



# Maschinelle Uebersetzung: UNITRAN

Ein linguistischer Ansatz.

25. / 26.6.2004

Brigitta Ingold

8/13/04      Blockseminar Computerlinguistik 2004: UNITRAN      1



# Inhalt

- UNITRAN in der MT-Landschaft
  - Ziele
  - Ergebnisse
- Architektur
  - Interlingua
  - Parametrisierung
  - Komponenten
- Divergenzproblem
  - Syntax / Lexikon
  - 7 Kategorien lexikalisch-semantischer Divergenzen

8/13/04      Blockseminar Computerlinguistik 2004: UNITRAN      2

## UNITRAN in der MT-Landschaft

- **Bonnie Jean Dorr (University of Maryland)**
  - Machine Translation: A View from the Lexicon, 1993
- **Linguistisch basierte Analyse:**
  - Syntax (Chomsky)
  - Prinzipien- und Parameterbasiert (PBMT)
  - Interlingua (Jackendoff: LCS)
  - Fokus auf Lexikon und Semantik (LBMT)
- **UNIversal TRANslator**

8/13/04 Blockseminar Computerlinguistik 2004: UNITRAN 3

## UNITRAN: Ziele

- **Allgemeine Lösung für ein spezifisches Unterproblem: sprachübergreifende Übersetzungsdivergenzen erforschen und klassifizieren**
- **Erproben verschiedener linguistischer Ansätze (Interlingua, Parametrisierung, Kompositionalität) in einer Anwendung**

8/13/04 Blockseminar Computerlinguistik 2004: UNITRAN 4

## UNITRAN: Ergebnisse

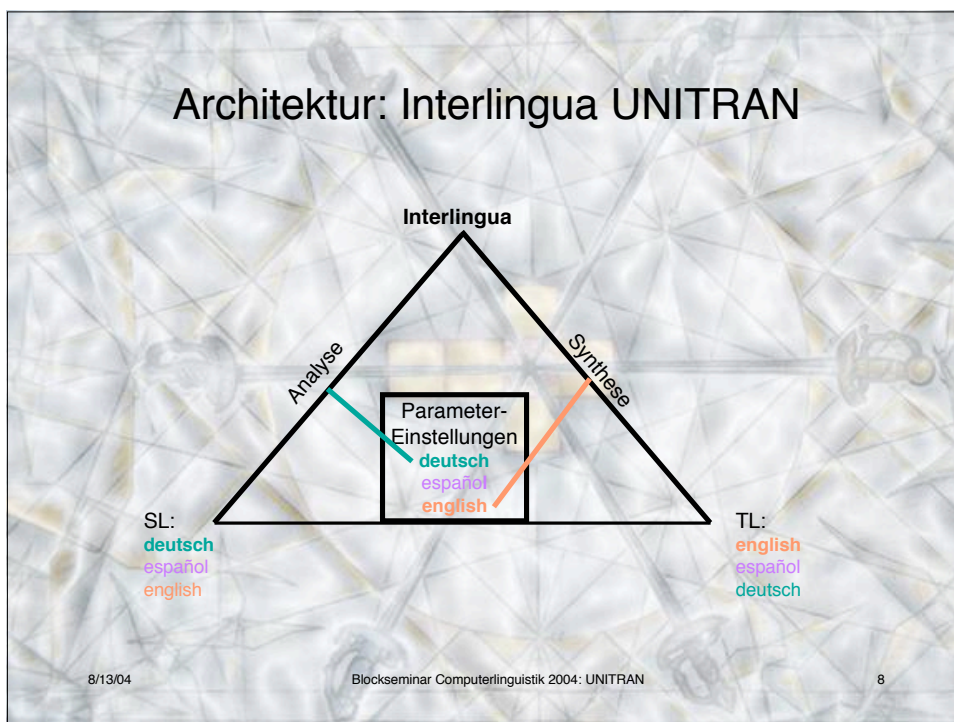
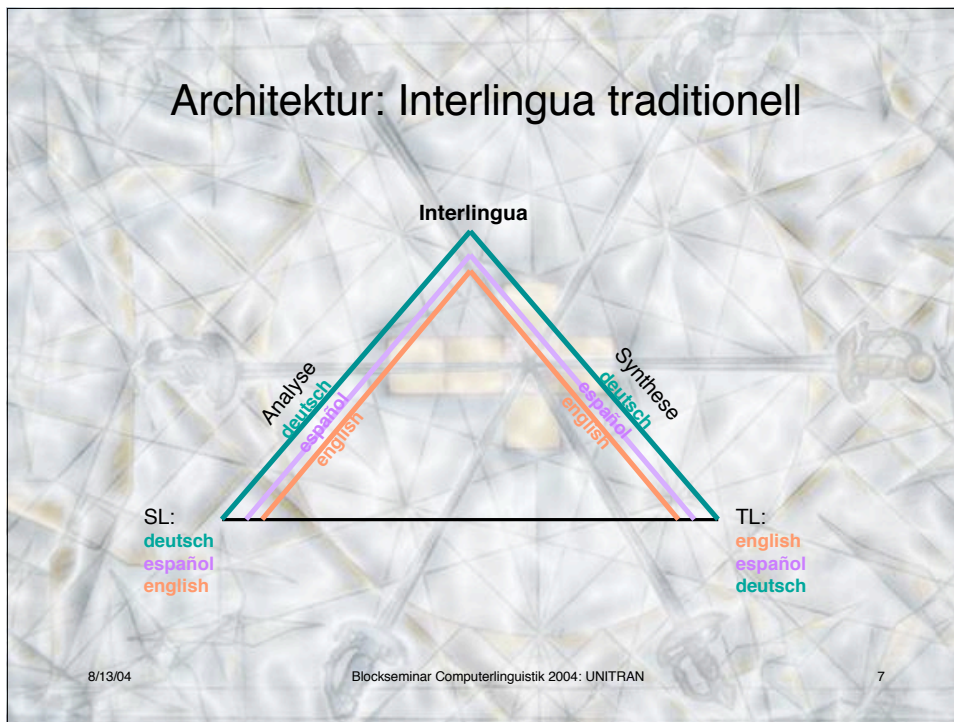
- **Praktisch**
  - Übersetzungssystem (Englisch, Spanisch, Deutsch), das als Testumgebung für linguistische Theorien dient
- **Linguistik**
  - Notation für Divergenzen
  - Lösungsansatz für das Divergenzproblem
- **Künstliche Intelligenz (AI)**
  - Neue Prozessgestaltung: Parser und Generator arbeiten als Tandem mit linguistischen Einschränkungen
  - Formalisierung der für MT notwendigen Operationen

