Dolmetscher mit dem Lesen der original Rede beschäftigt ist.

Als „ausführlicher Sitzungsbericht“ („CRE = Compte Rendu in Extenso“) wird die Zusammenstellung der Redebeiträge der Abgeordneten im Plenum bezeichnet. Es wird in den Amtssprachen von jeder Sitzung verfasst. Die üblicherweise „Regenbogen“ genannte Fassung der ausführliche Sitzungsberichte in der Originalsprache des Redners werden zunächst an die Web-Seite von Europäischen Parlament, EUROPARL, als PDF-Datei veröffentlicht. Auf die Texten kann man nur via Datum (Sitzungstag) zugreifen. Wenige Arbeitstage später wird eine HTML-Fassung vorgelegt. Sie ist in einzelne Reden unterteilt und somit leichter zu lesen. Nach noch einige Tage wird sie in die Suche nach Rednern und Wort im Debatentitel aufgenommen. Die ursprüngliche Fassung wird anschließend durch eine revidierte Version ersetzt, in der, bei Notwendigkeit, die von Redner gewünschte Korrekturen durchgeführt sind.

Während der Phase der vorläufige Übersetzung wird der „Regenbogen“ in den folgenden Wochen in alle EU-Amtssprachen übersetzt. Die Texte wurden von den freiberuflichen Übersetzern an das Europäisches Parlament zurückgeschickt und - vom zuständigen Dienst ungeprüft - im Internet veröffentlicht. Nachdem alle Sprachversionen überprüft werden, werden sie endgültig im Internet publiziert.

„Regenbogen“ Fassungen und endgültige Versionen von Texten sind auf EUROPARL seit April 1996 bis zu Gegenwart verfügbar [EUR].

Bild 1 gibt den Überblick über die Struktur von existierende EPPS Daten.

„Regenbogen“ Fassungen und endgültige Versionen von Texten dienen zur hohe Lesbarkeit und verschaffen nicht exakte Wort für Wort Übertragung. Die original Rede kann von Unschlüssigkeiten, Fehlerstarten und Wortunterbrechungen gereinigt werden. Auch Umstellung, Ersetzung, Streichung und Einfügung von Wörtern ist möglich zu beobachten. Ein Beispiel dafür ist im Bild 2 zu sehen.

Europe by Satellite (EbS) ist eine TV Nachrichten Agentur der Europäischen Union, die spezialisiert sich auf Übertragung der Europe bezogene Information via Internet und Satellite [EbS]. EbS überträgt Live-Sendungen von EP Plenartagungen in der original Sprache und simultane Übersetzungen auf verschiedene Audio Kanäle via Satellite. Dafür verwendet man ein Kanal für jeder offizielle Sprache der Europäische Union und ein Kanal für die original Rede. Diese Kanäle sind zusätzlich verfügbar als 30 Minuten Internet streams (in Deutsch?) für die Dauer von eine Woche nach die Plenartagung. Audio Übertragungen sind monaural (auf Deutsch?). Internet Audio streams (?) haben Abtastrate von 16 kHz und sind mit RealAudio Sipro codec mit Bit Rate von 16 kbit/s verschlüsselt. Satellite Audio streams (?) haben Abtastrate von 48 kHz und sind mit MPEG 1 layer II codec mit Bit Rate von 64 kbit/s verschlüsselt.

In May 2004 wurde an der RWTH Aachen mit der Data Sammeln angefangen. EPPS Sendungen wurden in fünf Sprachen (Englisch, Spanisch, Deutsch, Französisch und Italienisch) aufgezeichnet worden. Aus der EbS Übertragung sind EPPS Aufzeichnungen von May bis Juli (OKTOBER???) gemacht worden. In May wurden Aufzeichnungen von Internet Audio streams (?) gemacht, in Juli - von Internet und Satellite Audio streams (?). Für jeder Sprache wurde in diese Period ganze 25 Stunden (ODER 48 Stunden für EN und SP ???)aufgezeichnet. Für Transkription wurden Internet streams (?) von May (???) und Satellite stream (?) von Juli (???) ausgewählt. Die EPPS Daten (welche ? nur englisch oder alle, oder EN und SP?) werden manuell segmentiert und transkribiert. ...???

????? WOHIN a restricted set of training corpora:— Englisch TC-STAR EPPS - about 40h (auf tc-star seite 41) of transcribed data (May 3-Oct.14, 2004) — Englisch EPPS final transcriptions, about 36M words (from parallel texts)-zum LM?

3.2 Trainings- und Testdaten

Die Sprecher wurden disjunkt im Verhältnis 80:10:10 in allen Sprachen auf drei Mengen verteilt: Trainings-, Kreuzvalidierungs- und Evaluationsmenge. Kreuzvalidierungsdaten werden auch Entwicklungsdaten genannt (engl. development data) und statt dem Begriff Evaluierungsdaten wird manchmal der Begriff Testdaten verwendet.

