

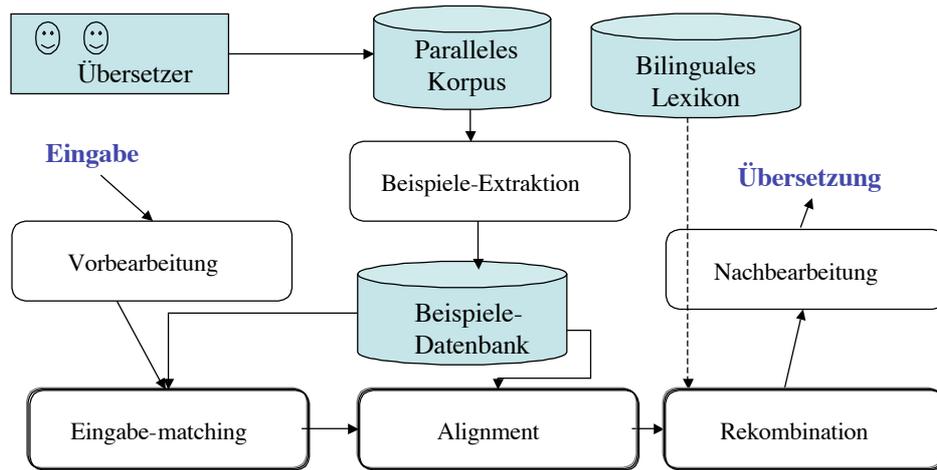
## Systemarchitektur

Cristina Vertan, Walther v. Hahn

### Was implementieren wir im Praktikum?

- Ein Deutsch-English-Deutsches beispielbasiertes Übersetzungssystem.
- Domäne: Autoreparatur
- Eingabetyp: simulierte Spracheingaben für Fragen, sowie Antworten („Telefongespräch mit einem VW-Service-Center“):
  - *Ich habe Winterdiesel. Mein Auto springt trotzdem nicht an.*
  - *Ich verliere Bremsflüssigkeit. Was soll ich machen?*

