

Anwendungen der Computerlinguistik

4. Übersicht der Themen

Pius ten Hacken

Block 1

Freitag 14-18	Samstag 9-12	Samstag 14-17
1. Einführung	3. MT	4. Themen
2. NLU-Modell	Diskussion	
Praktische Arbeit		

Themenbereiche für Referate

Systran

Themenbereiche für Referate

Systran

NLU-basierte MT

Themenbereiche für Referate

Systran

NLU-basierte

MT ohne NLU

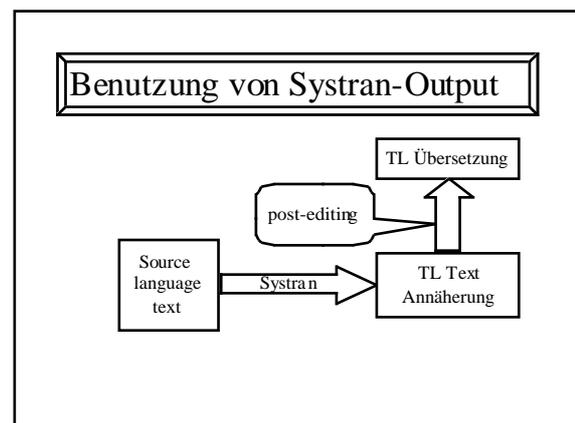
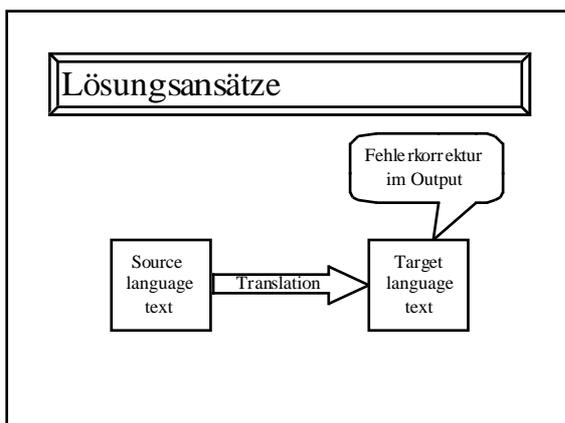
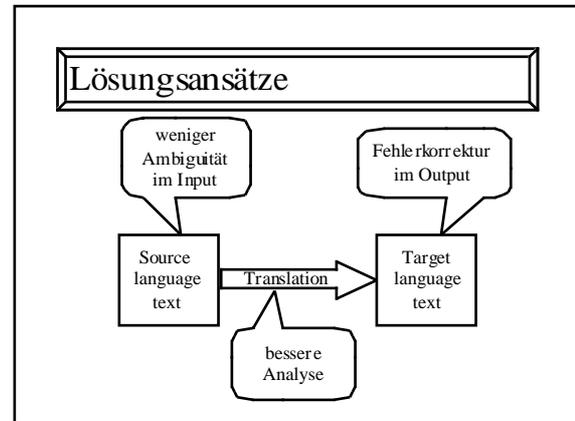
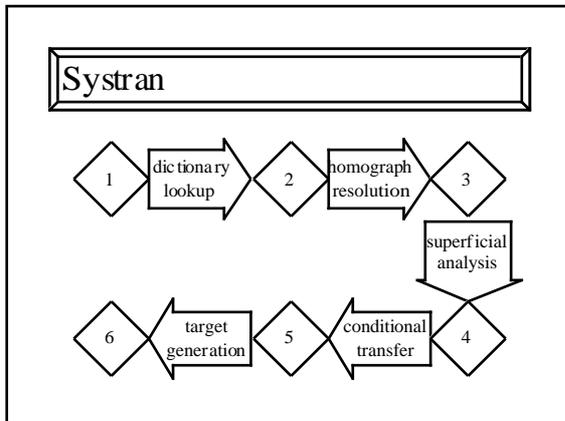
Themenbereiche für Referate