3.3 Textauswahl und Datenaufnahme

Die Daten für Englisch wurden nicht im Rahmen von GlobalPhone gesammelt, sondern vom Wall Street Journal-Korpus übernommen. Tabelle 3.1 und 3.2 [EUR] geben ein Überblick über die gesammten Datenbasis. 103 Sprecher haben Daten gespendet3.2. —————PEREDELAT', NE NUJNO—————

	#Äußerungen (Stunden)		
Sprache	Training	Kreuzvalidierung	Evaluation
EN	7,137 (15.0)	144 (0.4)	152 (0.4)
DE	9,259 (16.9)	199 (0.4)	250 (0.4)

Tabelle 3.1: Statistik über die Äußerungen des GlobalPhone-Korpus

Sprache	#Wörter
EN	9,461
DE	24,000

Tabelle 3.2: Größe des GlobalPhone Wörterbuchs

TRAINING

Zum Training und zur Evaluation beider Erkennen wurde der JANUS-Spracherkennung verwendet. JANUS (oder Janus) ist ein Projekt, das das Ziel hat, einen allgemein-nutzbaren Spracherkennungstoolkit für Forschung und Anwendungen zur Verfügung zu stellen. Die Software besteht aus dem JRTk (Janus Recognition Toolkit), das zur Entwicklung von Spracherkennungssystemen benutzt wird und dem Ibis Dekoder. JRTk wurde von Universität Karlsruhe (TH)(Deutschland) und Carnegie Mellon University (USA) entwickelt. JRTk ist in C implementiert und in eine Tcl/Tk-Umgebung eingebettet [Wik]. Für die monolingualen Erkennen wurde das JRTk-Spracherkennungstoolkit V 5.0 Patch-Level 13 verwendet [KSS03].

Um ein Training mit JANUS zu starten, braucht man eine Datenbasis, bestehend aus digitalisierten Aufnahmen und einer möglichst genauen Transkription des Gesprochenen. Die Datenbasis ist in eine Trainings-, Kreuzvalidierungs- und Evaluationsmenge aufgeteilt. Man braucht auch ein Aussprachewörterbuch, das alle im Training verwendeten und beim Evaluieren zu erkennenden Wörter beinhaltet.

TEST

Wenn der Erkennen trainiert ist, ist es wichtig zu wissen, wie gut er funktioniert. Neben dem akustischen Modell wird zusätzlich ein Sprachmodell benötigt und es muss ein Suchvokabular festgelegt werden.

Zur Bewertung eines Erkenners wird eine gesprochene Äußerung dekodiert und die resultierende Wortkette (Hypothese) abgespeichert. Um die Güte der Hypothesen zu bewerten, ist ein Fehlermaß erforderlich. Bei der Erkennung geht

es darum, aus der unendlichen Anzahl möglicher Satzhypothesen die richtige raus-zusuchen. Man spricht in diesem Zusammenhang auch von Suche.

Die Dekodierung für die monolingualen Erkennen wird mit Hilfe des Ibis Dekoder durchgeführt.

Der Ibis Dekoder ist Ein-Suchlauf Dekoder. Es ist möglich, den Dekoder für Lattice-Rescoring zu benutzen.

Tests werden normalerweise auf einer Kreuzvalidierungsmenge durchgeführt, um festzustellen, dass sich die Erkennungsrate verbessert und keine Probleme aufgetreten sind.

Kapitel 4

Experimente

4.1 Basiserkenner

...

4.2 Ansatz mit Filled Pauses

4.3 Ansatz von D. Van Uytsel

4.4 Ansatz von Umut

...

4.5 Maschinelle Übersetzung

Kapitel 5

Ergebnisse: Vergleich und Analyse

In diesem Kapitel werden die Ergebnisse der Experimente bewertet.

Kapitel 6

Zusammenfassung

Die Idee, ...

verwendet, nur statt dem diakritischen Zeichen „Umlautpunkt“, wird das wellenförmigen Zeichen „Tilde“ (\sim) vor dem Buchstaben eingegeben... *ch*, wie in *suchen*..... relativ geringe Änderungen erfahren. Beim schriftlich
Die Anzahl der verschiedenen, kontextabhängigen Modelle kann schon bei relativ kleinen Kontextbreiten sehr groß sein....

Die Ballungsalgorithmen unterscheiden sich nach folgende Kriterien [OB00]:

- welche Grundeinheiten zur Ballung verwendet werden,
- ob es ein agglomeratives (bottom-up) oder divisives (top-down) Verfahren ist,
- welches Distanzmaß benutzt wird (Likelihood-Distanz, Entropie-Distanz).

