

*PS Einführung in die Computerlinguistik
bzw.
SE aus Artificial Intelligence*

Anwendungen der Computerlinguistik

MMag. Gudrun Kellner



Überblick

- Parsing/Tagging
- Korrekturprogramme
- Computergestützte Lexikographie
- Volltextsuche und Text Mining
- Sprachsynthese
- Spracherkennung
- Dialogsysteme
- Maschinelle und computergestützte Übersetzung
- Elektronische Kommunikationshilfen

Parsing/Tagging

- Analyse sprachlicher Ausdrücke auf ihre grammatische Struktur

Der Hund bellt.

• $\langle \text{Der Hund bellt, } S_{[NP_{[DET[Der]_N[Hund]}]VP_{[V[bellt]]}]} \rangle$

- Analysestrategien:
 - Top-down-Verarbeitung (Grobstruktur vor Details)
 - Bottom-up-Verarbeitung (Wortanalyse vor Gesamtstruktur)
 - Wahrscheinlichkeitshypothesen
- Üblicherweise gekoppelt mit semantischen, pragmatischen und kontextuellen Analysekomponenten

Korrekturprogramme

- 3 Arten von Korrekturprogrammen:
 - Korrektur von „Nicht-Wörtern“
 - Kontextabhängige Korrektur
 - Grammatikkorrektur

- Methoden:
 - Listen von häufigen Fehlern, Verwechslungsmengen
 - Suche nach dem ähnlichsten String (Fehlertypen: Auslassung, Einfügung, Änderung und Transposition)
 - Parser-Einsatz
 - Wahrscheinlichstes intendiertes Wort (Worthypothesengrafen)

- Anwendungen: Textkorrektur, automatische Korrekturen von OCR- oder Audioanalyse-Ergebnissen

Computergestützte Lexikographie

- Lexikalische Akquisition:
 - Beschaffung von Wort-/Textmaterial

- Repräsentation linguistischer Daten:
 - Auswahl, Klassifikation und Kodierung von linguistischen Informationen
 - abhängig von dem Einsatz-/Anwendungsgebiet und der gewählten linguistischen Theorie
 - Format: Datenbank, Textdatei, Datenstrukturen von computerlinguistischen Formalismen (Merkmalstrukturen)

- Computational Terminology:
 - Extraktion von Fachwortschatz aus Texten
 - Identifikation von Äquivalentkandidaten in Paralleltexten

Volltextsuche und Text Mining

■ Volltextsuche

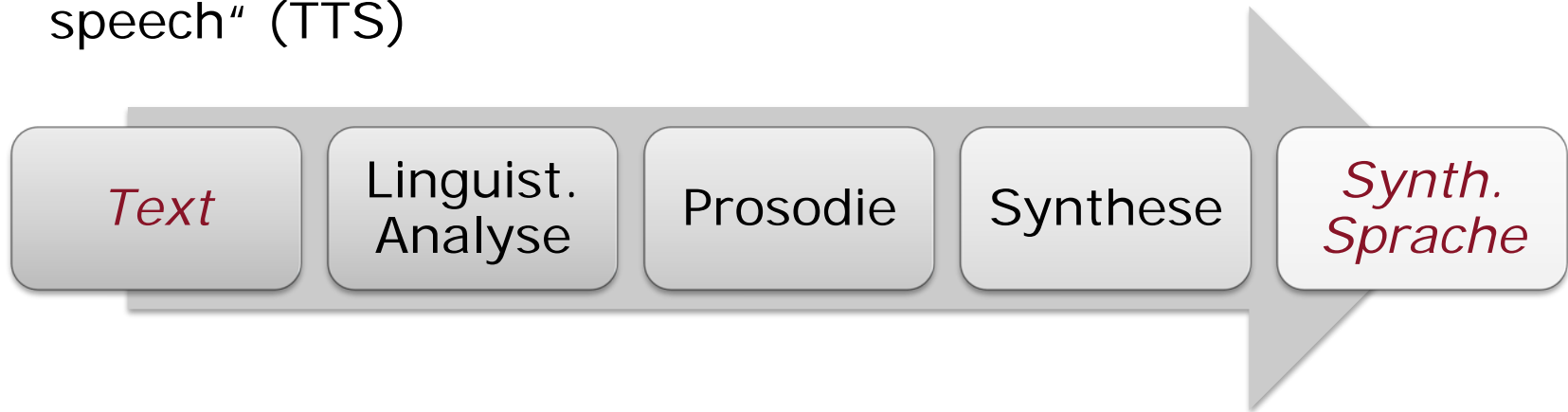
- Komplettindizierung statt Schlagwortverzeichnis (Printmedium)
- Umgang mit sehr großen Datenmengen
- Häufig: Normalisierung der Wörter auf Stammformen, Ausfiltern von „Stoppwörtern“ wie Artikel, Konjunktionen etc.
- Optionen: unscharfe, phonetische, Phrasen-, Feldersuche

