

Maschinelle Übersetzung

Regelbasierte Ansätze

Computerlinguistik II
Dr. Jürgen Hermes
Institut für Digital Humanities
Universität zu Köln

Referentinnen:
Anna Schmitz
Sümeyye Peru

Inhalt

1. Definition
2. Nutzen
3. Geschichte und Gegenwart
4. Methoden
5. Regelbasierte MÜ
6. Grade und eines RBMT-Systems
7. Problematiken
8. Vor- und Nachteile
9. Fazit
10. Quellen

Definition

- Übersetzen vs. Dolmetschen
- Übersetzung wird verstanden als “die schriftliche Übertragung eines Textes, der in einer Quellsprache vorliegt, in einen entsprechenden Text in einer Zielsprache” (Carstensen et al. 2010: 642)
→ Maschinelle Übersetzung = Übersetzung mittels eines Computers (Grein in Krenz und Ramlow 2008:20)

“Ein Maschinelles Übersetzungssystem ist ein Computerprogramm, das *weitgehend selbstständig* maschinenlesbare Texte von einer Ausgangssprache in eine Zielsprache übersetzt.” (Mügge 2002: 102)

- MÜ vs. CAT
- HAMT vs. MAHT

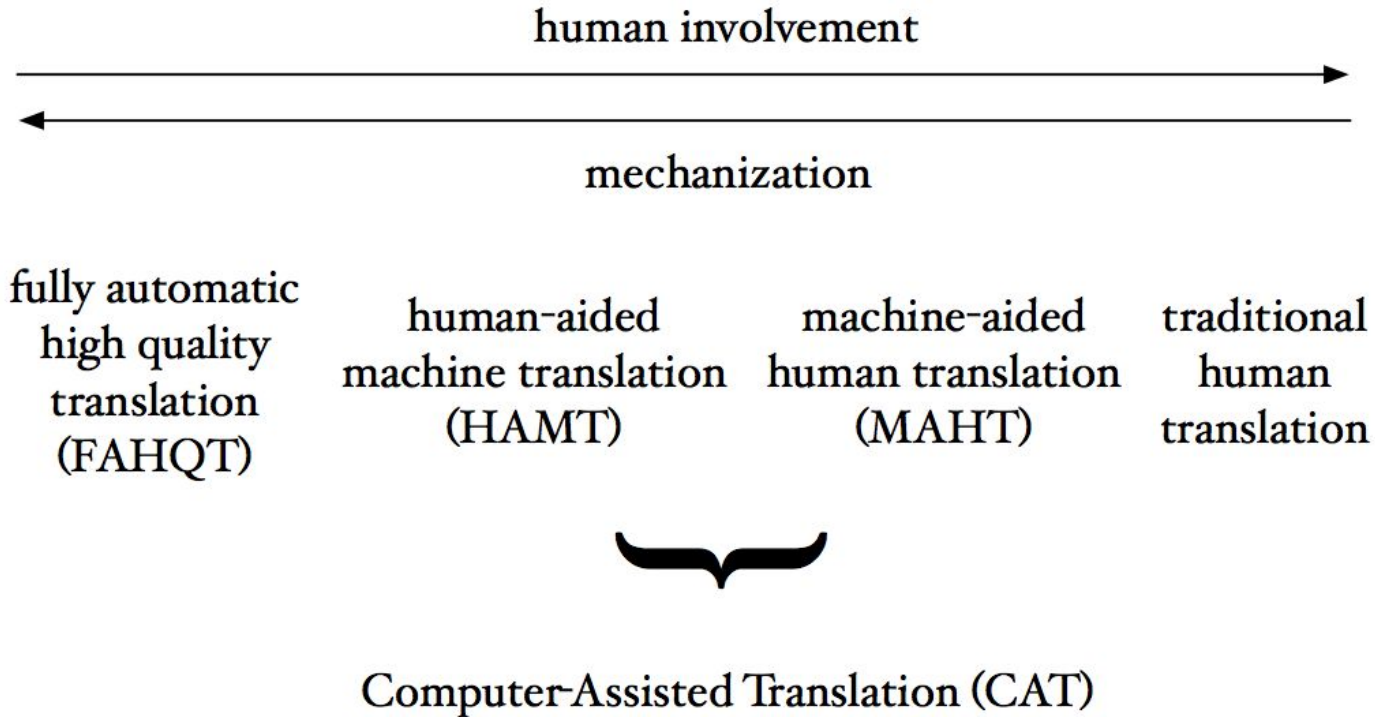


Abb. "Automatisierungsgrade nach Hutchins und Somers 1992" auf

<https://files.ifi.uzh.ch/cl/siclemat/lehre/hs10/ecl1/script/html/scripthse43.html#x63-4000010.2.3>

Nutzen

→ Ziel: FAHQT = Fully automatic high quality translation (Mügge 2002:101)
/ Fully accurate high quality translation (Carstensen et al. 2010: 643)