Kai-Fu Lee hat als erster ein agglomeratives Ballungsverfahren durchgeführt [SO96]. Als Distanzmaß (Beginn/Mitte/End) aller Phonemen konstruiert. (SLPM). SLPM ermöglicht alternative Realisierungen der [Shr94]

In letzter Zeit wurde vermehrt die Verwendung von Graphemen für Spracherkennungssysteme untersucht. Die Arbeiten aus diesem Gebiet sind in der Studienarbeit nachgeforscht worden.

Schillo, Fink und Kummert haben zwei Ansätze zum Bau eines graphembasierten Spracherkennungssystems präsentiert: Erkennen für kontextunabhängige graphemische Einheiten mit verschiedene untersucht:

1. Phonem-Graphem Fragen: Der für den phonembasierten Erkennen verwendeter Fragenkatalog wurde mit Hilfe sprachabhängiger Phonem-zu-Graphem Abbildungen umgewandelt und für die Erzeugung kontextabhängiger Modellen benutzt.
2. Bottom-up Entropy Fragen: Die Menge von Monographemen wurde mittels eines agglomerativen (bottom-up) Ballungsverfahren ...
3. Hybrid Entropy Fragen: Das Ballungsverfahren stellt sich als Hybrid von top-down (divisive) und bottom-up Verfahren dar.
4. Singleton: Man fragt, welche Art von Graphemen im linken oder rechten Kontext sind. Jede Frage besteht aus einem Graphem.

Die Singleton Fragen liefern niedrigere Wortfehlerrate als Phoneme-Grapheme Fragen in Englisch und in Deutsch, aber nicht in Spanisch...

Die Tabelle 6.1 fasst die Ergebnisse der Experimenten für phonembasierte

Sprache	Kreuzvalidierungsdaten
DE	85,56
EN	90,3

Tabelle 6.1: Wortakkuratheit in % für phonembasierte Basiserkennung

Basiserkennung zusammen.

Im Laufe der Studienarbeit wurden graphembasierte Basiserkennung unter Verwendung der bisherigen ...

Die Ergebnisse des Tests sind in Tabelle 6.2 zu sehen:

Sprache	Kreuzvalidierungsdaten
DE	85,28
EN	81,88

Tabelle 6.2: Wortakkuratheit in % für graphembasierte Basiserkennung

In der vorliegenden Arbeit habe ich die ... Tabelle 6.3 gibt Auskunft über Verfahren:

	Deutsch	
	Kreuzvalidierungsdaten	
# Gaussiens für semikont. Erkennung	Baum ohne Unterteilung und Singleton	Baum mit Unterteilung und ergänzte Singleton
256	85,15	85,31
1024	85,35	85,69
1500	-	85,78
2048		85,51

Tabelle 6.3: Wortakkuratheit in % für vollkontinuierlichen graphembasierte Spracherkennung in Deutsch bei Anwendung des flexiblen Ballungsverfahrens

1. Singleton.
2. Ergänzte Singleton.
3. Phonem-Graphem Fragen.

Literaturverzeichnis

- [EbS] European Union’s TV news agency, “Europe by satellite,”. <http://europa.eu.int/comm/ebs/>.
- [EUR] The Secretariat of the European Parliament, “EUROPARL: aus dem Plenum,”. <http://www.europarl.eu.int/plenary/>.
- [KSS03] M. Killer, S. Stüker, and T. Schultz. Grapheme Based Speech Recognition. In *Proceedings of the 8th European Conference on Speech Communication and Technology (Eurospeech-2003)*, Geneva, Switzerland, September 2003.
- [OB00] D. Oppermann and S. Burger. What Makes Speech Data Spontaneous? In *P....*, volume 4, pages 584–587, ..., 2000.
- [Pak99] S. V. Pakhomov. Modeling Filled Pauses in Medical Dictations. In *Proceedings....*, pages 845–848, AAA, May 9999.
- [Shr94] E. Shriberg. *Preliminaries to a Theory of Speech Disfluencies*. Phd thesis, University of California at Berkeley, 1994.
- [SO96] M. Siu and M. Ostendorf. Modeling Disfluencies in Conversational Speech. In *Proc. International Conference on Spoken Language Processing*, volume I, pages 386–389, Atlanta, U.S.A., October 1996.
- [St’03] Sebastian Stüker. Multilingual Articulatory Features. Master’s thesis, Carnegie Mellon University, USA und Universität Karlsruhe (TH), Germany, 2003.
- [Wik] “Wikipedia, der freien Enzyklopädie,”. http://de.wikipedia.org/wiki/Amtssprachen_der_Europäischen_Union.