

Referat

Automatisches Übersetzen

Michael Pointner, Hubert Mayer

Inhalt

- Geschichte des Automatischen Übersetzens
- Gründe für Automatisches Übersetzen
- Worin unterscheiden sich Sprachen?
- Übersetzungsprobleme
- Übersetzungs-Methoden
- Übersetzungsqualität

Geschichte der Automatischen Übersetzung

- Anfänge in einem russisch-englisch Übersetzungsprogramm für das US-Militär
- 1966: US-Verteidigungsministerium bescheinigte Unrealisierbarkeit
- 1980er: System "SUSY,,
- 1990er: BMBF-Leitprojekt Verbmobil
- Google Übersetzer seit 2006

Gründe

- Viele Texte sind heute digital verfügbar
- Der Markt für Übersetzung verdoppelt sich alle vier Jahre
- Kommerziell wichtig: die ostasiatischen Sprachen Chinesisch, Koreanisch und Japanisch; sowie Thai.
- Militärisch wichtig: Sprachen der internationalen Konfliktregionen, vor allem mit Beteiligung des US-Militärs

Grundproblematik einer Übersetzung

- **Doppelte Bindung**
 - Ein Text soll sowohl den Stil in der Ausgangssprache, aber auch die Anforderungen des Lesers der Zielsprache erfüllen.
 - Bestimmte Grammatikstrukturen der Ausgangssprache sollen in der anderen Sprache nachempfunden werden.
- **Subjektivität**
 - Beim Übersetzungsvorgang fließt immer ein gewisser Charakter des Übersetzers ein.



*Worin unterscheiden
sich zwei Sprachen ?*

- Lautsystem
- Silben
- Vokabular
- Wortstellung
- Morphologie (Wortendungen, Affixe, Suffixe)
- Fälle
- Artikel
- Satzglieder/Phrasen
- Syntax
- Wortgeschlecht (Genus)
- Singular/Plural/Dual
- Konnotationen (gelbes Haar vs blondes Haar)
- Modi (Indikativ, Imperativ, Konjunktiv)

Schriftsysteme

Аа Бб Вв Гг Дд Ее
 Ёё Жж Зз Ии Йй
 Кк Лл Мм Нн Оо
 Пп Рр Сс Тт Уу Фф
 Хх Цц Чч Шш Щщ
 Ъъ Ыы Ьь Ээ Юю
 Яя (İi Øø Vv Ъь)

あ	い	う	え	お
a	i	u	e	o
か	き	く	け	こ
ka	ki	ku	ke	ko
さ	し	す	せ	そ
sa	shi	su	se	so
た	ち	つ	て	と
ta	chi	tsu	te	to
な	に	ぬ	ね	の
na	ni	nu	ne	no
は	ひ	ふ	へ	ほ
ha	hi	fu	he	ho
ま	み	む	め	も
ma	mi	mu	me	mo
や		ゆ		よ
ya		yu		yo
ら	り	る	れ	ろ
ra	ri	ru	re	ro
わ				を
wa				wo
ん				
n				

ア	イ	ウ	エ	オ
a	i	u	e	o
カ	キ	ク	ケ	コ
ka	ki	ku	ke	ko
サ	シ	ス	セ	ソ
sa	shi	su	se	so
タ	チ	ツ	テ	ト
ta	chi	tsu	te	to
ナ	ニ	ヌ	ネ	ノ
na	ni	nu	ne	no
ハ	ヒ	フ	ヘ	ホ
ha	hi	fu	he	ho
マ	ミ	ム	メ	モ
ma	mi	mu	me	mo
ヤ		ユ		ヨ
ya		yu		yo
ラ	リ	ル	レ	ロ
ra	ri	ru	re	ro
ワ				ヲ
wa				wo
ン				
n				

- Unterschiede zwischen Sprachen.
- Ein Symbol bedeutet
 - Ein Laut Deutsch (mit Ausnahmen)
 - Eine Silbe Japanisch
 - Ein ganzes Wort Chinesisch
- Wortenden in manchen Sprachen schwer definierbar
 - Mehrere Schriftzeichen können 1 oder mehreren Wörtern entsprechen.
 - Die Maschine muss die Enden der Wörter herausfinden.

- Auf **Walisisch** kann ein Wort eine lange Nominalphrase beschreiben
- Llanfairpwllgwyngyllgogerychwyrndrobwllllantysiliogogoch
- „Marienkirche in einer Mulde weißer Haseln in der Nähe eines schnellen Wirbels und in der Gegend der Thysiliokirche, die bei einer roten Höhle liegt“.