- Institutionen (internationale Institutionen, Unternehmen, Übersetzungsagenturen)
- Privat
- Öffentlich (mehrsprachige Informations- und Auskunftssysteme)
- Online-Übersetzungen, Maschinelles Dolmetschen

Gründe:

- Zeit (100.000 bis über 1.000.000 Wörter pro Stunde MÜ vs. 3000 Wörter pro Tag HÜ) und Kostenfaktoren
- sozial-politischer Faktor
- kommerzieller Faktor
- wissenschaftlicher Faktor
- philosophischer Faktor

Geschichte

- erste Überlegungen können bis in 17. Jh zurückverfolgt werden
- Ursprünglich handelte es sich um die Wissenschaft einer Universalsprache (Religion & Weltfrieden)

- Beginn 1933 - Petr Smirnov-Troyanskii- Patent auf Übersetzungsmaschine
- 1942 - Entwicklung des ersten Computers an der Harvard University
→ spezifische Ideen zur Umsetzung Automatischer Übersetzungssysteme
- 1948 - Vorstellung eines Übersetzungssystems im Versuchsstadium von RH Richens und Andrew Donald Booth
- 1949 - Memo von Warren Weaver (auch als Weaver Memorandum bekannt)
→ weltweit weitere Forschungen und Arbeiten in diesem Gebiet
- 1951 - Bar Hillel als zuständiger Forscher berufen; 1952 - Kritik; Aussprache für CAT mit Prä- und Postedition
- 1954 - Vorstellung eines MÜ Prototyps Russisch - Englisch durch Zusammenarbeit von Dostert und IBM
- 1964 - Gründung Automatic Language Processing Advisory Committee (ALPAC)
 - Prüfung der Kriterien: Einsatznotwendigkeiten, schnelle Verfügbarkeit, schnelle Kostenreduzierung, deutliche Leistungsverbesserung
 - 1966 - ALPAC Report - Förderung der CL und KI

- in USA, GB und Japan, wie auch Sowjetunion wurde die Forschung eingestellt
- Verlagerung der Forschung andernorts wie Deutschland, Frankreich und Kanada
 - Zweisprachigkeit Kanadas, Europäischen Gemeinschaft

Geschichte

- erste Überlegungen können bis in 17. Jh zurückverfolgt werden
- Ursprünglich handelte es sich um die Wissenschaft einer Universalsprache (Religion & Weltfrieden)

- Beginn 1933 - Petr Smirnov-Troyanskii - Patent auf Übersetzungsmaschine
- 1942 - Entwicklung des ersten Computers an der Harvard University
 - spezifische Ideen zur Umsetzung Automatischer Übersetzungssysteme
- 1948 - Vorstellung eines Übersetzungssystems im Versuchsstadium von RH Richens und Andrew Donald Booth
- 1949 - Memo von Warren Weaver (auch als Weaver Memorandum bekannt)
 - weltweit weitere Forschungen und Arbeiten in diesem Gebiet
- 1951 - Bar Hillel als zuständiger Forscher berufen; 1952 - Kritik; Aussprache für CAT mit Prä- und Postedition
- 1954 - Vorstellung eines MÜ Prototyps Russisch - Englisch durch Zusammenarbeit von Dostert und IBM
- 1964 - Gründung Automatic Language Processing Advisory Committee (ALPAC)
 - Prüfung der Kriterien: Einsatznotwendigkeiten, schnelle Verfügbarkeit, schnelle Kostenreduzierung, deutliche Leistungsverbesserung
 - 1966 - ALPAC Report - Förderung der CL und KI

- in USA, GB und Japan, wie auch Sowjetunion wurde die Forschung eingestellt
- Verlagerung der Forschung andernorts wie Deutschland, Frankreich und Kanada
 - Zweisprachigkeit Kanadas, Europäischen Gemeinschaft

Geschichte

- 70er: MÜ Forschung vor allem spezifische Arbeitsbereiche und Sachgebiete statt beliebige Texte
- erst direkte Übersetzung, dann Interlingua, dann Transfer-basierte
- seit Mitte 70er Jahre: Konzentration auf drei Bereiche:
 - Entwicklung von transferbasierten Systemen mit Einbezug der Erfahrungen von früheren Interlinguasystemen
 - Entwicklung neuartiger Interlingua-basierter Systeme
 - Einbeziehung der KI Forschung in MÜ
- 1976 - METEO; Kanada, Wetterberichte
- 1976 - Erwerb Systran von EG für EN-FR
- 1982 - Beginn des EUROTRA Projektes der EG
- 1983 - Beginn von wissensbasierten MÜ in den USA
- Japan: Entwicklung von direkte und transfer-basierte Systeme für JAP und EN
- METAL DE - EN; leistungsfähigste kommerzielle Übersetzungssystem