8/13/04 Blockseminar Computerlinguistik 2004: UNITRAN 5

## Inhalt

- **UNITRAN in der MT-Landschaft**
  - Ziele
  - Ergebnisse
- **Architektur**
  - Interlingua
  - Parametrisierung
  - Komponenten
- **Divergenzproblem**
  - Syntax / Lexikon
  - 7 Kategorien lexikalisch-semantischer Divergenzen

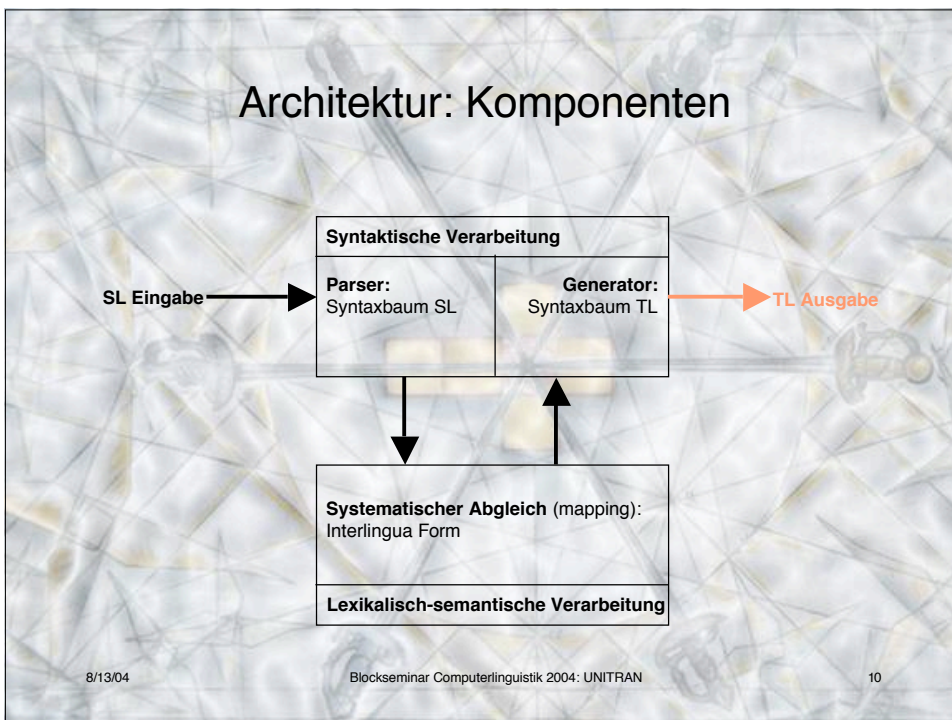
8/13/04 Blockseminar Computerlinguistik 2004: UNITRAN 6



