

Anwendungen der Computerlinguistik

3. Maschinelle Übersetzung

Pius ten Hacken

Block 1

Freitag 14-18

1. Einführung

2. NLU-Modell

Praktische
Arbeit

Samstag 9-12

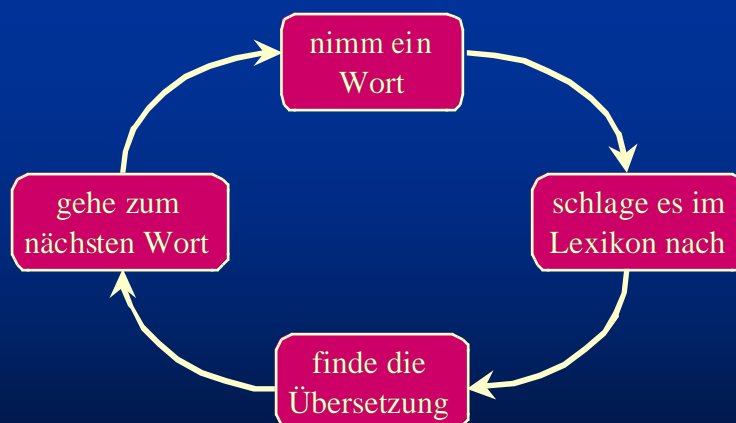
3. MT

Samstag 14-17

Systran

- 1960: erste Vorführung von GAT an der Georgetown University
- 1964: Entwicklung einer Version von Systran, Universität des Saarlandes
- 1968: Gründung der Firma Systran von Peter Toma
- 1975: Kontrakt mit der Commission of the European Communities (CEC)

Wort für Wort Ersetzung



Beispiel: Wort für Wort

Le succès de cette méthode est évident.



The success of this method is obvious.

Beispiel: Wort für Wort

Cette méthode primitive ne fonctionne pas.



This primitive method does not work.

Korrespondenz der Wörter

Cette méthode primitive ne fonctionne pas.

This primitive method does not work.

Problem: Ambiguität

Le succès de cette méthode est évident.

The success of this method is obvious.

the	le
	la
this	ce
	cette
obvious	évident
	évidente
	évidents
	évidentes

Problem: Ambiguität

Le succès de cette méthode est évident.



The success of this method is obvious.

succès:

- a. success
- b. bestseller
hit
- c. conquest

méthode:

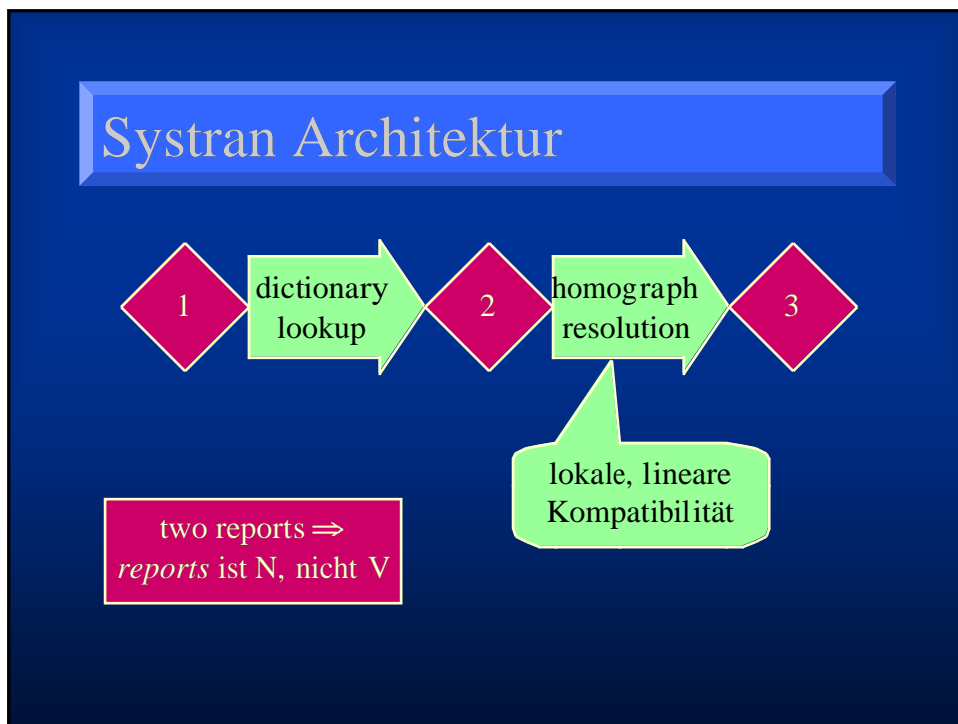
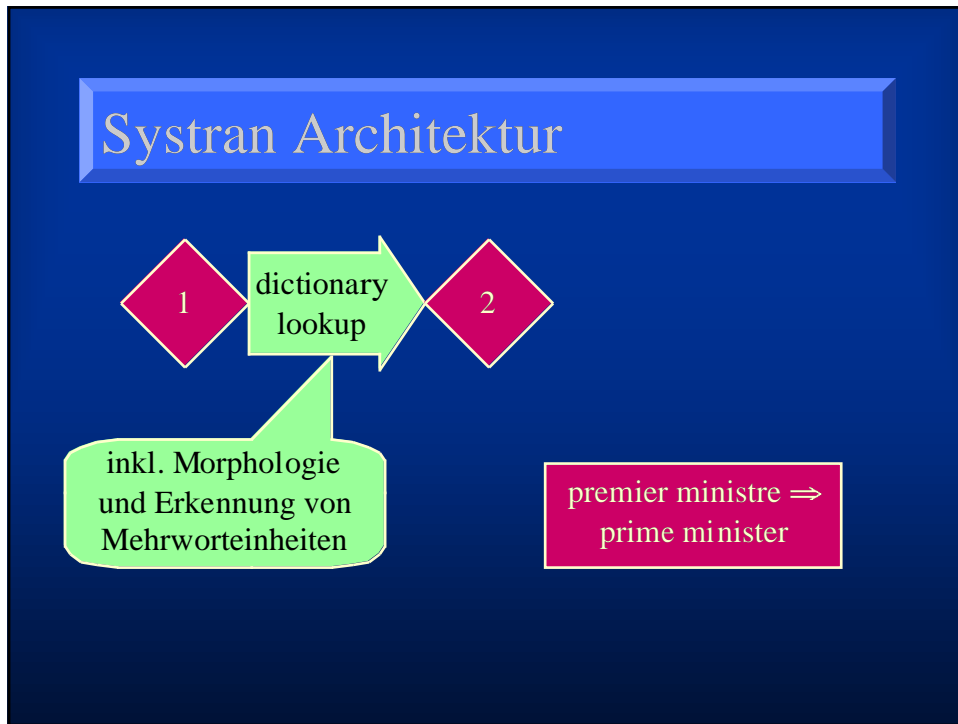
- a. method
approach
- b. manual
tutor
primer

