

Maschinelle und computergestützte Übersetzung

Computerlinguistik SS 2010

12.07.2010

Patricia Naumann

Yuanfeng Zhang

Ziele und Herausforderungen

- Text soll automatisch aus einer Quellsprache in eine Zielsprache übertragen werden
- Dabei soll die Funktion des Textes erhalten bleiben
- Es werden sowohl computerlinguistische Analysen als auch eine hochwertige Textgenerierung benötigt

□ *Maschinelle Übersetzung*

Systemarten

- Bilingual vs. Multilingual
- Unidirektional vs. bidirektional
- Batch-Systeme vs. Interaktive Systeme

□ *Maschinelle Übersetzung* Übersetzungsstrategien

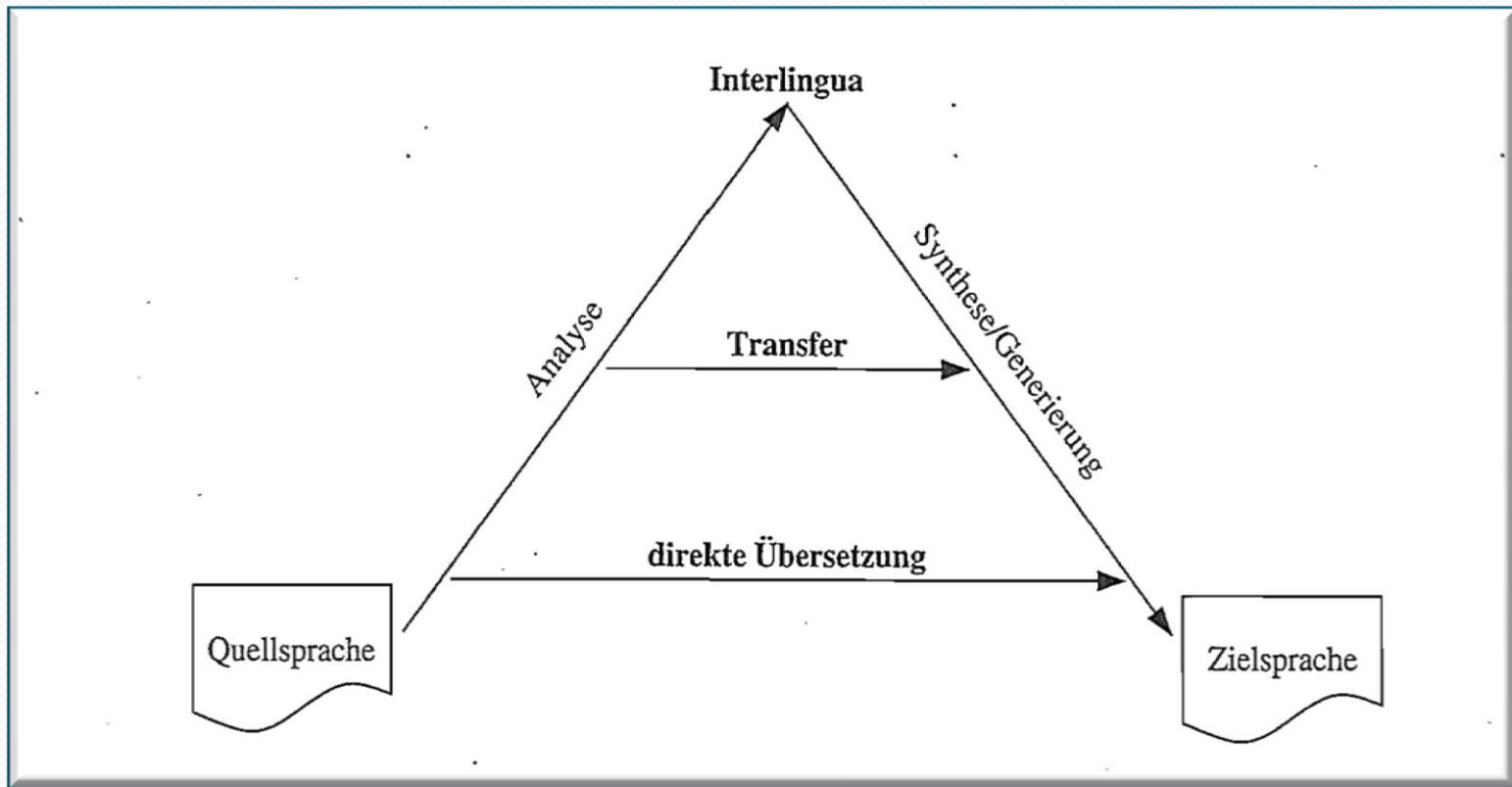
- Formen direkter Übersetzung
 - Beispielbasierte Übersetzung
 - Statistische Übersetzung

