# Einführung in die Computerlinguistik: Maschinelle Übersetzung

WS 2012/2013

Manfred Pinkal

## Goethe

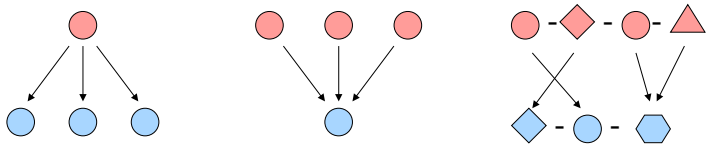
- Über allen Gipfeln ist Ruh. In allen Wipfeln spürest du kaum einen Hauch
- Over all summits is rest. In all treetops you do not feel breath.
- Über allen Gipfeln ist Rest. In allen Treetops glauben Sie nicht Atem.

## Können Computer übersetzen?

Babel Fish
Google

- Über allen Gipfeln ist Ruh. In allen Wipfeln spürest du kaum einen Hauch
- Over all summits is rest. In all treetops you do not feel breath.
- Über allen Gipfeln ist Rest. In allen Treetops glauben Sie nicht Atem.

## Übersetzungsäquivalenz: Elementare Probleme



## Lexikalische Ambiguität

- Homonymie:
  - engl. *rest* → *Rest/Ruhe*
  - *dt. Warte"→ wait/control room*
- Polysemie:
  - *breath* → *Atem/Hauch*
  - *Termin* → *appointment / time slot*
- *"gehen"* in Verbmobil (6 von 15 Varianten)
  - *Gehen wir ins Theater?* – gehen_move
  - *Gehen wir essen?* – gehen_act
  - *Mir geht es gut.* – gehen_feel
  - *Es geht um einen Vertrag.* – gehen_theme
  - *Das Treffen geht von 3 bis 5.* – gehen_last
  - *Geht es bei Ihnen am Montag?* – gehen_passen

## Disambiguierung

... durch satzinternen Kontext (Selektionsbeschränkungen)

- *Wir treffen uns vor dem Frühstück*
  → *before*
- *Wir treffen uns vor dem Hotel*
  → *in front of*

Aber:

- *Wir treffen uns nach Hamburg*
  → *?*

## Disambiguierung

*… durch den Diskurskontext:*

- *Geht das bei Ihnen?*
- *Sollen wir uns am Fünften treffen? Geht das bei Ihnen?*
  → *Is this ok for you?*
- *Wo sollen wir uns treffen? Geht das bei Ihnen?*
  → *Can we meet at your place?*

## Idioms und Kollokationen

- Idioms und Kollokationen: Sprachspezifische, konventionelle Kookkurenzen von Wörtern
  - *Karten geben*
    → *to deal cards*
  - *eine Prüfung ablegen*
    → *to take an exam*
  - *eine Prüfung abnehmen*
    → *to give an exam*
  - *den Fahrschein entwerten*
    → *to validate the ticket*
- WSD reicht nicht aus: Zielsprachliche Entsprechung muss explzit gegeben sein/ gelernt werden.

## Semantische Granularität: Lexikalische Unterschiede

- *I will go to Hamburg tomorrow.*
  → *fahren/fliegen*
- *Ich fahre mit der Bahn nach Hamburg. In Frankfurt muss ich umsteigen.*
  → *change trains*
- *Ich fliege nach Hamburg. In Frankfurt muss ich umsteigen.*
  → *change planes*

## Semantische Granularität: Grammatische Unterschiede

- Geschlechtsspezifische Personenbezeichnungen im Deutschen
  - *doctor → Arzt / Ärztin*
  - *teacher → Lehrer / Lehrerin*
- Präsens und Futur im Englischen
  - *Ich fahre nach Hamburg → I am going / I will go to Hamburg*
- Verbaspekt:
  - *Simple Present/ Progressive im Engl.*
  - *Vollendete/unvollendete Form im Russ.*

## Granularität D/E - J

Deutsch/Englisch → Japanisch

- J: Höflichkeitsformen
- J:Topikmarkierung (gegeben/ neu)

Japanisch → Deutsch/Englisch

- D: Artikel/ Definitheit (bestimmt/ unbestimmt), J hat keine Artikel
- J: „Null-Anapher“: Satzteile werden tendenziell weggelassen, wenn aus dem Kontext erschließbar ("Null-Anapher")