Wortstellung

Wortstellung

Subjekt – Verb – Objekt (SVO)

Subjekt – Objekt – Verb (SOV)

Objekt – Verb – Subjekt (OVS)

Objekt – Subjekt – Verb (OSV)

Verb – Subjekt – Objekt (VSO)

Verb – Objekt – Subjekt (VOS)

Beispielsatz

„Peter schreibt Romane.“

„Peter Romane schreibt.“

„Romane schreibt Peter.“

„Romane Peter schreibt.“

„Schreibt Peter Romane.“

„Schreibt Romane Peter.“

Beispielsprachen

Englisch, Französisch, Spanisch,
Swahili

Japanisch, Koreanisch, Türkisch,
Persisch

Hixkaryana, Apalai, Bacairi

Xavante

inselkeltische Sprachen,
Hawaiisch, Klassisches Arabisch

Fidschi, Malagasy

Grammatik und Morphologie

- Sprachen haben unterschiedliche Konzepte um Fälle und Korrelationen zwischen Wörtern wiederzugeben.
- Ausdrücken von Zeiten
 - Durch Hilfsörter wie „werden“ oder „will“ Deutsch, Englisch
 - Durch Flexionen (Wortänderungen) Deutsch, Englisch, Japanisch
- Ausdrücken von Fällen
 - Deutsch: Artikel, Präpositionen und Flexionen (für meinen Freund)
 - Englisch: Wortreihenfolge und Präpositionen (for the friend)
 - Japanisch: Partikel nach dem Wort (tomodachi ni)
- Ausdrücken von Mehrzahl
 - Deutsch: Flexionen: Tasche/Taschen,
 - Englisch: Plural s : Bag/Bags
 - Japanisch: Keine Mehrzahl, wenn nötig durch Hilfsörter wie „viele, mehrere, drei“ ausgedrückt
- Verneinungen
 - Deutsch, Englisch: Durch Hilfsörter „nicht, not“Deutsch, Englisch
 - Japanisch: Durch Flexionen し・ません shimasen (nicht tun)
 つめた・くない tsumetai (nicht kalt)

Übersetzungsprobleme

Mehrdeutigkeit/Synonyme

- Viele Wörter haben mehr als nur eine Bedeutung
- Welche Bedeutung ist die Richtige?
 - I drew a castle! Schloss oder Burg?
- Auch Sätze haben diese Charakteristik
 - I saw the man on the hill with a telescope.
- Homonyme
 - Wörter mit gleicher Aussprache und Schreibweise aber unterschiedlicher Bedeutung
 - Swallow: Die Schwalbe, schlucken
 - Date: Datum, Verabredung, die Person mit der ich verabredet bin. „Thats my date“
 - Right: Rechts, richtig „You’re right.“
- Es gibt viele Synonyme für Wörter und die Übersetzung muss das passende für den Kontext verwenden.
 - übersetzen
 - To translate, interpret, compile, decode, ferry over, cross over, localize



Homonymie

unterschiedliche
Bedeutung und auch oft
unterschiedliche Herkunft



Homographie

dieselbe Schreibweise bei
unterschiedlicher
Aussprache, z.B.
„modérn“ = „fortschrittlich“,
„módern“ = „verwesen“

Homophonie

dieselbe Aussprache bei
unterschiedlicher
Schreibweise, z.B. „Wende“
und „Wände“

Quelle:

<https://de.wikipedia.org/wiki/Homograph#/media/File:Homonymie.svg>

- Eigenamen
 - Mohammed Kellner
 - Unterschiedliche Lautsysteme
 - Michael -> Mihaeru (Deutsch -> Japanisch)
- Ortsnamen
 - Baden bei Wien
- Anglizismen und Lehnwörter aus anderen Sprachen
 - Jugendsprache

Problem: Einsprachige Wörter

- Manche Ausdrücke existieren nur in der Ausgangssprache
 - Schadenfreude
 - Diese Wort gibt es im Englischen nicht.
 - Option A: Umschreiben
 - Option B: Stehen lassen
 - It's raining cats and dogs
 - Dieses Sprichwort gibt es nicht so auf Deutsch
 - Ist es genügend ein deutsches Sprichwort zu verwenden?
 - Geht Information beim Übersetzen verloren?

Dialekte und Umgangssprache

- Dialekte können bei automatischen Übersetzungen nur schlecht oder gesondert berücksichtigt werden.
 - „Du Hubert, bringst ma di Leita.“
 - generalise - generalize (RP vs American)
- Umgangssprache
 - Suit yourself.
 - <https://translate.google.at/?hl=de#en/de/suit%20yourself>

Veraltete Ausdrucksarten

- Viele syntaktische Formen sind veraltet und werden heutzutage üblicherweise nicht mehr verwendet.
- Die Maschine kennt die Grammatik nicht oder hat zu wenige Beispiele für eine Übersetzung.
- Eingabe: Auf dass du auf deinem Wege nicht verzweifeln mögest.
- Übersetzung: mayest On that you not despair in your ways.

