

Sprachgenerierung

Referat von Christian, Marc-Björn und Peter
im Rahmen des studentischen Projektes
ESPRESSO

- 1 -

Christian Nawroth
Peter Becker
Marc-Björn Sedel
© 2004

Motivation

- ☉ Grundlegende Motivation:
 - ☉ Mensch-Maschine-Interaktion
- ☉ Ziel: menschliche Artikulationsweise
- ☉ Bisher: nur "wenige" statische Sätze vom Rechner
- ☉ Lösungsansätze → NLG

- 2 -

Christian Nawroth
Peter Becker
Marc-Björn Sedel
© 2004

Gliederung

- ☉ Geschichtliche Entwicklung
- ☉ NLG (Natural Language Generation)
- ☉ Linguistische Textanalyse
- ☉ Sprachsynthese
- ☉ Arten der Sprachgenerierung
- ☉ Prosodie
- ☉ Beispiel: Verbmobil
- ☉ Anwendungsgebiete
- ☉ Zusammenfassung
- ☉ Quellen

- 3 -

Christian Nawroth
Peter Becker
Marc-Björn Sedel
© 2004

Einordnung des Themas

- ☉ Sprachgenerierung hat viele Teilgebiete:
 - ☉ Textanalyse
 - ☉ Textgenerierung
 - ☉ Sprachsynthese
- ☉ Sprachgenerierung steht im großen Kontext:
 - ☉ Computer linguistic / Sprachtechnologie
- ☉ Älteste Forschungsrichtung der Sprachtechnologie:
maschinelle Sprachsynthese

- 4 -

Christian Nawroth
Peter Becker
Marc-Björn Sedel
© 2004

Geschichtliche Entwicklung

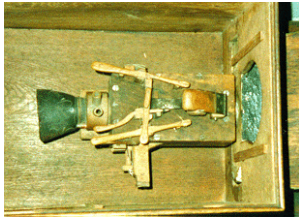
Mechanische Systeme

- C.G. Kratzenstein (1779): Vokale



FIG. 9. Kratzenstein's resonators for synthesis of vowel sounds. The resonators are actuated by blowing through a free, vibrating reed into the lower end. The i sound is produced simply by blowing into the lower pipe without a reed.¹

- Baron Wolfgang von Kempelen (1791): Einzellaute



Geschichtliche Entwicklung (II)

- Sir Charles Wheatstone (1838): Lautübergänge
- Faber (1846): "Euphonia"



Geschichtliche Entwicklung (III)

Elektische Systeme

- Homer Dudley (1939): Voder

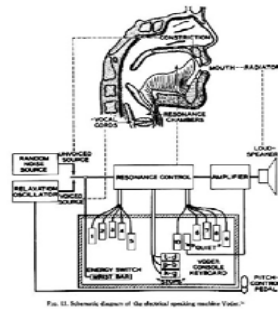
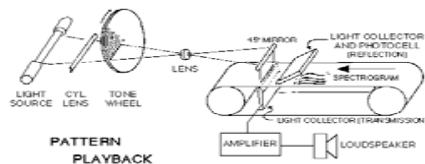


FIG. 11. Schematic diagram of the electrical speaking machine Voder.¹

- Frank Cooper (1951): Pattern Playback von Spektrogrammen

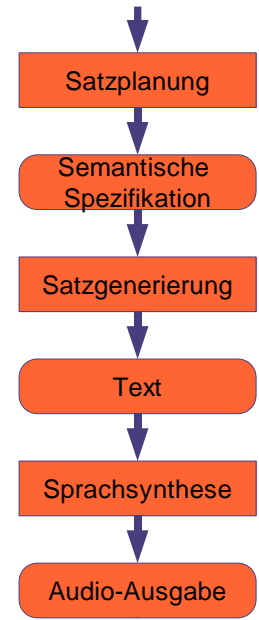


Geschichtliche Entwicklung (IV)