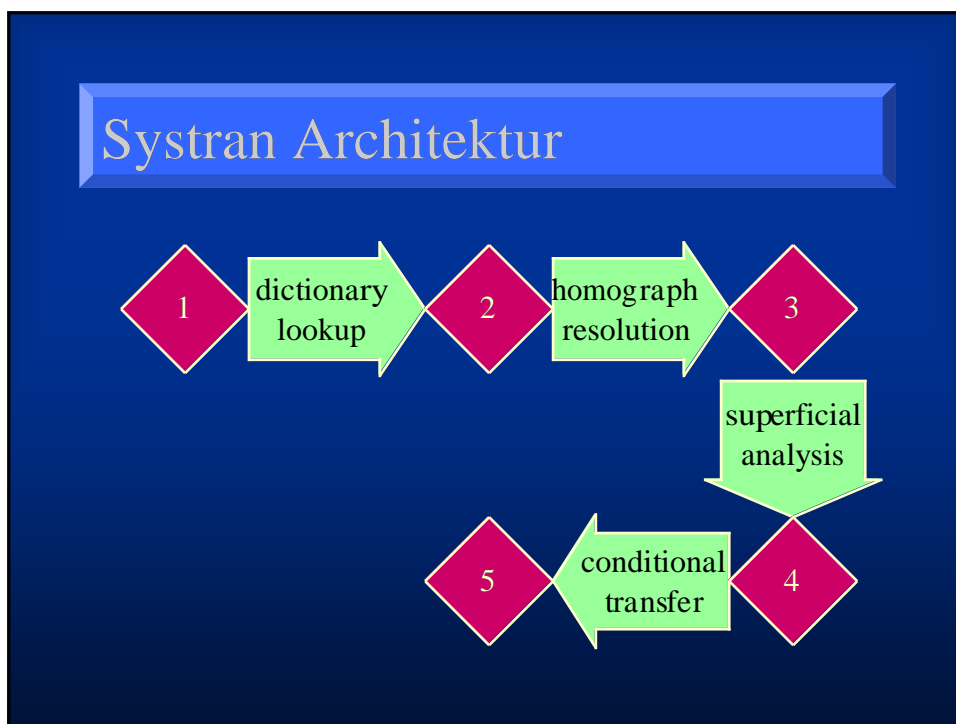
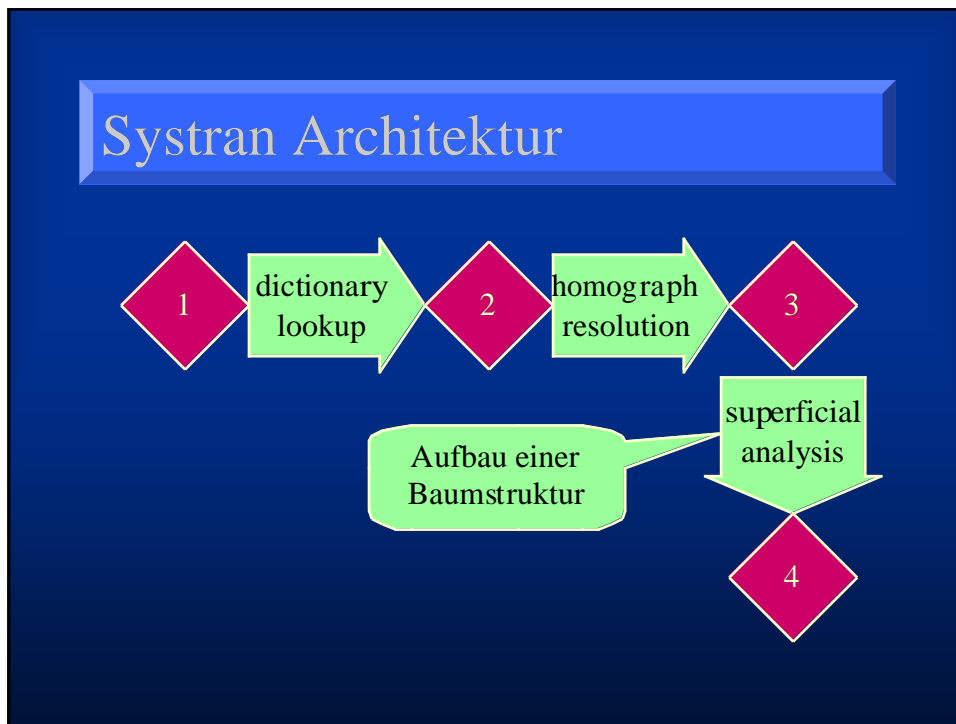
évident:

- obvious
- evident
- self-evident

Zusammenfassung: Probleme

- **Fehlende Korrespondenz zwischen Wörtern**
 - Reihenfolge der Wörter, z.B. N+A, A+N
 - Mehrworteinheiten, z.B. Negation
- **Fehlende Korrespondenz zwischen Wortformen**
 - Morphologische Ambiguität, z.B. Masc.
 - Lexikalische Ambiguität

