

Einführung in die Computerlinguistik: Maschinelle Übersetzung

WS 2014/2015

Andrea Horbach

Mit Folien von Manfred Pinkal

Babel Fish , ca. 2007

- 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.

Babel Fish , ca. 2007

- 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.

Lexikalische Mehrdeutigkeit

- Homonymie:
 - engl. *rest* → *Rest/Ruhe*
- 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

Ambiguitätsauflösung

... durch satzinternen Kontext (Sortenbeschränkungen)

- *Wir treffen uns vor dem Frühstück*
→ before

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

Aber:

- *Wir treffen uns nach Hamburg*
→ ?

Ambiguitätsauflösung

... durch den Diskurskontext

- *Geht es bei Ihnen?*
- *Wo sollen wir uns treffen? Geht das bei Ihnen? → at your place*
- *Sollen wir uns am Fünften treffen? Geht das bei Ihnen? → for you*

Idioms und Kollokationen

- *Spielkarten 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

- Sprachspezifische, konventionelle Mehrwortausdrücke, die gelernt bzw. im Lexikon explizit vorgegeben werden muss – i.d.R. keine semantische Mehrdeutigkeit

Granularitätsunterschiede im Lexikon

- *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*

Systematische Granularitäts-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?"

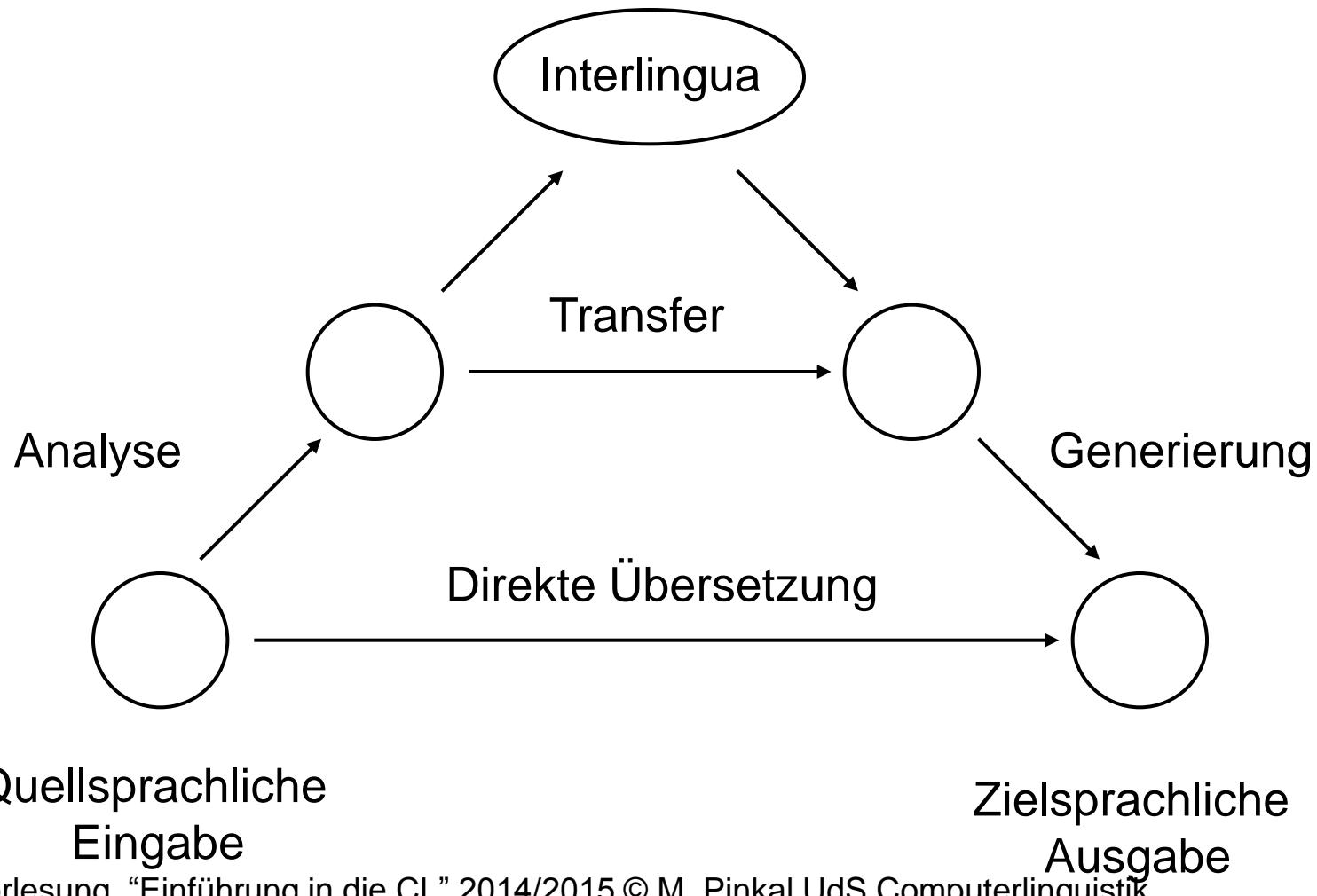
Yotei-wa kimemashita ka. → Hat er (mit Ihnen) einen Termin ausgemacht?