- Akustische Theorie der Sprachlauterzeugung
 - Fant (1953), Walter Lawrence (1953): OVE I, serielle Filter
 - Holmes (1973): parallele Filter
 - Klatt-Synthesizer (1980): hybride Filterstruktur
- Joseph Olive (1970): aufgenommene Segmente menschlicher Sprache zusammenfügen
- Amiga (1985): rsynth text-to-speech (TTS) translator

Generieren

- ☉ Sätze erzeugen...
 - ☉ ...die gemäß einer Phrasenstrukturgrammatik syntaktisch wohlgeformt sind
 - ☉ ...die einer gegebenen Syntaxstruktur entsprechen



Natural Language Generation (NLG)

- ☉ Entwicklung von computerbasierten Systemen, die Texte generieren.
 - u.a. FB10 der Uni Bremen → John Bateman
- ☉ Generierung natürlicher Sprache
- ☉ maschineninterne Repräsentationen → natürlichsprachige Oberflächenform
- ☉ Realisierung der Oberflächenstruktur:
 - ☉ morphologische, lexikalische und grammatische Phänomene, wie z.B. Wort- und Satzgliedstellung, Tempus, Modus, Satzform, ...

Natural Language Generation (NLG)

- ☉ Komplexe Aufgabe der Generierung zerlegen
 - ☉ "What to say?"
 - ☉ "How to say (it)?"
 - ☉ → dadurch semantische und syntaktische Umschreibung
- ☉ andere Einteilung: kognitive und linguistische Domäne

Natural Language Generation (NLG)

- ☉ Erweiterung des Dualismus → Trikolon
 - ☉ "Why say it?" → benötigt pragmatisches Wissen



- ☉ viele Systeme aber oft "dualistisch"
 - ☉ Pragmatik nicht immer im Vordergrund
 - ☉ Erzeugung konsistenter Texte auch ohne Pragmatik

Natural Language Generation (NLG)

- ☉ Drei zentrale Fragestellungen bei der Generierung:
 - ☉ Annäherung an den menschlichen Spracherzeugungsprozess
 - ☉ Modellierung der Generierung durch Planungsprozesse
 - ☉ Suche nach zulässigen Endketten einer Grammatik / Semantik

Architekturen:

☉ sequentiell



☉ integriert



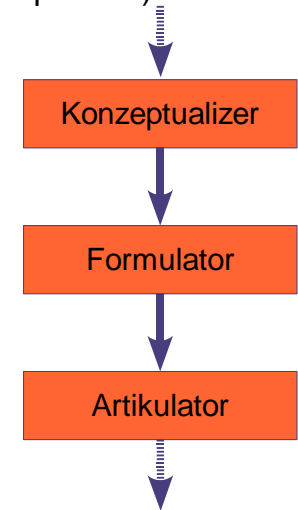
☉ interaktiv



Natural Language Generation (NLG)

☉ LEVELT: dreistufige Architektur (sequentiell)

- ☉ Konzeptualizer
 - inhaltsbestimmende Generierung
- ☉ Formulator
 - natürlichsprachige Oberflächenstruktur
- ☉ Artikulator
 - akustische Äußerung



Natural Language Generation (NLG)

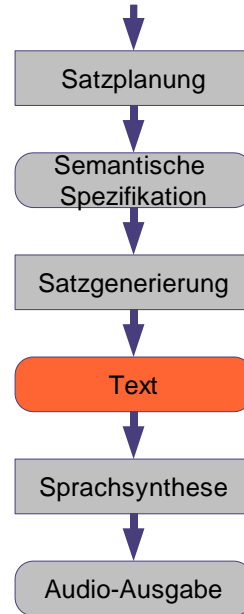
- ☉ spezielle, weiterführende Ziele:
 - ☉ Multilingualität: ein System für mehrere Sprachen
 - ☉ Inkrementalität: schrittweises Arbeiten ("Output" stückweise generieren)
 - steigert Geschwindigkeit (kürzere Antwortzeiten)
 - psychologisch adäquat

Natural Language Generation (NLG)