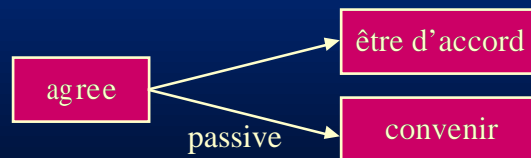




Übersetzung

- Inhaltswörter

- feste Ausdrücke (*in order to*) wurden schon erkannt (und übersetzt)
- Komposita (*washing machine*) wurden schon erkannt (und übersetzt)
- Conditional Semantics Lexikon:



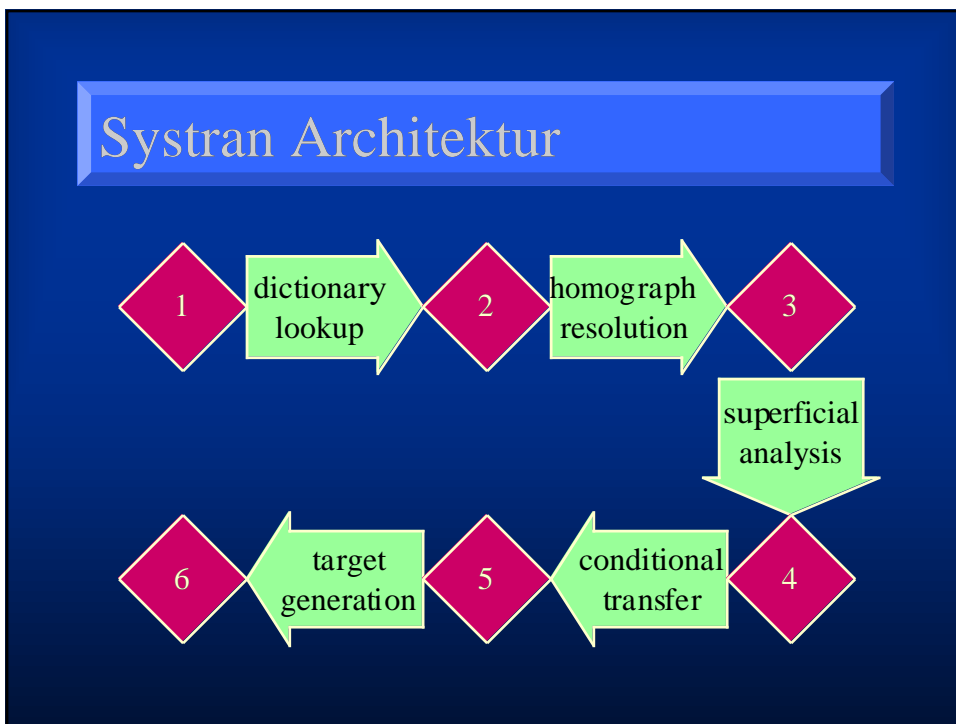
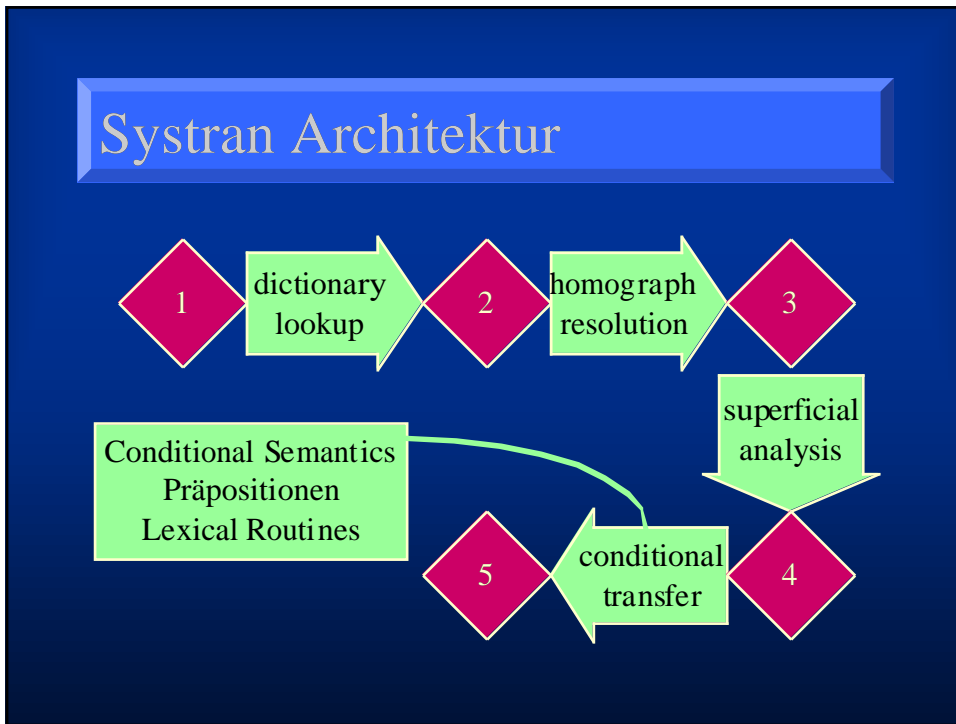
Übersetzung