Systran

NLU-basierte

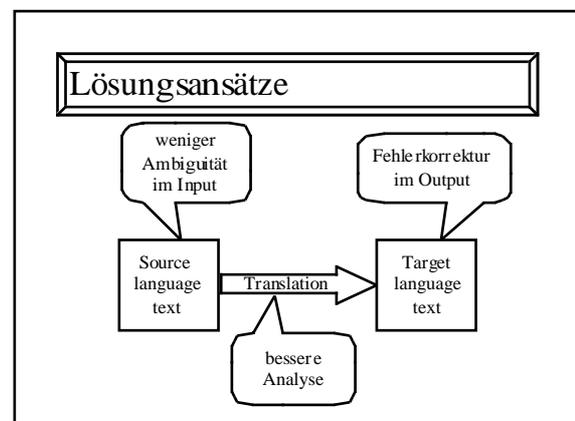
MT ohne NLU

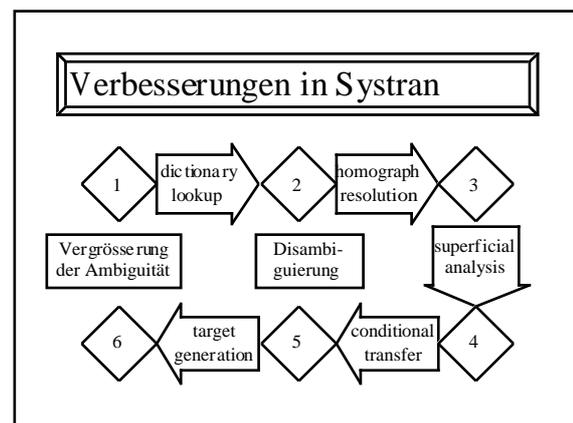
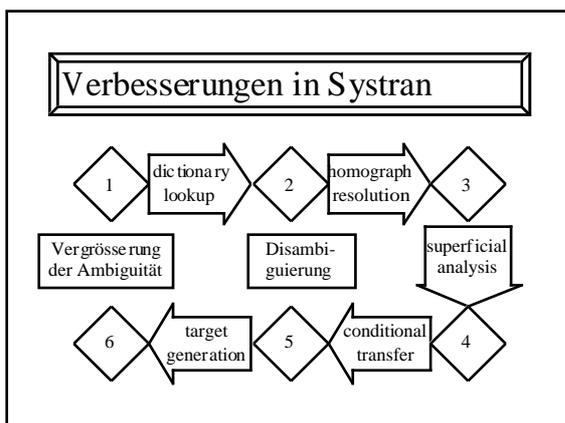
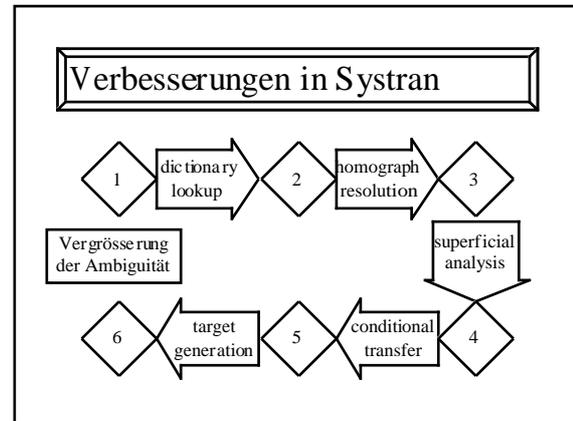
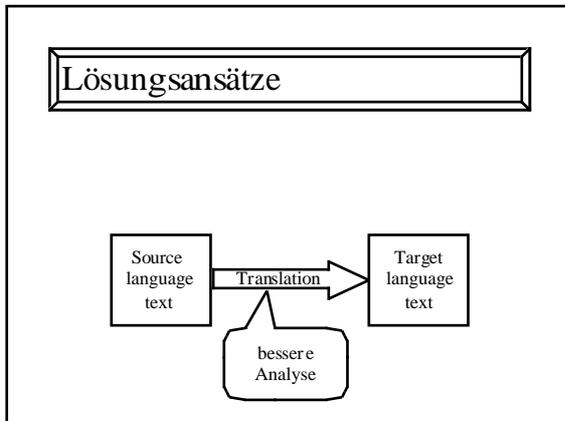
NLU für andere Bereiche als MT



Post-editing

- übliche Praxis in der professionellen Übersetzung
- effizientes Post-Editing von MT Output erfordert:
 - andere Herangehensweise (minimalistisch statt perfektionistisch)
 - Unterstützung durch editing software





- ### Reduktion der Ambiguität
- Lösung der Ambiguität innerhalb der Systran-Architektur:
 - mehr oder bessere Kontexte für homograph resolution
 - mehr oder bessere Kontexte für conditional semantics, Präpositionen, lexical routines
 - Vermeidung der Ambiguität:
 - weniger Homonyme im Lexikon
 - weniger Strukturen in der Analyse

- ### Ambiguität und Kenntnisse
- Je mehr wir über den Input wissen, desto mehr können wir die Ambiguität einschränken.
 - Wissensquellen:
 - allgemeine Eigenschaften der Sprache
 - Generalisierungen über den spezifischen Input, den wir erwarten können
 - Manipulation des Inputs

Generalisierungen

Le succès de cette méthode est évident.