Weitere Problemgebiete

- Die germanischen und romanischen Sprachen haben geschlechtsspezifische Artikel und Relativpronomen
 - Bei Übersetzung aus einer anderen Sprache müssen diese „aus der Luft“ generiert werden.
- Wörter werden kreiert, verändert, geborgt, ersetzt und vergessen
 - Handyakku, Waschmaschinenreparaturdienstangestellter
- Wörtliche Bedeutung VS gemeinte Bedeutung
 - → Siehe letzte Einheit „Mir ist kalt.“

Die Ausgangssprache bestimmen

- Manche Sprachen teilen sich bestimmte Wörter in ihrer Schriftweise, daher ist es wichtig zu wissen um welche Sprache es sich handelt.
- Beispiel: Chinesisch und Japanisch
 - Diese Sprachen teilen sich viele Schriftzeichen
 - Diese Zeichen haben aber oft ähnliche oder ganz unterschiedliche Bedeutungen

Das Tegami Problem

私は手紙を書きます。

手
紙
手紙

Übersetzungs-Methoden - Überblick

- Direkte Maschinenübersetzung
- Transfer
- Interlingua
- Example-Based Machine Translation
- Statistics-Based Machine Translation
- Human-Aided Machine Translation

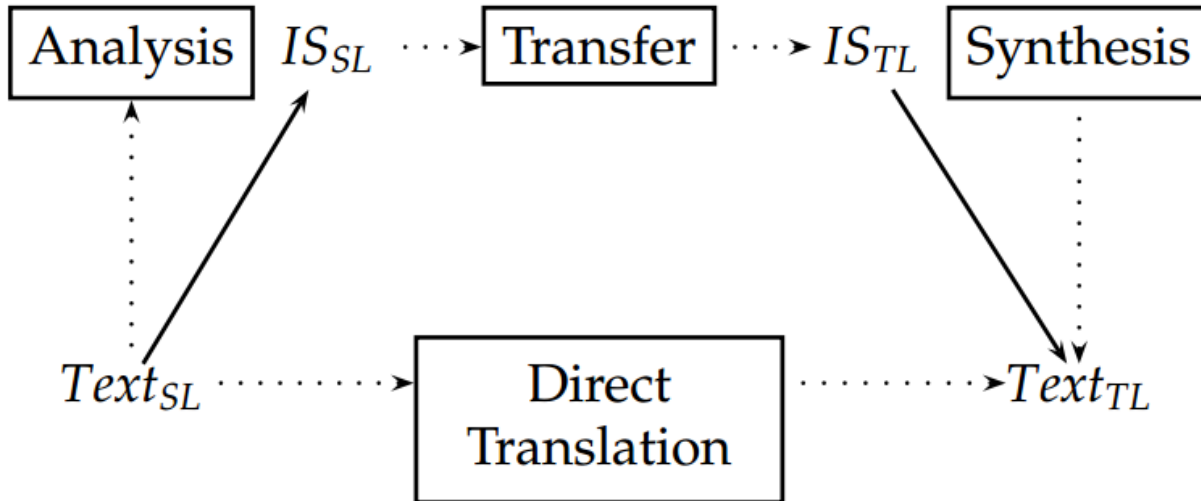
Übersetzungs-Methoden

- **Direkte Maschinenübersetzung: Wort-zu-Wort Übersetzung**



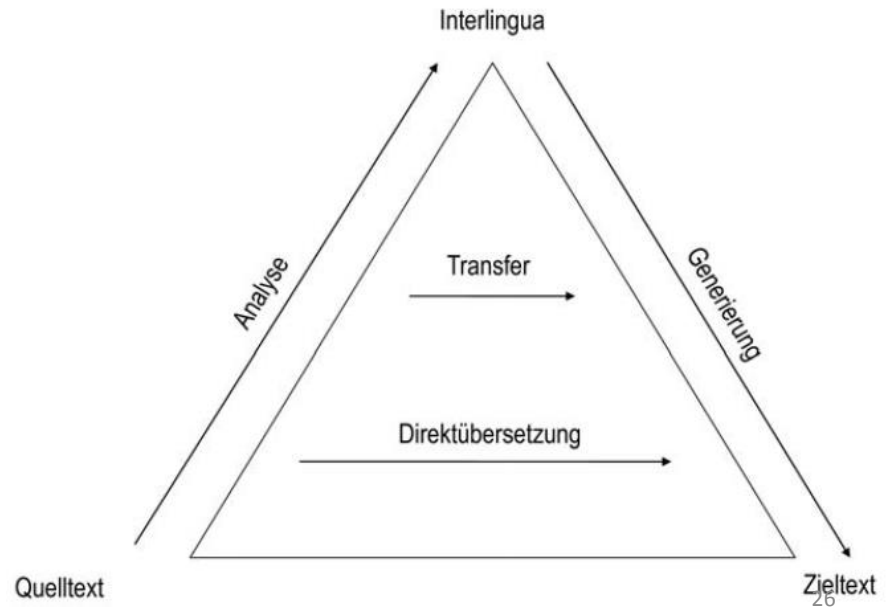
Übersetzungs-Methoden

- **Transfer:** Übertragen der Grammatikstrukturen in die Zielsprache



Übersetzungs-Methoden

- **Interlingua: Übersetzen über eine Zwischensprache**



Übersetzungs-Methoden

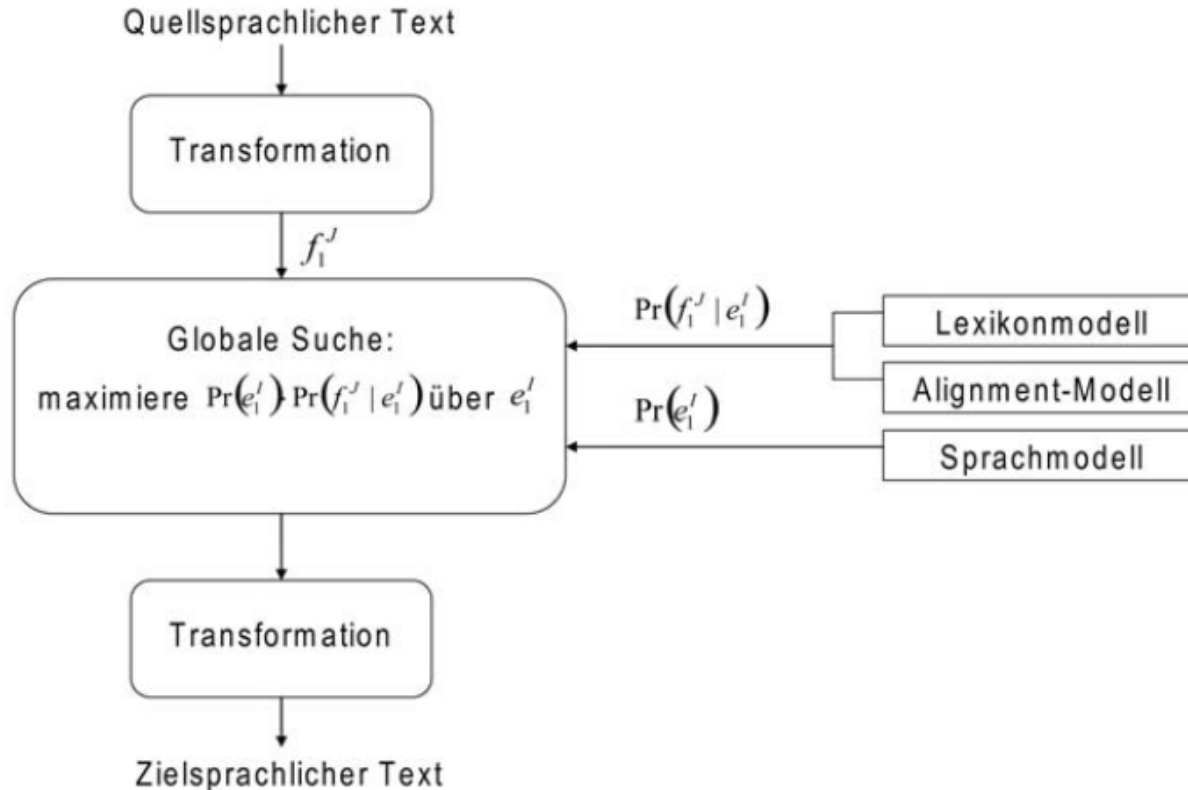
- **Example-Based Machine Translation:**
Speichern von häufig wiederkehrenden Sätzen

Julie bought a notebook	Julie compró una libreta
Ann read a book on economics	Ann leyó un libro de económica

Übersetzungs-Methoden

- **Statistics-Based Machine Translation:** Wörter und grammatische Formen in Ausgangs- und Zielsprache werden, aufgrund ihrer Häufigkeit und gegenseitigen Nähe, einander zugeordnet