- Übersetzung der (verbleibenden) Präpositionen

- (*rely*) *on* ⇒ (*compter*) *sur*

- Lexical routines

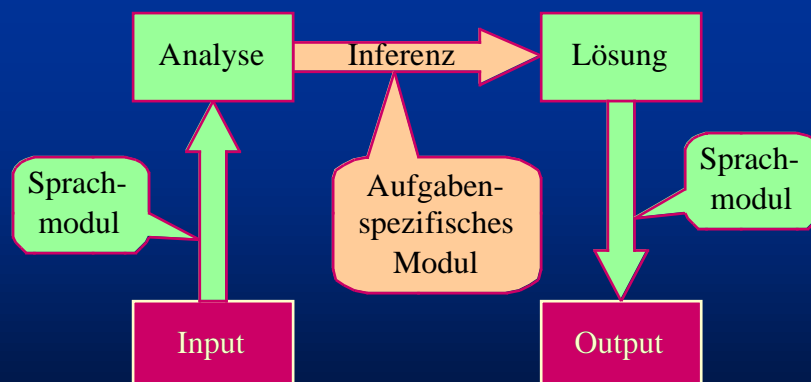
- Umformulierung wegen eines spezifischen Wortes
 - *expects to V [inf]* ⇒ *s'attend à ce qu'il/elle V[subj]*