Und regelbasierte Ansätze:

- Transfer
- Interlingua

□ Maschinelle Übersetzung

Übersetzungsstrategien



Direkte Übersetzung

- Oberflächliche (morphologische) Analyse
- Wort-für-Wort-Übersetzung mit Wörterbuch
- Einfache Wortstellungskorrekturen

+ Sehr schnelle, aber

— schlechte Übersetzung

□ *Direkte Übersetzung*

Beispielbasierte Übersetzung

- Zuordnung von Übersetzungseinheiten aus einem bilingualen Korpus
 - Analogiebildung in neuen Kontext aus Daten der Translation Memory
- + Kommt ohne Regeln und Wissen über die Sprachen aus
- Schlechtere Ergebnisse

□ *Direkte Übersetzung*

Statistische Übersetzung

- Modell für Übersetzung aus bilingualem alignierten Korpus
- Häufigkeit und Nähe der zu übersetzenden Wörter
- Wahrscheinliche Übersetzungen auf Satzebene, daraus zu Phrasen und Wörtern

- Auch hier keine Regeln benötigt

Transfer

1. Text in Quellsprache geparst und semantisch analysiert
 2. Abstrakte Repräsentation
 3. Abstrakte zielsprachliche Repräsentation durch Transferregeln
 4. Natürlichsprachliche Ausgabe
- + Keine vollständige Bedeutungsanalyse, einfache Phasen
 - Wortwahl begrenzt, nur für eingearbeitete Sprachen

Interlingua

- Analyse der Eingabe erzeugt sprachunabhängige Zwischenrepräsentation
 - Bisher nicht komplett
 - Auf Domänen und spezifische Grammatiken beschränkt
-
- + Schnelle Integration neuer Sprachen ins System
 - Bisher keine neutrale Repräsentation für alle Sprachen gefunden

□ *Interlingua*

Wissensbasierte Übersetzung

- System kann auf Weltwissen zurück greifen
 - Spezialisiertes Wissen auf Domänen
 - Vermeidet Übersetzungsmehrdeutigkeiten
- + Bessere Übersetzungsergebnisse, besonders bei Ambiguität
- Nicht universell anwendbar

□ *Maschinelle Übersetzung* Anwendung

- Für Rohübersetzungen genutzt, die schneller nachbearbeitet werden können als die Erstellung einer kompletten Übersetzung
- Wird kontrollierte Sprache eingesetzt (begrenzte Domäne, Grammatik, Lexikon; Eindeutige Sätze), können schnell mehrere übersetzte Texte angefertigt werden

□ *Maschinelle Übersetzung* Evaluation

- Menschliche Beurteilung
- Algorithmen
 - BLEU (Durchschnitt des Niveaus der Sätze)
 - NIST (komplexere Übersetzungen werden höher gewertet)
 - WER / PER (Zahl falsch übersetzter Wörter / ohne Beachtung der Wortreihenfolge)
 - METEOR (beachtet Synonyme und Lemmata)

Computergestützte Übersetzung

- Hilfe für professionelle Übersetzer
- Musterabgleich mit der Datenbank gibt Übersetzungsvorschlag
- Könnten maschinelle Übersetzung verbessern

- Translation Memories
- Terminologie-Datenbanken
- Alignment-Tools zum Speicher übersetzter Texte

□ CAT

Translation Memories

- Zweisprachige Textsegmente (Absatz, Satz)
 1. Paare von Referenztexten (Kontext bleibt erhalten)
 2. Datenbankorientierte Systeme (Übersetzungseinheiten in Paaren, ohne Kontext)
- Gefundene Segmente können automatisch übersetzt werden

□ *CAT*

Terminologiedatenbanken

- Mehrsprachige Glossare, Thesauri
 - Termdefinitionen
 - Linguistische Analyse
 - Konkordanzen
 - Entsprechung
- Einheitliche Verwendung der Terminologie
- Leichter Zugang zum Inhalt