Statistics-Based Machine Translation



Übersetzungs-Methoden

- **Human-Aided Machine Translation:**
dabei wird ein menschlicher Benutzer gebeten, mehrdeutige oder schwierig zu übersetzende Konstruktionen selbst zu übersetzen

Übersetzungsqualität

- Wird oft von Muttersprachlern der Zielsprache durchgeführt und per Kennziffer ausgedrückt.
 - Japan beispielsweise verwendet ein Punktesystem von 0-4 Punkten
 - 4P. Sehr gut verständlich
 - 3P. Ein paar falsche Wörter, sonst gut verständlich
 - 2P. Mit gutem Willen kann man sich denken was gemeint war.
 - 1P. Der Satz wird in einem falschen Sinn verstanden, wenn überhaupt
 - 0P. Der Satz ergibt keinen Sinn; Eine zufällige Ansammlung an Wörtern.

Bleu-score

- Algorithmus der die automatische Übersetzung mit einer Referenz eines Menschen vergleicht.
- Wurde kritisiert für ihre Unverlässlichkeit, da nur auf Satzebene und nicht auf Ebene der Semantik überprüft werden kann.
- Die **Trefferrate**, beschreibt wie viele Terme einer Automatischen Übersetzung ein menschlicher Übersetzer unverändert übernehmen kann.
 - Die Katze isst das Futter → The cat is the food.
Referenz: The cat eats the food.
Trefferrate von: 80% 4 von 5 Wörtern übernehmbar.

Evaluation

Funktioniert das Automatische Übersetzen?

- Ja, die sinngemäße Bedeutung wird meist vermittelt.
- ABER: Die genaue Grammatik und Formulierung kann oft nicht wiedergegeben werden.
- Faustregeln:
 - Je verbreiteter eine Sprache ist desto eher kann man den automatischen Übersetzungen vertrauen. → Corpus
 - Eine Sprache ist dann gut übersetzbar, wenn die Zielsprache weniger komplexen Regeln folgt als die Ausgangssprache.

- *Dieser Satz wird nun in 5 verschiedene Sprachen übersetzt und danach wieder zurück auf Deutsch übersetzt.*
- *Die Reihenfolge lautet: Deutsch, Englisch, Französisch, Italienisch, Spanisch*
- Dieser Satz wird in 5 Sprachen übersetzt und ins Deutsche übersetzt wieder. Die Reihenfolge ist: Deutsch, Englisch, Französisch, Italienisch, Spanisch
- Ich übersetze diese Satz jetzt auf Englisch und danach 2mal hin und her zwischen Englisch und Chinesisch und danach wieder zurück auf Deutsch. Mal schauen was passiert.
- mein Englisch jetzt und dann dieser Satz in zwei hin und her zwischen Englisch und vereinfachtes Chinesisch% 2can wird dann nach Deutschland. mal sehen, was passiert.

Wie funktioniert Google Translate

- <https://www.youtube.com/watch?v=h11GzWmenag>
 - Statistische Maschinelle Übersetzung
 - Mustererkennung (Corpus)

Demo

- It's raining cats and dogs (EN -> DE)
Es schüttet wie aus Kübeln (Richtig)
- Baila corazón, baila (ES -> DE)
Tanz Schatz, tanz
- Wenn ich arbeiten würde, würde ich ein Auto kaufen (DE -> EN)
If I were working, I would buy a car.

Demo

- padded hazelnut ice cream (EN -> DE)
geschlagenes Haselnuss-Eis
- мороженое в асc (RU -> DE)
Eis in einem Esel
(Aussprache: Maroschn newas)
- Mein Englisch ist nicht das Gelbe vom Ei (DE
-> EN)
My English is not exactly brilliant (Richtig)

Quellen

- <http://dietrich-schaeckermann.de/html/Grammat>
- [https://de.wikipedia.org/wiki/Syntax#Stellung der Syntax in der Grammatik](https://de.wikipedia.org/wiki/Syntax#Stellung_der_Syntax_in_der_Grammatik)
- http://media.dwds.de/jlcl/2009_Heft3/ldv-forum-24-3.pdf
- <http://orb.essex.ac.uk/lg/lg619/mt/index.pdf>
- [https://de.wikipedia.org/wiki/Maschinelle %C3%9Cbersetzung](https://de.wikipedia.org/wiki/Maschinelle_%C3%9Cbersetzung)
- <https://de.wikipedia.org/wiki/Homograph>
- <https://www.youtube.com/watch?v=h11GzWmenag>

Danke für eure Aufmerksamkeit!