■ Text Mining

- Pendant zum „Data Mining“ (Informationsextraktion aus Datenbanken) für Textdokumente
- Problem: Informationen sind nicht strukturiert
- Textklassifikation
- Schlüsselwort-/Informationsextraktion
- Automatische Textzusammenfassung

Sprachsynthese

- Automatische akustische Ausgabe von Texten: „text to speech“ (TTS)



- Linguistische Analyse: lexikalisch, morphologisch, syntaktisch
- Prosodie: Bestimmung des Satzmodus, Festlegung von Silben-, Wort-, Phrasen- und Satzakzenten
- Synthese: Vollformenwörterbuch vs. Ausspracheregeln

Spracherkennung



- Signalanalyse: Vorverarbeitung, Merkmalsextraktion
- Unit Matching: Erkennen von akustischen „Minimalblöcken“
- Ausschlussverfahren in mehreren Schritten:
 - Lexikalische Dekodierung: Zuordnung von Lautfolgen zu akustischen Wörterbucheinträgen (basierend auf den Units)
 - Syntaktische Analyse: mögliche Wortkombinationen werden auf syntaktische Korrektheit geprüft
 - Semantische Analyse: mögliche syntaktisch richtige Kombinationen werden auf semantische Korrektheit geprüft

Dialogsysteme

- Kommunikation zwischen Benutzer und Maschine
 - Sprachliche Ein- und Ausgabe
 - Textbasiert oder akustisch

Ablauf:

- Interpretation natürlicher Sprache
- Aktionsplanung
 - Aufruf entsprechender Programme/Funktionen
 - Anfrage an internes Datenbank-System
- Ausgabeplanung
 - Formulieren der entsprechenden Antwort und/oder
 - Durchführen der gewünschten Aktion

Übersetzungsprogramme

- Unterscheidungen:
 - maschinell/computergestützt
 - bi-/multilingual
 - uni-/bidirektional

- Ablauf:
 - Analyse des eingegebenen Texts in der Quellsprache
 - Umwandlung in abstrakte Repräsentationen der Inhalte
 - Generierung der Aussage in der Zielsprache

- Schwierigkeiten:
 - Lexikalische Lücken (gaps)
 - Nichtentsprechungen (mismatches)
 - Divergenzen (Unterschiede im Ausdruck einer Tatsache)

Elektron. Kommunikationshilfen

- Ziel: Ausgleich v. eingeschränkter Kommunikationsfähigkeit

- Anwendungen:
 - Zeigetafeln mit Symbolen, Bildern, Text
 - Kommunikationshilfen mit Sprach- oder Textausgabe
 - Texttelefone für Hör- oder Sprechgeschädigte

- Methoden:
 - Textkomplettierung
 - Wortvorschlagessystem
 - Textretrieval (Speichern v. vorgefertigten Äußerungen in thematischen Hierarchien)
 - Verwendung von Symbolsystemen