- ☉ Zufällige Generierung
 - ☉ Inversion vom Parsing
 - ☉ mit reversiblen Grammatiken
- ☉ Allgemeiner:
 - ☉ Input: strukturierte Daten
 - ☉ Output: passender natürlich-sprachlicher Text

Generieren

- ☉ Sätze erzeugen...
- ☉ ...die gemäß einer Phrasenstrukturgrammatik syntaktisch wohlgeformt sind
- ☉ ...die einer gegebenen Syntaxstruktur entsprechen



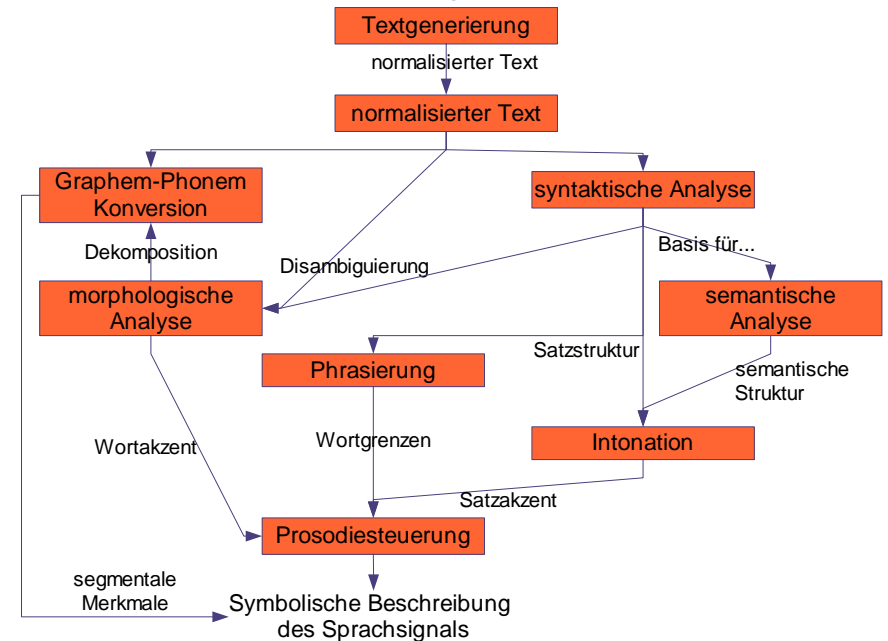
Linguistische Textanalyse

- ☉ Tokenisierung
- ☉ Textnormalisierung
- ☉ Lexikalische und morphologische Analyse
 - ☉ morphologisches Zerteilen der Wörter in Präfixe, Stamm-Morpheme und Suffixe
 - ☉ Kontextauswertung (und Disambiguierung)
- ☉ Phrasierung und Akzentuierung
- ☉ Syllabifizierung
- ☉ uvm.

Gliederung

- ☉ Geschichtliche Entwicklung
- ☉ NLG (Natural Language Generation)
- ☉ Linguistische Textanalyse
- ☉ Sprachsynthese
- ☉ Arten der Sprachgenerierung
- ☉ Prosodie
- ☉ Beispiel: Verbmobil
- ☉ Anwendungsgebiete
- ☉ Zusammenfassung
- ☉ Quellen

Sprachsynthese



Exkurs: Phonetik

- ☉ Kleinste artikulatorische Einheit
- ☉ Verschiedenste Arten von Lauten
 - ☉ Konsonanten, Vokale, Affrikate, Diphtonge etc
- ☉ Einheiten können je nach Wortstellung und Dialekt variieren oder ganz wegfallen
 - ☉ Ehe, Post/Polizei, reden→“redn”, Garten→“Gachten”, fünf→“fümf”, ich/ach, ...
- ☉ Abhängigkeit von Kontext, Morphologie, Sprechstil
- ☉ Verschiedene Bedeutungen je nach Betonung
 - ☉ 'umfahren, um'fahren; Wach'stube, Wach'stube
- ☉ Laute können in verschiedenen Sprachen andere Gewichtung haben

- 21 -

Exkurs: Phonetik

- ☉ Allophone
 - ☉ Varianten eines Phonems in spezifischen Kontexten, z.B. kurzes oder langes o
- ☉ Diphone
 - ☉ Verbindungen zweier Phoneme, im Sprachsignal meist von Mitte zu Mitte
 - ☉ Phonetik: fo – on – ne – et – ti – ik
- ☉ Triphone, Halbsilben, ...