## Architektur: Parametrisierung

- Parametereinstellungen durch Menü-Operationen
- Parser und Generator müssen SL und TL via Parameter verarbeiten können
- Vorteile
  - Trennung von sprachübergreifendem und sprachspezifischem Wissen
  - Vereinfachung der Grammatik-Beschreibung
  - Einfache Veränderung / Vergrößerung des Systems

8/13/04 Blockseminar Computerlinguistik 2004: UNITRAN 9

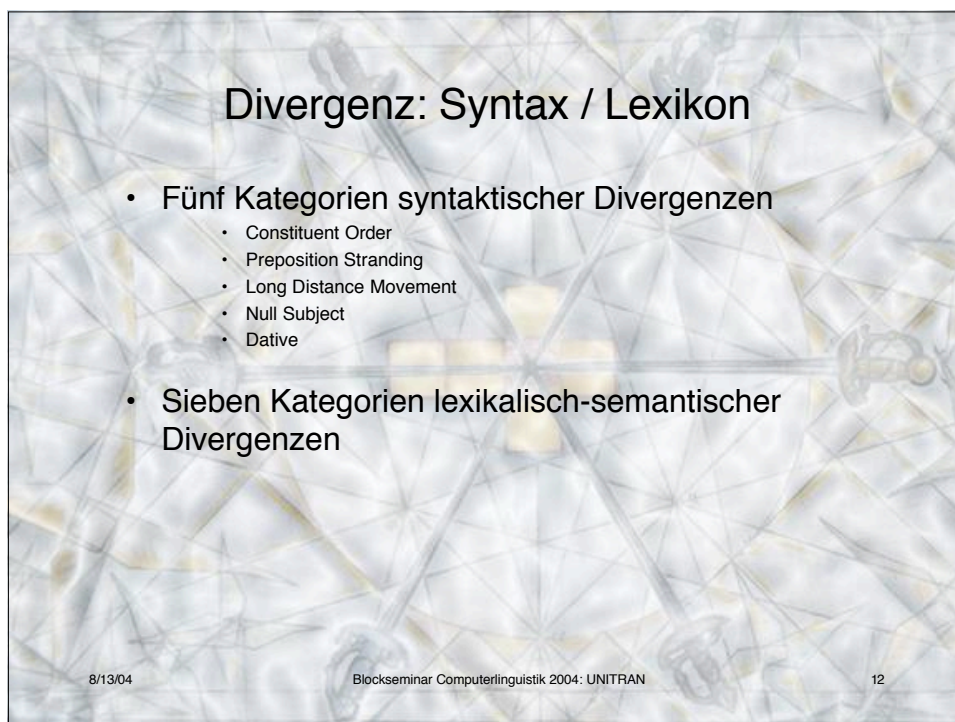




## Inhalt

- UNITRAN in der MT-Landschaft
  - Ziele
  - Ergebnisse
- Architektur
  - Intellingea
  - Parametrisierung
  - Komponenten
- Divergenzproblem
  - Syntax / Lexikon
  - 7 Kategorien lexikalisch-semantischer Divergenzen

8/13/04 Blockseminar Computerlinguistik 2004: UNITRAN 11



## Divergenz: Syntax / Lexikon

- Fünf Kategorien syntaktischer Divergenzen
  - Constituent Order
  - Preposition Stranding
  - Long Distance Movement
  - Null Subject
  - Dative
- Sieben Kategorien lexikalisch-semantischer Divergenzen

8/13/04 Blockseminar Computerlinguistik 2004: UNITRAN 12

## Divergenz: Lexikalisch-semantisch

Conflational	E: I stabbed John. S: Yo le di puñaladas a Juan.
Structural	E: John entered the house. S: Juan entró en la casa.
Thematic	E: I like Mary. S: Me gusta María.
Categorial	E: I am hungry. G: Ich habe Hunger.