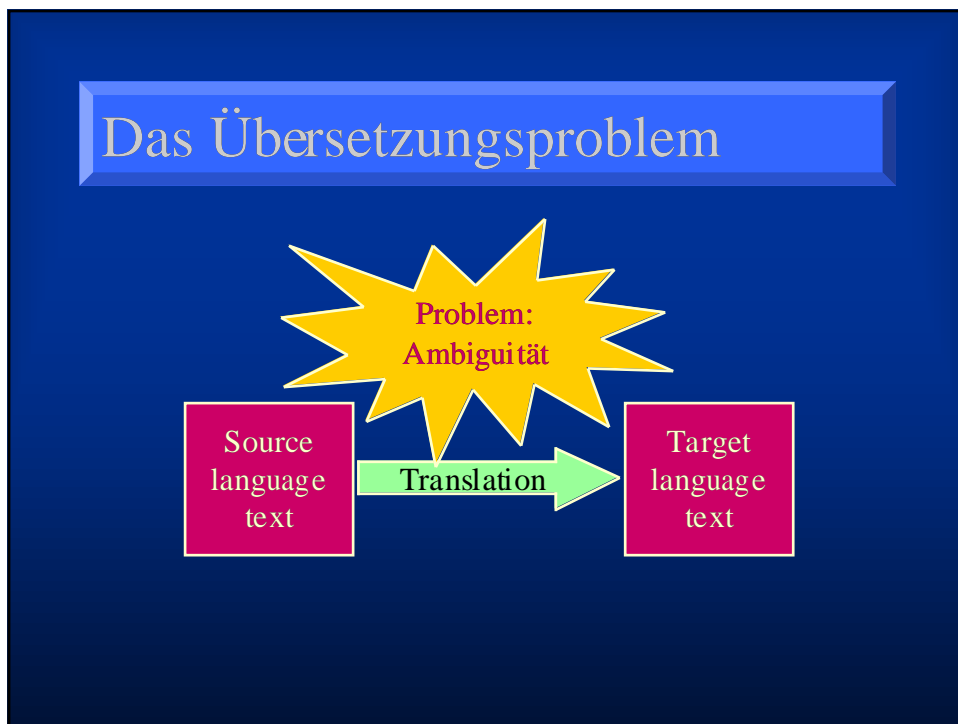
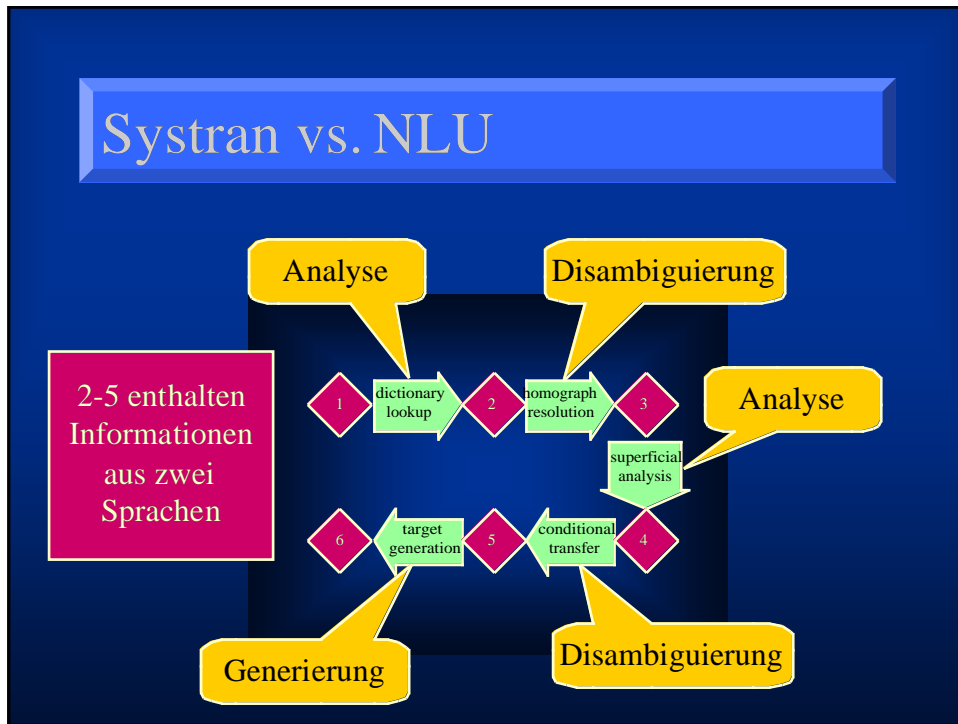


Generierung

- Übersetzung der Wörter
 - Festlegung der Übersetzung für die verbleibenden Wörter
- Morphologische Generierung
 - *spiel* + 3sg präs indic \Rightarrow *spielt*
- Allgemeine Umstellungen
 - Wortfolge: Fr. N+A zu Eng. A+N
 - Orthographie: *a* oder *an* im Englischen

Natural Language Understanding (NLU-Modell)





Problem: Ambiguität

- Phänomen: Eine Form hat zwei oder mehr Analysen
- Ebene der Ambiguität: morphologisch, syntaktisch, semantisch, pragmatisch
- Lösung: mehr Kontext, tiefere Analyse

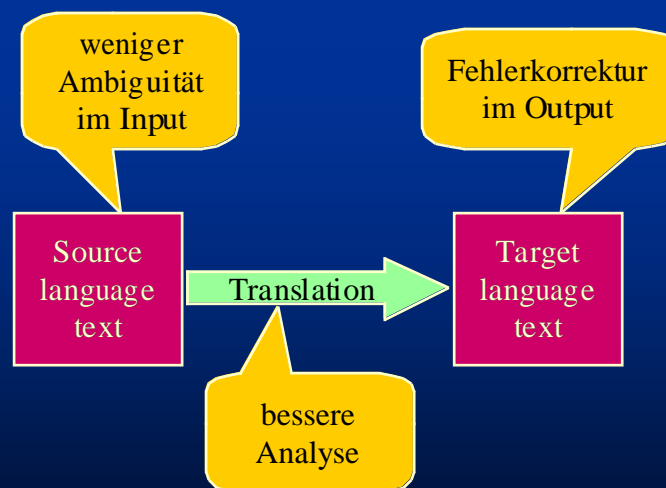
Morphologische Ambiguität

- Phänomen: Eine Form hat zwei oder mehr morphologische Analysen
- Beispiele: *Wogen, Macht*
- Lösung: Berücksichtigung des syntaktischen Kontexts