- 22 -

Syntheseverfahren

- ☉ Formant- bzw. Regelsynthese
 - ☉ Erzeugung der Sprachsignale aus modifizierten Sinuswellen
 - Anregungssignal, div. Filter
 - ☉ Imitation der akustischen Spracheigenschaften
 - ☉ Benötigt sehr viele Regeln
 - ☉ Intonation und Lautdauer leicht zu modifizieren
 - ☉ Mangelnde Natürlichkeit der Stimme

- 23 -

Syntheseverfahren (2)

- ☉ Artikulatorische Synthese
 - ☉ Weiterführung der Formantsynthese
 - ☉ Nachbildung des kompletten menschlichen Sprechapparats
 - ☉ Sehr rechenaufwändig
 - ☉ Z.Z. nur für Forschung interessant

- 24 -

Syntheseverfahren (3)

- ☉ Konkatenationssynthese
 - ☉ Aufgenommene Sprachsignale werden in einzelne morphologische Segmente zerteilt und katalogisiert
 - ☉ Segmente werden wieder zu Wörtern zusammengebastelt
 - ☉ Natürlicher Klang der Stimme
 - ☉ De-facto Standard für bestehende Systeme

Syntheseverfahren (4)

- ☉ Konkatenationssynthese
 - ☉ Probleme:
 - Je "einfacher" die Einheiten, desto geringer ihre Anzahl und desto geringer die Sprachqualität
 - Fließende Übergänge bei Allophonen schwierig, daher heute oft: Diphone bzw. Halbphoneme
 - Nachträgliche Änderung von Lautdauer und Tonhöhe schwierig, verfälscht u.U. den Klang
 - u.U. sehr viele Segmente benötigt
 - Homographie (gleiche Wörter, unterschiedliche Aussprache) und Prosodie (Satzmelodie)

Arten von Sprachgenerierung

- ☉ Text To Speech
- ☉ Concept To Speech
- ☉ Document To Speech
- ☉ Canned Text

Text To Speech (TTS)

- ☉ Text → Linguistische Analyse → Prosodie (Lautdauer, Intonation) → Synthese (Einheitenauswahl, Einheitenverkettung, eigentliche Synthese) → synthetische Sprache
- ☉ Oft mit Hilfe von "Konservenphrasen"
- ☉ Probleme:
 - ☉ Je "kleiner" die Einheiten, desto geringer die Anzahl und desto geringer die Qualität
 - ☉ Fließende Übergänge schwierig, daher heute oft: Diphone bzw. Halbphoneme
 - ☉ Homographie (gleiche Wörter, unterschiedliche Aussprache) und Prosodie (Satzmelodie)

Concept To Speech (CTS)

- ☉ “Erzeugung von synthetischer Sprache auf der Basis von pragmatischem, semantischem und Diskurswissens”
- ☉ Inhalte des Textes sind dem System bekannt, dadurch:
 - ☉ linguistische Repräsentation der Daten nötig
 - ☉ Pragmatik, Semantik, Syntax und Phonetik fließen mit ein
 - ☉ Generierung von geschriebenem Text überflüssig
- ☉ Interdisziplinäres Forschungsobjekt: Linguistik, KI, Kognitionsforschung, Signalverarbeitung
- ☉ Ziele: Einbeziehung von Prosodie, um Sinn des Textes hervorzuheben
- ☉ Anwendung: Intelligente Auskunftssysteme

Canned Speech

- ☉ Wiedergabe von größeren Sprachbausteinen
- ☉ sehr gute Sprachqualität möglich
- ☉ oft eingesetzt bei Auskunftssystemen
- ☉ relativ “einfach” zu realisieren
- ☉ basiert auf Domänenwissen
- ☉ nicht sehr flexibel, da nur geringer Wortschatz verfügbar → Probleme bei Erweiterung

Document To Speech

- ☉ wie TTS, nur zusätzlich Einbeziehung von Struktur- und Stilelementen, z.B. Tabellen, Grafiken etc.
- ☉ Diese Daten müssen strukturiert werden, mögliche Ansätze in SGML, XML etc.