Go-yotei wa okimeni nari mashita ka. → Haben Sie (mit ihm) einen Termin ausgemacht?

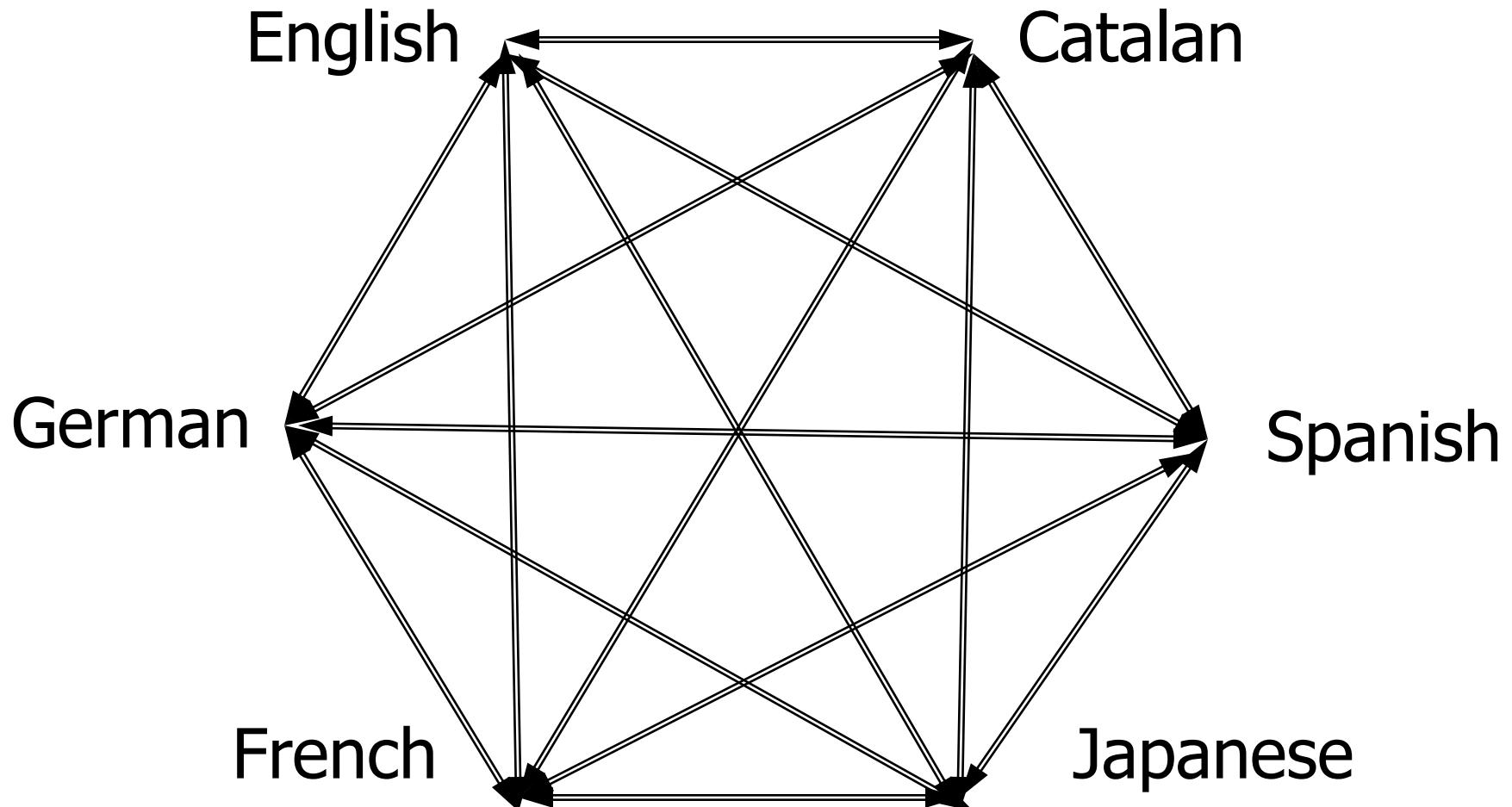