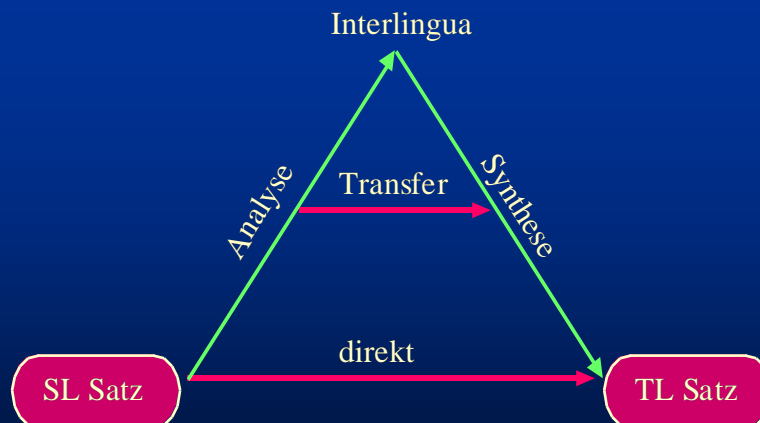
Syntaktische Ambiguität

- Phänomen: Ein Satz hat zwei oder mehr syntaktische Analysen
- Beispiele:
 - *Er sah die Frau mit den blonden Haaren*
 - *Er sah die Frau mit seinen verschlafenen Augen*
- Lösung: Berücksichtigung der semantischen Bedingungen

Lösungsansätze



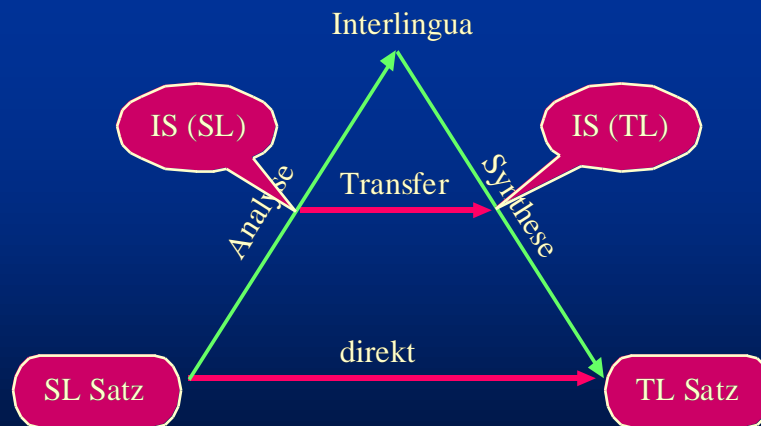
Transfer und Interlingua



Zwischenrepräsentationen

- **Interlingua (IL):**
 - neutrale Darstellung, ohne direkter Bezug zu einer Sprache
- **Interface-Struktur (IS):**
 - Endprodukt der Analyse
 - Anfangspunkt der Synthese
 - sprachspezifisch

Transfer und Interlingua



Fazit (1)

- Systran ist nicht auf die NLU-Methode basiert.
- Verbesserung der Übersetzungsqualität kann erreicht werden durch:
 - Einschränkung des Inputs
 - Ausschluss bestimmter Möglichkeiten in der Analyse
 - Korrektur der Fehler im Output

Fazit (2)

- Die NLU-Methode sieht für die maschinelle Übersetzung ein Transfer-Modell vor.
 - Die allgemeine Analyse produziert eine TL- und Aufgaben-unabhängige IS.
 - Die Inferenz besteht aus Übersetzungs- und Sprachpaar-spezifische Disambiguierung.

Block 1