Prosodiegenerierung

Prosodie

- ☉ beschreibt:
 - ☉ Betonung, Satzmelodie, rhythmisches Satzmuster
- ☉ Ziel:
 - ☉ Verständlichkeit und Natürlichkeit
- ☉ Motivation:
 - ☉ Entwicklung neuer anspruchsvoller Anwendungen
 - Charakteristik eines Sprechers
- ☉ Stand:
 - ☉ schon weit erforscht
 - ☉ Weg vom Text zur synthetischen Prosodiegenerierung noch weit

Prosodie – automatisch generieren?

- ☉ Frage: Kann man einen Text vorlesen, ohne ihn zu verstehen?
- ☉ Problem:
 - ☉ Prosodie stark abhängig von lexikalischen Faktoren, Syntax/ Semantik, Pragmatik und Weltwissen.
- ☉ Lösungsansatz:
 - ☉ Keine allgemein gültige Methode bekannt.
 - ☉ Via Syntax
 - ☉ Via Satzoberflächenstruktur
 - ☉ Akzentuierung?
 - ☉ Lautdauer?

Gliederung

- ☉ Geschichtliche Entwicklung
- ☉ NLG (Natural Language Generation)
- ☉ Linguistische Textanalyse
- ☉ Sprachsynthese
- ☉ Arten der Sprachgenerierung
- ☉ Prosodie
- ☉ Beispiel: Verbmobil
- ☉ Anwendungsgebiete
- ☉ Zusammenfassung
- ☉ Quellen

Verbmobil

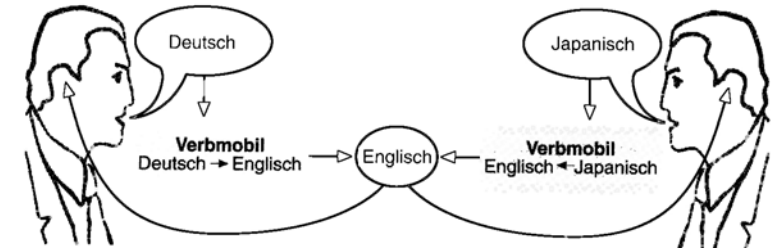
Ein Beispiel aus der Praxis
der Sprachgenerierung

Verbmobil - Motivation?

- ☉ unter anderem diese:
 - ☉ 11 offizielle Sprachen in der EU
Englisch, Deutsch, Französisch, Italienisch, Spanisch, Portugiesisch, Griechisch, Holländisch, Dänisch, Schwedisch, Finnisch
 - ☉ das ergibt: 110 Übersetzungsrichtungen
 - ☉ Automatisierung höchst erwünscht
 - ☉ bisher: über 1000 Übersetzer bei der EU
 - ☉ System: Euro Systran
- ☉ Maschinelle Übersetzung bisher: Babelfish, T1, und andere
→ unbefriedigend!

Verbmobil (1)

- ☉ Verbundprojekt des BMFT mit DFKI
- ☉ Ziel: Unterstützung eines Dialogs mit
fremdsprachlichen Gesprächspartnern
- ☉ Beispiel:
 - ☉ Unterstützung eines Dialogs zwischen einem Japaner und
einem Deutschen, die "ein bisschen" Englisch
sprechen können