□ *Demonstration*

Englisch → Japanisch

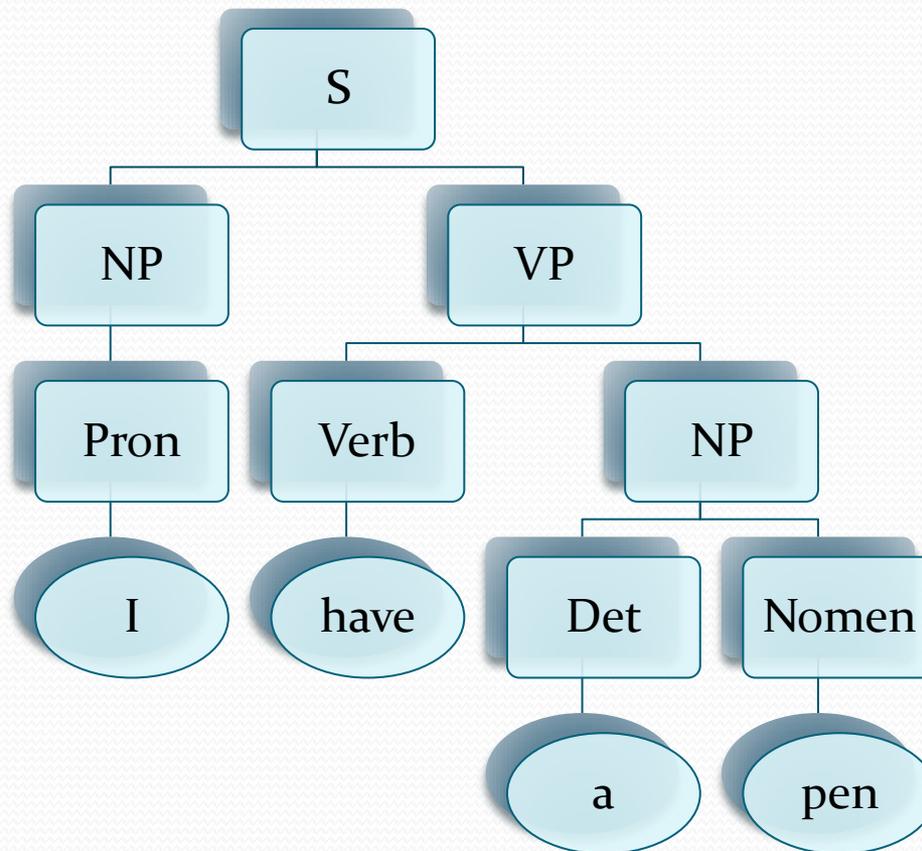
I have a pen.

watashi wa pen o motteiru.

□ *Demonstration*

Englisch → Japanisch

- Syntaktische Analyse



□ *Demonstration*

Englisch → Japanisch

- Ersetzung der Satzeinheiten mittels des Lexikons

Englisch	Japanisch
I	watashi
have	motteiru
a	-
pen	pen

- Konvertierte Satzeinheiten

(S (NP (pron **WATASHI**)) (VP (verb **MOTTEIRU**) (NP (det -) (noun **PEN**))))))

□ *Demonstration*

Englisch → Japanisch

- Rekonstruierung der Syntaxstruktur und Einfügung der Partikel

Englisch	Japanisch
S → NP VP	S → NP WA VP
VP → Verb NP	VP → NP O Verb

- Konvertierter Syntaxbaum

(S (NP (pron **WATASHI**)) **WA** (VP (NP (det -) (noun **PEN**)) **O** (verb **MOTTEIRU**)))

→ watashi wa pen o motteiru.