The success of this method is obvious.

succès:	méthode:	évident:
a. success	a. method	obvious
b. bestseller	approach	evident
hit	b. manual	self-evident
c. conquest	tutor	
	primer	

Customization

- Entfernung von irrelevanten Bedeutungen im Lexikon
- Generalisierungen für die verbleibenden Bedeutungen
- Fehleranalyse im Post-Editing

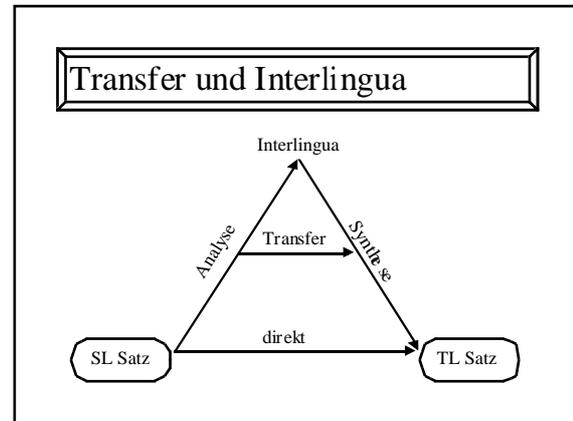
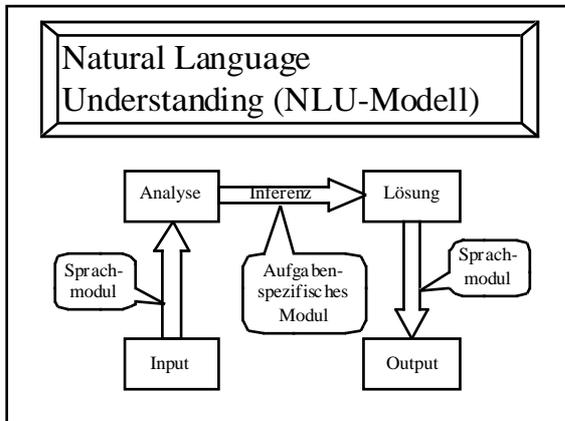
Lösungsansätze

Lösungsansätze

Controlled language

- selbst-erstellte Texte
- beschränkte Domäne
- genügende Textmenge
- Ausschluss von Ambiguität im Input
- Probleme:
 - genügend Ausdrucksmöglichkeiten
 - Unterstützung bei der Einhaltung der Einschränkungen

Themenbereiche für Referate

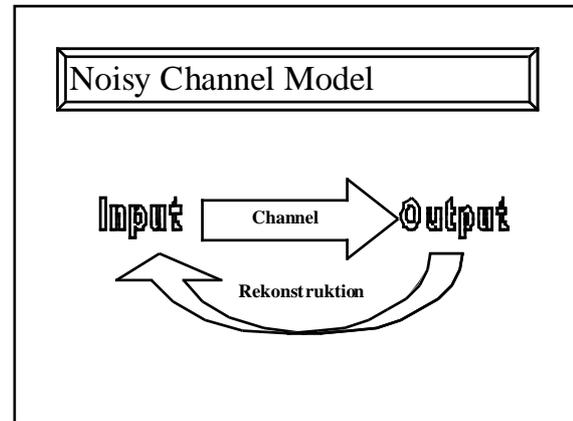
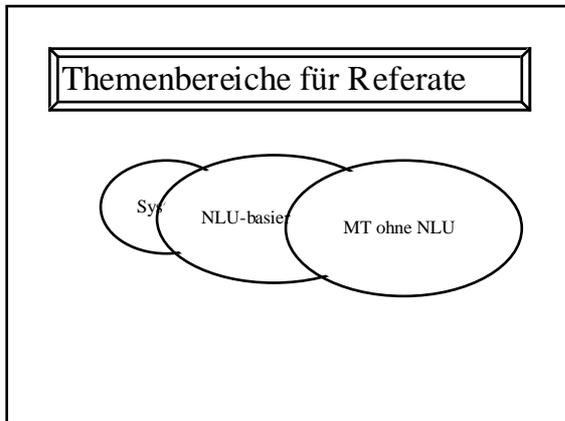