Paradigmen der Maschinellen Übersetzung

- Wissensbasierte MÜ
- Statistische MÜ

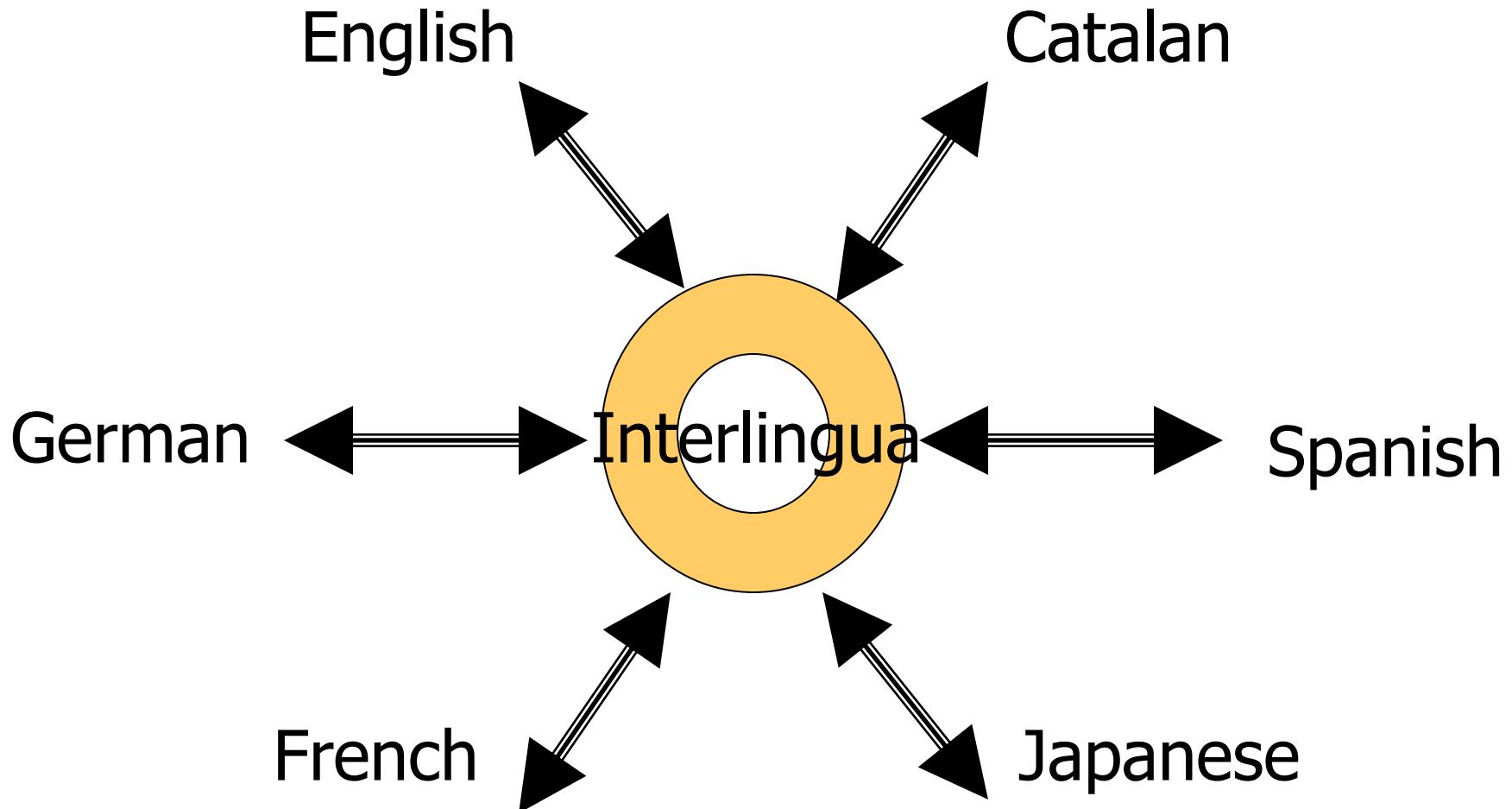
Das "Vauquois-Dreieck"