## Beispiel

*"Termin ausgemacht?"*

*Yotei-wa kimemashita ka.*

→ *Hat er (mit Ihnen) einen Termin ausgemacht?*

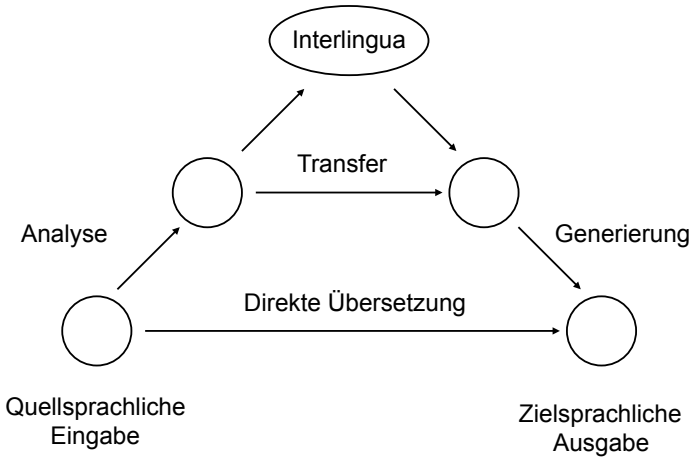
*Go-yotei wa okimeni narimashita ka.*

→ *Haben Sie (mit ihm) einen Termin ausgemacht?*

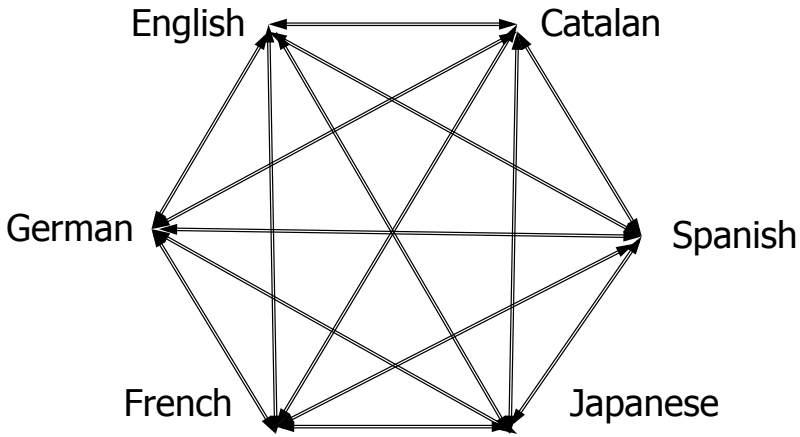
## Paradigmen der Maschinellen Übersetzung

- Wissensbasierte MÜ
- Statistische MÜ

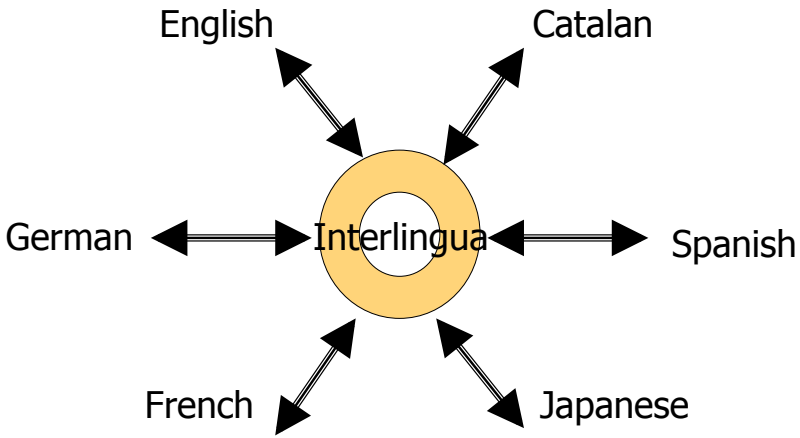
## Das "Vauquois-Dreieck"




## Das Transfer-Modell



## Das Interlingua-Modell




## Interlingua und Transfer

- Die Übersetzung in die / aus der Interlingua muss für jede neue Sprache nur (je) einmal bereitgestellt werden. – Wenn im Transfermodell zu n Sprachen eine neue hinzukommt, müssen 2n neue Übersetzungsrichtungen bereitgestellt werden.
  - Beispiel: Durch die letzten EU-Erweiterungen wachsen die EU-Amtssprachen von 11 auf 23 an.
  - Statt 110 Übersetzungspaaren benötigt man 506.
- Interlingua muss extrem feingranular sein, da alle Unterschiede in allen Sprachen darstellbar sein müssen. Das erfordert bei der Übersetzung einen immer gleich hohen und für viele, insbesondere eng verwandte Sprachpaare unnötigen Übersetzungsaufwand.
  - Beispiel: Übersetzung D-E benötigt keine detaillierte Bestimmung von Höflichkeitsinformation