Grafik entnommen aus "Spektrum der Wissenschaft" 03/94; vgl. Quellen

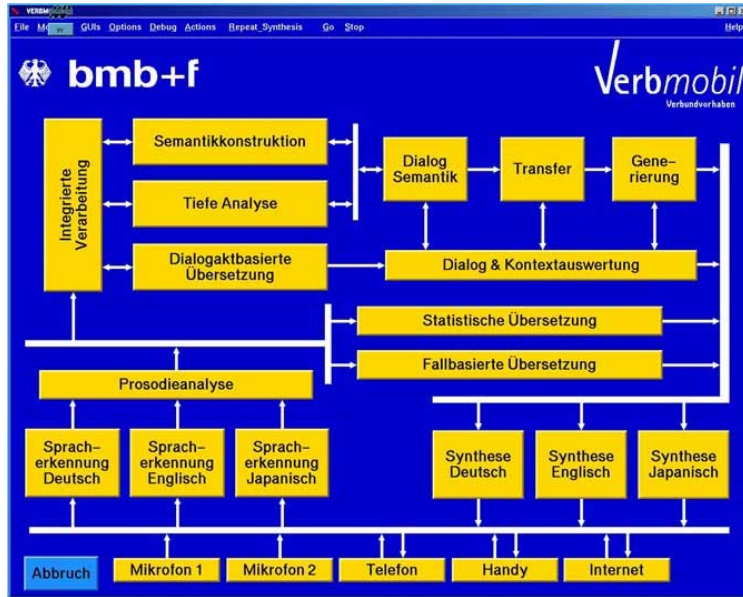
Verbmobil (2)

- ☉ Übersetzungsprozess:
 - ☉ Erkennung gesprochener Spontansprache
 - ☉ Analysieren der Eingabe
 - ☉ Übersetzung in Fremdsprache
 - ☉ Satzgenerierung
 - ☉ Ausgabe
- ☉ 1. Phase: 1993 – 1996
- ☉ 2. Phase: 1997 – 2000

Verbmobil (3)

- ☉ 2. Phase:
 - ☉ Domänen: Reiseplanung & Hotelreservierung
 - ☉ 10.000 Sprachpaare, Deutsch-Englisch
2.500 Sprachpaare, Deutsch-Japanisch
 - ☉ Klärungsdialoge
- ☉ Integration der Satzmelodie:
 - ☉ Wortfolge "Ja-zur-Not-geht-es-auch-am-Samstag"
 - "Ja, zur Not geht es auch am Samstag."
 - "Ja, zur Not! Geht es auch am Samstag?"
- ☉ Prosodiemodul erkennt Satz- und Phrasengrenzen
- ☉ 70% approximativ korrekte Übersetzungen

Verbmobil (4)



Grafik: <http://verbmobil.dki.de/>

- 41 -

Verbmobil (5)

- ☉ Multifunktionalität
- ☉ Multilingualität
- ☉ Multimedialität
- ☉ Mobilität
- ☉ Multiparty-Funktionalität

- 42 -

Anwendungsgebiete

- ☉ Hilfsmittel zur Computernutzung für blinde und sehbehinderte Menschen
- ☉ Auskunfts- / Buchungssysteme
- ☉ Übersetzer
- ☉ Zugriff auf Datenbanken
- ☉ Sprachlernprogramme
- ☉ Sprechende Bücher/ Spielzeuge
- ☉ Medizinischer/ Industrieller Bereich

- 43 -

Gängige Programme

- ☉ Webreader
 - ☉ WebSpeech (G Data Software)
- ☉ Screenreader
 - ☉ Blindows 3.01 (Audiodata)
 - ☉ Window-Eyes 4.2 (FluSoft)
- ☉ Mails/ Dokumente - Reader
 - ☉ Logox 4 Pro (G Data Software)
 - ☉ CoolSpeech 5 (ByteCool Software)
- ☉ Übersetzen von Schriftdokumenten in Audiofiles:
 - ☉ TextSound (ByteCool Software)

- 44 -

Stimmen

- ☉ Nicht nur Engines wichtig, sondern auch Stimmen:
 - ☉ Alle Voices für SAPI (Speech Application Programming Interface)
 - ☉ Sehr unterschiedliche Ergebnisse:
 - Microsoft
 - TruVoice
 - L&H
 - AT&T und Cepstrahl: Fast schon natürlichsprachlich