Transfer-Modell



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 offiziellen EU-Sprachen von 11 auf 24 an.
 - Statt 110 Übersetzungspaaren benötigt man 552.
- 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

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

Statistische MÜ

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

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

- 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 $P(D|E)$?

- Bayes-Regel :

Spracherkennung:

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

$$\begin{aligned} \max_W P(W | O) &= \max_W \frac{P(O | W) \cdot P(W)}{P(O)} \\ &= \max_W P(O | W) \cdot P(W) \end{aligned}$$

Übersetzung:

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

$$\begin{aligned} \max_D P(D | E) &= \max_D \frac{P(E | D) \cdot P(D)}{P(E)} \\ &= \max_D P(E | D) \cdot P(D) \end{aligned}$$

Übersetzungsmodell und Sprachmodell

$$\max_D P(D | E) = \max_D P(E | D) \cdot P(D)$$

Die Güte einer Übersetzung wird bestimmt durch:

- Die Korrektheit oder „Originaltreue“, approximiert durch das **Übersetzungsmodell** $P(E|D)$.
- Die Natürlichkeit oder „Flüssigkeit“ des zielsprachlichen Satzes, approximiert durch das zielsprachliche **Sprachmodell** $P(D)$.

Übersetzungsmodell

$$\max_D P(D | E) = \max_D P(E | D) \cdot P(D)$$

- Als Datengrundlage verwenden wir Parallelkorpora, die Texte und ihre Übersetzungen in verschiedenen Sprachen enthalten.
- Wichtigste Datengrundlage ist [Europarl](#): Die Akten des europäischen Parlaments mit Parallelversionen aller Parlamentsdokumente in den offiziellen EU-Sprachen.
- Vorverarbeitungsschritt ist Alignierung auf Satzebene. Europarl ist bereits satzweise aligniert.
- Aufgabe ist die Abschätzung der Wahrscheinlichkeit eines englischen Quellsatzes E , gegeben eine deutsche Übersetzung D .
- Häufigkeitsschätzung auf Satzebene: Sparse-Data-Problem!
- Lösung: Approximation der Satzwahrscheinlichkeit durch Übersetzungswahrscheinlichkeiten auf Wort- oder Konstituenten-Ebene

Übersetzungsmodell: Erster Versuch

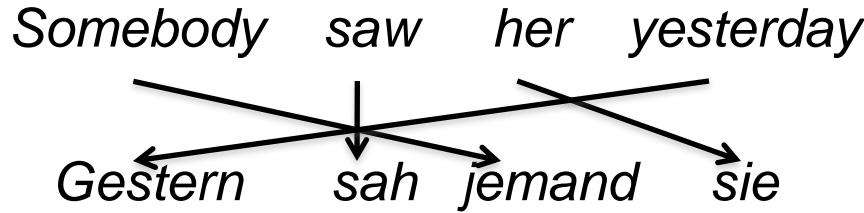
- Wir gehen von Wort-für-Wort-Übersetzung aus, nehmen an, dass die Übersetzungs wahrscheinlichkeiten $t(d|e)$ für einzelne deutsch-englische Wortpaare unabhängig voneinander sind, und erhalten für die Satzlänge n :

$$P(E | D) \approx \prod_{i=1}^n t(e_i | d_i)$$

- Beispiel:

Somebody saw her
↓ 0.8 ↓ 0.3 ↓ 0.5
Jemand sah sie

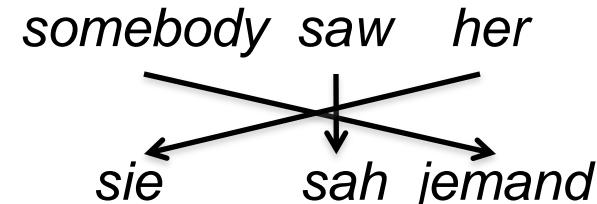
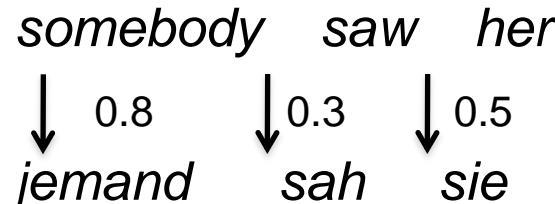
Übersetzungsmodell: Wortalignierung



- Zweiter Versuch: Wir verwenden dieselbe Gleichung, wobei i nicht über Positionen in der Wortkette, sondern über Alignierungspaare (im Beispiel: (1,3), (2,2), (3,4), (4,1)) läuft.

$$P(D | E) \approx \prod_{i=1}^n t(d_i | e_i)$$

Problem:



Übersetzungsmodell: Umordnungskosten

somebody saw her

↓ 0.8 ↓ 0.3 ↓ 0.5

jemand sah sie

somebody saw her

~~sie~~ → ~~sah~~ → ~~jemand~~

$$P(E | D) \approx \prod_{i=1}^n t(e_i | d_i) * d(d_i, e_i)$$

$$d(d_i, e_i) = \frac{1}{2^{|i - pos(e_i)|}}$$

Dabei ist $pos(e_i)$ die Position des mit d_i alignierten Wortes e_i und $pos(d_i)$ die Position des i-ten Wortes in d , d.h. $pos(d_i) = i$

Übersetzungsmodell: Probleme

- Somebody saw her yesterday
- Gestern sah jemand sie

- Somebody saw her the day before yesterday
- Jemand sah sie vorgestern

- I guess somebody saw her
- Ich vermute, dass jemand sie sah

- I guess somebody saw her
- Ich vermute, dass jemand sie gesehen hat

- Somebody saw her
- Jemand hat sie gesehen

Übersetzungsmodell und Sprachmodell

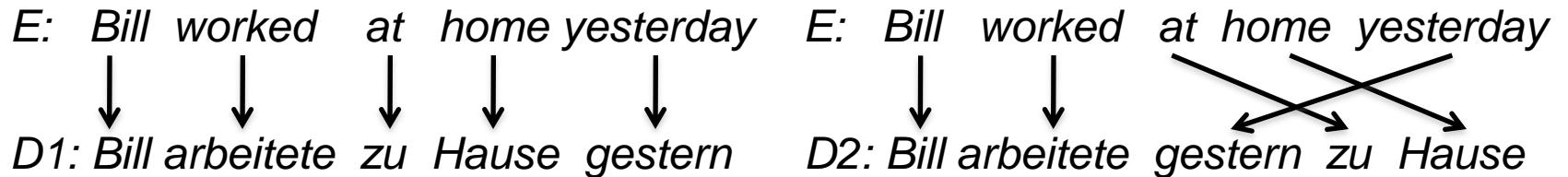
$$\max_D P(D | E) = \max_D P(E | D) \cdot P(D)$$

Die Güte einer Übersetzung wird bestimmt durch:

- Die Korrektheit oder „Originaltreue“, approximiert durch das Übersetzungsmodell $P(E|D)$.
- Die Natürlichkeit oder „Flüssigkeit“ des Zielsprachlichen Satzes, approximiert durch das Zielsprachliche Sprachmodell $P(D)$.

Sprachmodell für die Zielsprache

$$\max_D P(D | E) = \max_D P(E | D) \cdot P(D)$$



$$P(E | D1) > P(E | D2) \quad P(E | D1) * P(D1) < P(E | D2) * P(D2)$$

- n-Gramm-Approximation für Sprachmodelle, Beispiel: Bigramm

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

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

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.

Können Computer übersetzen?

Antwort hängt ab von

■ **Textsorte:**

Dichtung, Rechtstext vs. Gebrauchstexte

■ **Erwartung des Nutzers:**

Präzise Information vs. Abschätzung des Inhalts

■ **„Leidensdruck“ des Nutzers:**

MÜ Englisch → Deutsch vs. MÜ Chinesisch → Deutsch