## Wissensbasierte MÜ

- Techniken: Stemmer/Morphologien, Grammatiken, Lexika für Quell- und Zielsprache, Transferregeln, sprachunabhängige Ontologien, Weltwissen, Inferenzregeln
- Probleme
  - Abdeckung: Ungeheure Vielfalt von syntaktischen und semantischen Phänomenen und Übersetzungsäquivalenten
  - Präzision: Ambiguität und Granularitätsunterschiede
- Klassisches (und noch immer aktuelles) Beispiel:
  - SYSTRAN (Babel Fish)

## Statistische MÜ

- Gesucht: Der wahrscheinlichste zielsprachliche Satz, gegeben ein quellsprachlicher Satz; z.B.: eine englische Wortkette (*E*), gegeben eine deutsche Wortkette (*D*).

$$\max_E P(E \mid D)$$

- Das erinnert an das Problem der Spracherkennung: Gesucht ist die wahrscheinlichste Wortkette, gegeben eine Folge akustischer Merkmalsmuster:

$$\max_W P(W \mid O)$$

# Wie bestimmen wir *P(E|D)*?

- Bayes-Regel :

Spracherkennung:

$$P(W \mid O) = \frac{P(O \mid W) \cdot P(W)}{P(O)}$$

$$\max_W P(W \mid O) = \max_W \frac{P(O \mid W) \cdot P(W)}{P(O)}$$

$$= \max_W P(O \mid W) \cdot P(W)$$

Übersetzung:

$$P(E \mid D) = \frac{P(D \mid E) \cdot P(E)}{P(D)}$$

$$\max_E P(E \mid D) = \max_E \frac{P(D \mid E) \cdot P(E)}{P(D)}$$

$$= \max_E P(D \mid E) \cdot P(E)$$

# Übersetzungsmodell und Sprachmodell

$$\max_E P(E \mid D) = \max_E P(D \mid E) \cdot P(E)$$

- Die Güte einer Übersetzung wird bestimmt durch:
  - Den Grad der Entsprechung von zielsprachlichem und quellsprachlichem Satz, approximiert durch das Übersetzungsmodell *P(D|E)*.
  - Die "Normalität" oder "Flüssigkeit" des zielsprachlichen Satzes, approximiert durch das zielsprachliche Sprachmodell *P(E)*

# Übersetzungsmodell

$$\max_E P(E \mid D) = \max_E P(D \mid E) \cdot P(E)$$

- Problem ist die Zuordnung von quellsprachlichen und zielsprachlichen Wörtern: Tilgungen, Einfügungen, Mehr-zu-Eins-Entsprechungen, unterschiedliche Wortstellung.
- Alignierung großer Parallelkorpora (z.B. Europarl: Akten des Europäischen Parlaments)
- Wortalignierung mithilfe eines Alignierungsalgorithmus, der Wörter so aligniert, dass die "Alignierungskosten" (durch Tilgung, Einfügung, Permutation) minimiert werden (vgl. Levenshtein-Distanz)
- Übersetzungsmodelle erhält man aus dem alignierten Parallelkorpus.

# Sprachmodell für die Zielsprache

$$\max_E P(E \mid D) = \max_E P(D \mid E) \cdot P(E)$$

- Wie berechnen wir $P(Z) = P(w_1 w_2 \ldots w_n)$ ?
- n-Gramm-Technik, z.B. Bigramm-Approximation:
  - $P(w_n|w_1 w_2 \ldots w_{n-1}) \approx P(w_n|w_{n-1})$

    $P(w_1 w_2 \ldots w_n) \approx P(w_1) * P(w_2|w_1) * P(w_3|w_2) * \ldots P(w_n|w_{n-1})$

# Statistische MÜ

- Liefert im Allgemeinen Resultate, die den besten wissensbasierten Systemen vergleichbar sind.
- Systeme lassen sich vergleichsweise schnell trainieren und auf neue Sprachen/ Domänen adaptieren.
- Für bestimmte Anwendungen sehr hochwertige Übersetzungen, weil Muster aus Parallelkorpora komplett übernommen werden können.
- Beispiel: Google Translate