- Eurotra**
- Transfersysteme
 - neun Sprachen, dezentrale Entwicklung
 - Minimierung des Transfers durch Angleichung der IS
 - linguistisch informierte Analyse
 - eklektische Behandlung der Theorie
 - keine absolute Vereinheitlichung des Analyseprozesses

- Rosetta**
- Interlingua-System
 - drei Sprachen, zentrale Entwicklung
 - Gesamtarchitektur basiert auf Logik:
 - Kompositionalität
 - "Menge der möglichen Übersetzungen"
 - eklektische Einpassung von Analysen einzelner Phänomene

- Unitran**
- Transfersysteme
 - one-woman project
 - Linguistisch basierte Analyse
 - Syntax gemäss Chomsky
 - Semantik gemäss Jackendoff
 - Klassifikation und Lösung von Divergenzen in der Übersetzung

- Knowledge-Based MT**
- Interlingua-Ansatz
 - Künstliche Intelligenz als Basis
 - Sprachtheorie als Hilfsmittel
 - Repräsentation einer Domäne mit sprachunabhängigen Konzepten

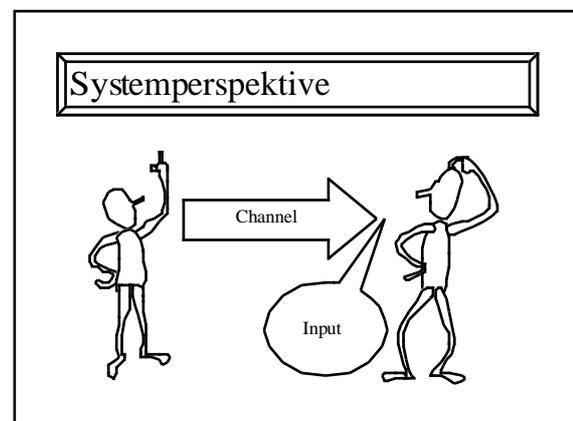
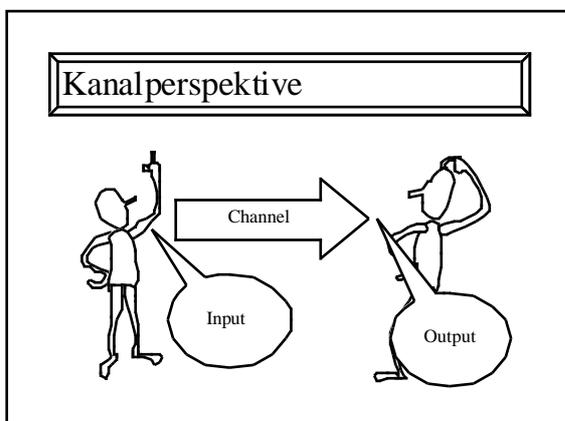


Information Theory

- Claude Shannon,
 - Bell Laboratories, 1948
- Information als Problem für Engineering
- Rekonstruktion auf der Basis von probabilistischen Verfahren

Form und Bedeutung

- "Frequently the messages have *meaning* that is they refer to or are correlated according to some system with certain physical or conceptual entities.
- These semantic aspects of communication are irrelevant to the engineering problem." [Shannon (1948)]



Rekonstruktion

- Der Output des Kanals wird bestimmt von:
 - Eigenschaften des Inputs
 - Eigenschaften des Kanals

Tagging im Noisy Channel Model

MT im Noisy Channel Model

Schritte in der Entwicklung

- Corpußsammlung
- Alignment
- Evaluation der Corpora
 - Korrespondenz der Wörter aus SL mit TL
 - Abfolge der Wörter in der TL
- Maximierungsalgorithmus

Example-Based MT

- Alternative zur Anwendung von Bayes's Theorem
- Datenbank von Beispielen mit korrekten Übersetzungen
- Algorithmus zur Erkennung der richtigen Beispiele

Themenbereiche für Referate