□ *Demonstration*

Englisch → Japanisch

- Ergebnisse von den Übersetzungswebseiten

Webseite	Ergebnis	Beurteilung
SYSTRANet	watashi ni pen ga arimasu.	○
WorldLingo	watashi ni pen ga aru.	○
Google	watashi wa pen o motte.	X
excite	watashi wa pen o motteimasu.	○

□ *Demonstration*

Englisch → Japanisch

- Ergebnisse von den Übersetzungswebseiten

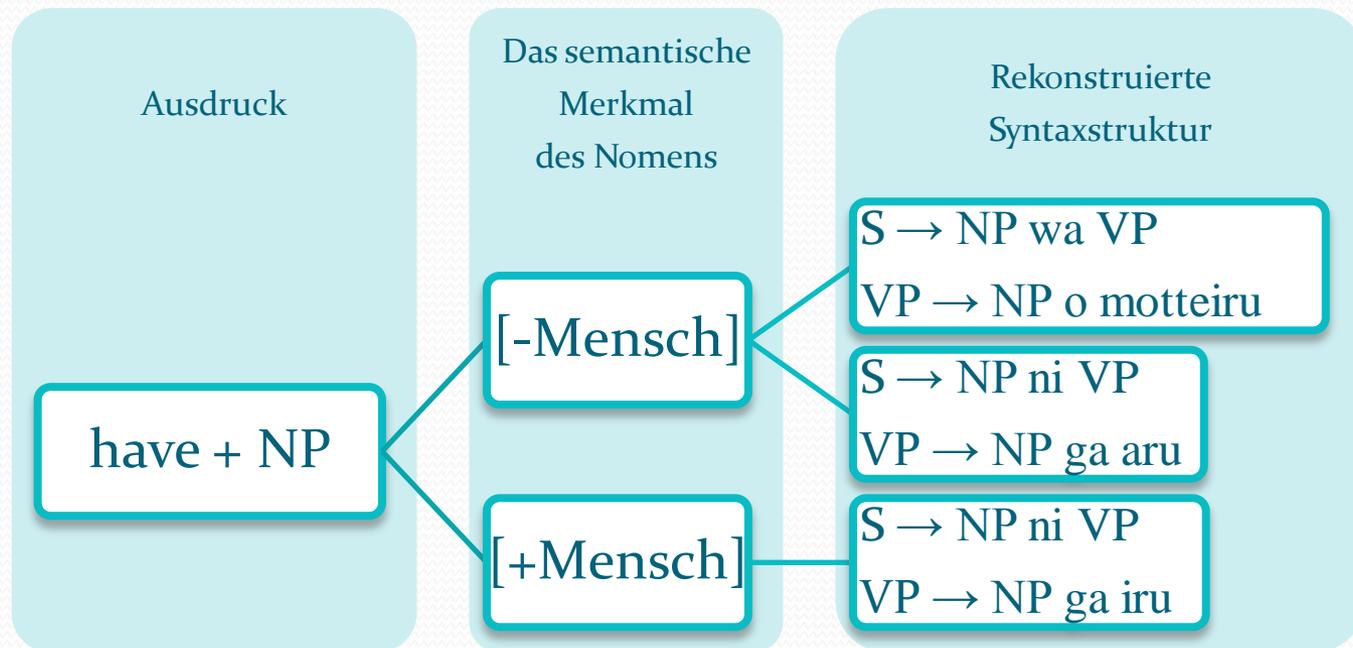
Webseite	Ergibnis	Beurteilung
SYSTRANet	watashi ni pen ga arimasu.	○
WorldLingo	watashi ni pen ga aru.	○
Google	watashi wa pen o motte.	X
excite	watashi wa pen o motteimasu.	○

- Vergleich mit dem Satz „I have a cousin.“

Webseite	Ergibnis	Beurteilung
SYSTRANet	watashi ni itoko ga imasu.	○
WorldLingo	watashi ni itoko ga iru.	○
Google	watashi wa itoko ga arimasu.	X
excite	watashi niwa itoko ga imasu.	○

□ *Demonstration* Englisch → Japanisch

- Semantische Analyse und Syntaktische Verarbeitung



Quellen

- K.-U. Carstensen et al. (2004): Computerlinguistik und Sprachtechnologie. Eine Einführung. Spektrum, Akademischer Verlag, Seite 563 - 571
- <http://www.globalsecurity.org/intell/systems/mt-techniques.htm>
- http://www.translationdirectory.com/articles/article1814.php#Automatic_evaluation
- <http://de.wikipedia.org/wiki/Terminologie-Datenbank>
- http://de.wikipedia.org/wiki/Maschinelle_%C3%9Cbersetzung
- <http://de.wikipedia.org/wiki/%C3%9Cbersetzungsspeicher>
- http://de.wikipedia.org/wiki/Computerunterst%C3%BCtzte_%C3%9Cbersetzung