8/13/04

Blockseminar Computerlinguistik 2004: UNITRAN

13

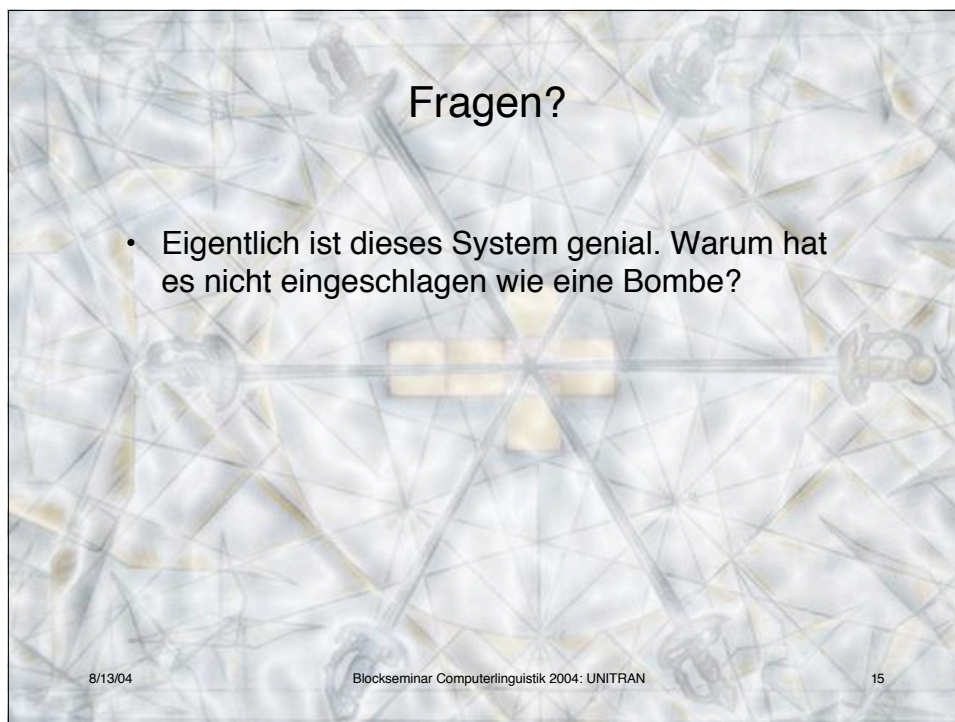
## Divergenz: Lexikalisch-semantisch

Demotional	E: I like to eat. G: Ich esse gerne.
Promotional	E: John usually goes home. S: Juan suele ir a casa.
Lexical	E: John broke into the room. S: Juan forzó la entrada al cuarto.

8/13/04

Blockseminar Computerlinguistik 2004: UNITRAN

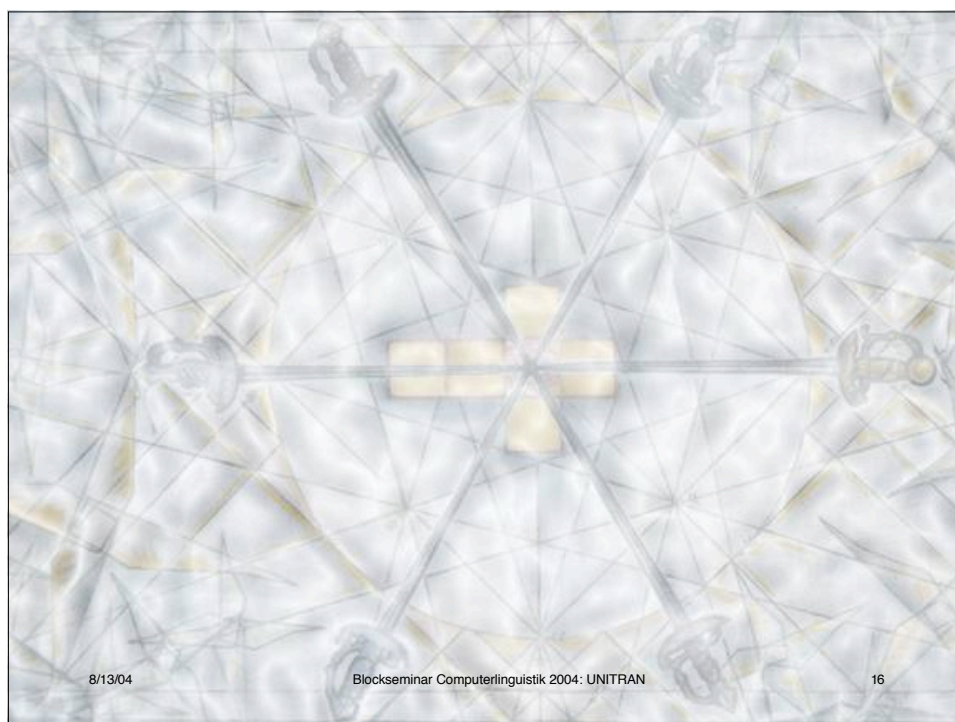
14



Fragen?

- Eigentlich ist dieses System genial. Warum hat es nicht eingeschlagen wie eine Bombe?

8/13/04 Blockseminar Computerlinguistik 2004: UNITRAN 15



8/13/04 Blockseminar Computerlinguistik 2004: UNITRAN 16



## Interlingua-Repräsentation

- ‚Types‘:
  - Event, State, Position, Path, Place...
- Primitives Dimensionen: Örtliche (spatial), Kausale (causal), Feld (field).
  - Spatial: GO, STAY, BE
  - Causal: CAUSE, LET

8/13/04

Blockseminar Computerlinguistik 2004: UNITRAN

17

## Interlingua-Repräsentation

Beispiel:

'John went home'

[Event GO ([Thing JOHN],  
[Path TO  
([Place HOME])])]

8/13/04

Blockseminar Computerlinguistik 2004: UNITRAN

18