## Architektur der EBMT-Systeme

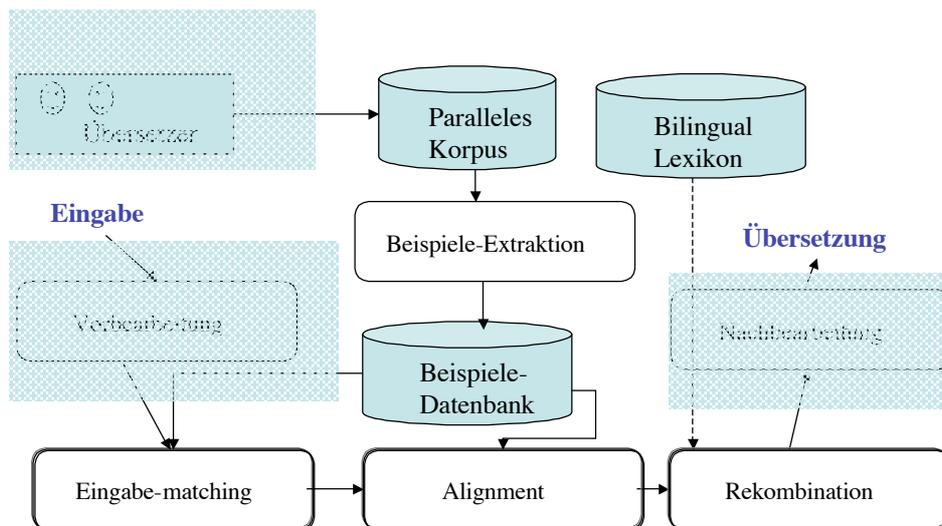


03.11.2004

MTPraktikum WiSe04/05

3

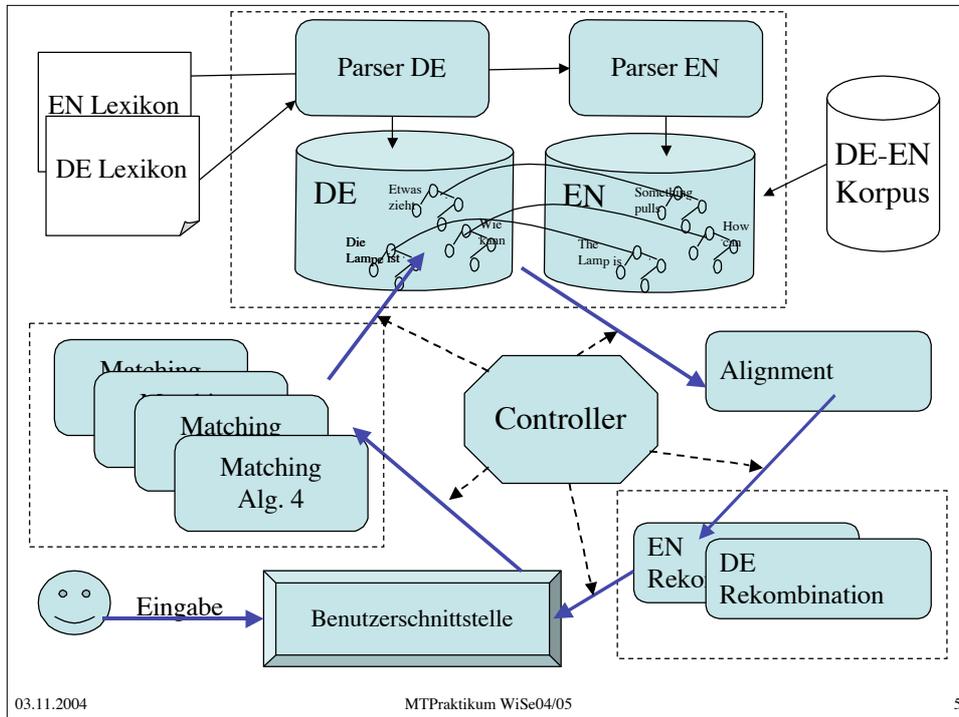
## Architektur der Autoreparatur HILfsystEm (AHILE)



03.11.2004

MTPraktikum WiSe04/05

4



## Aufgabe I Datenbankerstellung

- Lexikoneinträge (Deutsch-Englisch) basierend auf dem Korpus
- Beispieldatenbank-Erstellung.
  - Alinierte Zeichenketten (Sätze) (von dem Korpus + eventuell häufige Gesprächsausdrücke : “Wie kann ich?”)
  - Annotation von eindeutigen Delimitatoren (Quantifikatoren, Konjunktionen)
- Implementierung von folgenden Verfahren:
  - Lexikonsuche (Extraktion von Einträgen und Eintragsmerkmalen z.B. Wortarten) - Nötig für Parser
  - Zugriff auf die Datenbankeinträge

## **Aufgabe II und III Parser für DE und EN**

- Die Parser sollen die Strukturen der Sätze in der Datenbank extrahieren
- Benötigt: Lexikon (DE respektive EN), mindestens mit Einträgen über Wortarten.
- Gemeinsame Aufgabe (II+III): Korrespondenz zwischen den Strukturen.

03.11.2004

MTPraktikum WiSe04/05

7

## **Aufgaben IV, V, VI, VII Matching-Algorithmen**

- Aufgabe IV:
  - längste Zeichenkette (EN, DE)
  - längste Zeichenkette (EN, DE) mit Annotationen
- Aufgabe V: Edit-Distance (EN, DE)
- Aufgabe VI: Wort-Matching (mit semantischem Abstand) (DE, EN)
- Aufgabe VII Wort-Matching (Angle of Similarity) (DE, EN)

03.11.2004

MTPraktikum WiSe04/05

8

## **Aufgabe VIII Alignment**

- Suche von syntaktischen Teilstrukturen in der Zielsprache
- Suche von entsprechenden Zeichenketten in der Zielsprache bei annotierten Zeichenketten

03.11.2004

MTPraktikum WiSe04/05

9

## **Aufgaben IX, X und XI Rekombinierung**

- Rekombinierung von Teilstrukturen (DE)
- Rekombinierung von Teilstrukturen (EN)
- Rekombinierung von Zeichenketten in der Zielsprache (das entspricht dem Fall von annotierten Datenbankeinträgen)

03.11.2004

MTPraktikum WiSe04/05

10

## **Aufgabe XII**

### **System-Architektur**

- Design und Implementierung der Benutzungsschnittstelle
- Design und Implementierung des Controllers, Spezifikation der Modulschnittstellen

## **Vorhandenes Material**

- Englische und Deutsche Versionen des Volkswagen-Handbuchs (als paralleles Korpus)
- Die konzeptuelle Struktur eines deutsch-englischen Lexikons
- Test-Korpus (Beispieleingaben)
- Einführungsmaterial über Parsing
- Beispiel Integration Java und Prolog-Code