State of the art

- ☉ Sprachgenerierung gewinnt in allen Lebensbereichen immer mehr an Bedeutung
- ☉ Hohes Niveau
 - ☉ Hoher Spezialisierungsgrad
- ☉ Noch Mängel vorhanden (bsp: Cepstrahl)
 - ☉ Umgang mit unbekanntem Wörtern
 - ☉ Interpunktion am Satzende

Ausblick

- ☉ Erschaffen der perfekten Stimme?
- ☉ Charakteristik der Sprecher phonetisch beschreibbar?
- ☉ Erzeugung von Gesang (1996 Alexander Guelfenburg "Debut!")
- ☉ Automatische Texttypberücksichtigung bei Sprachsynthese
 - ☉ Gedichte (emotional) vs. Gebrauchsanweisung (sachlich)
 - ☉ Bsp.: Logox
- ☉ Natürlichsprachlicher Dialog

Quellen / Literaturhinweise

1. Möbius, B.: "VL Sprachsynthese 1".
<http://www.ims.uni-stuttgart.de/Lehre/teaching/2002-WS/Synthese1/index.html> (22.01.2004)
2. Breitenbücher, M.: "Datenbasierte Methoden der Sprachsynthese". Stuttgart, 1999
(<http://elib.uni-stuttgart.de/opus/volltexte/1999/448/>) (22.01.2004)
3. Schilder, F., Guhe, M.: "VL Semantische Sprachverarbeitung".
http://www.informatik.uni-hamburg.de/WSV/teaching/vorlesungen/SemSprach_SoSe03.html
(22.01.2004)
4. Gómez, R.: "Akustisch-phonetische Aspekte der Spracherkennung". Berlin, 2003
(<http://user.cs.tu-berlin.de/~ramiro/downloads/Spracherkennung.pdf>) (22.01.2004)
5. Fellbaum, K.: "VL Ausgewählte Gebiete der Sprachverarbeitung".
<http://www.kt.tu-cottbus.de/teleteaching/> (22.01.2004)
6. Wahlster, W.: "Verbmobil - Übersetzungshilfe für Verhandlungsdialekte" in Spektrum der Wissenschaft 03/94 (S. 99 ff.). Heidelberg, 1994.
7. Hess, Prof. Dr. M.: "Einführung in die Computerlinguistik (I)". Vorlesungsscript, Universität Zürich. Zürich, 2003.
8. Klamer, M.: "Textgenerierung". Studienarbeit im Fach Informatik, Universität Erlangen. Erlangen, 1996.
9. Verbmobil. <http://verbmobil.dfki.de/overview-de.html> (23.01.2004)
10. Bateman, J.: <http://www.fb10.uni-bremen.de/anglistik/langpro/webSPACE/fb/info-pages/nlg/> (23.01.2004)

Quellen / Literaturhinweise (2)

1. Traunmüller, H.: "Geschichte der Sprachsynthese". <http://www.ling.su.se/staff/hartmut/kempln.htm> (21.01.2004)
2. Switek, S.: "Sprachsynthese". http://www.spinfo.uni-koeln.de/lehre/CLGrundlagen/Referate/sadeghian_switek/folie_002.htm (21.01.2004)
3. "Computer und Linguistik". http://www.spr-fh-koeln.de/C1-MT/cl_intro_de.html (11.12.2003)
4. Antrup, J. W.: "Aspekte der Computerlinguistik". <http://www.ifi.unizh.ch/groups/CL/CLBuch/KAPIT12.PDF> (21.01.2004)
5. Klankert, T.: "Hybride Unit Selection für ein Sprachsynthesystem". Diplomarbeit, Universität Stuttgart, Stuttgart, 2003.
6. Tchchieva, S., Kojobergenova, M. und Sáiz V.: "Einführung in die maschinelle Erzeugung gesprochener Sprache". http://www.cl.uni-heidelberg.de/kurs/ws03/ecl/referate/r_s12.pdf (22.01.2003)
- 7.

Sprachgenerierung / NLG

Zusammenfassung

Ende

Danke für die Aufmerksamkeit.