→ zunehmende Kommerzialisierung

- 1988 - Vorstellung des statistisch basierten Ansatzes von IBM Wissenschaftler Peter Brown
 - Aufschwung durch technischen Fortschritt in Computerentwicklung und Datenspeicherung wie auch Verfügbarkeit von Texten in digitaler Form bedingt
- ab Ende 80er Diversifizierung der Forschungsaktivitäten
- 90er: Maschinelles Dolmetschen Bsp. Verbmobil

Gegenwart

- Regelbasierte und statistische Übersetzungen galten demnach lange Zeit als Konkurrenten
- angestrebte 4. Generation: hybrides System

- Zielgruppen:
 - Sprachindustrie
 - Lokalisierungsindustrie
 - international Unternehmen
 - Entwickler und Anwender

- Anwendungsbeispiele: Google und Microsoft

Methoden

→ Intellektuelle MÜ

- Constraint-basierte MÜ
- Wissensbasierte MÜ
- Lexikalisch basierte MÜ
- Regelbasierte MÜ
- Prinzipien-basierte MÜ
- Shake and Bake MÜ

→ Korpus-basierte MÜ

- Beispiel-basierte MÜ
- Statistisch-basierte MÜ

→ hybride MÜ

- unidirektionale vs. bidirektionale MÜS

Regelbasierte MÜ

- engl. Rule-Based Machine Translation - RBMT
- „Maschinelle Übersetzung (MÜ), oder genauer gesagt regelbasierte Maschinelle Übersetzung (RBMÜ) bezeichnet die Analyse eines Textes anhand **linguistischer Regeln** und die Übersetzung in eine andere Sprache mit Hilfe eines komplexen Computerprogramms.“

(<https://www.dicits.com/de/technologie/faq/wie-funktioniert-mt.html>)

- direkte vs. indirekte (Transfer, Interlingua)
 - indirekte werden auch als Strategien der 2. Generation bezeichnet (Eberle 2008 in Werthmann und Witt 2014: 88)
- Regelbasierte Verfahren beruhen auf manuell erstellten zweisprachigen Wörterbüchern sowie lexikalischen und grammatischen Entsprechungsregeln.

Grade eines RBMT-Systems

- Ergebnisse reichen von kurios bis nützlich
- Abhängigkeiten (Sprachpaar, Fachsprachenunterstützung, allgemeinsprachliches System etc.)

Analyse → Transfer → Synthese

- Direkte Übersetzung
- Transferübersetzung
- Interlingua Übersetzung

Direkte Übersetzung

- Wort für Wort Übersetzung
- keine oder nur minimale strukturelle und semantische Analyse der Quellsprache (Stein 2009 in Werthmann und Witt 2014: 88)

Abb. 2: Schema für direktes MÜ-System (Carstensen et al. 2004, 566)



Abbildung von Werthmann und Witt 20144: 89

Transfer-Ansatz

- Analyse
 - quellsprachliche Eingabe geparkt, ansatzweise semantisch analysiert
- Verwendung von unifikationsbasierten Grammatikformalismen wie Head-Driven Phrase Structure Grammar (HPSG) oder Lexical Functional Grammar (LFG)
- Transfer
 - quellsprachliche Repräsentation durch Anwendung von Transferregeln in abstrakte zielsprachliche Repräsentation abgebildet
 - Transferregeln beschreiben Übersetzungsentsprechungen, die in ihrer Verwendung auf sinnvolle Kontexte eingeschränkt werden; Kontextbeschreibungen kann die Qualität der Übersetzung beeinflussen
 - entspricht dem klassischen regelbasierten Ansatz in konventionellen Systemen

Beispiel:

E-N1 E-N2 → F-N2 de F-N1 (installation configuration - La configuration d'installation)

Transfer-Ansatz

- Generierung
 - Das Generierungsmodul hat die Aufgabe, aus der Strukturbeschreibung (zielsprachliche Repräsentation) einen wohlgeformten Satz in der Zielsprache zu erzeugen, indem es den Regeln entsprechend Wörter beugt und deren Stellung im Satz anpasst, sprich: es wird eine natürlichsprachliche Ausgabe erzeugt

Nachteil: Für Integration neuer Sprachen ins System oder eine Änderung der Übersetzungsrichtung müssen mehrere neue Komponente erstellt werden.



Abbildung 5.5: Transfer-basierte Übersetzung

Interlingua

- Verwendung sprachunabhängiger Zwischenrepräsentationen, die durch eine Analyse der Eingabe erhalten wird, und selbst als Eingabe für eine zielsprachliche Textgenerierung dient. (Carstensen et al. 2010: 646)
- Repräsentation neutral bzgl. aller im System verarbeiteten Sprachen; erlauben daher keinen direkten Bezug mehr zwischen Quell- und Zielsprache.

BSP: Guten Tag, Herr Prof. Klabunde
→ <Greet>; <Title>; <name:Klabunde>

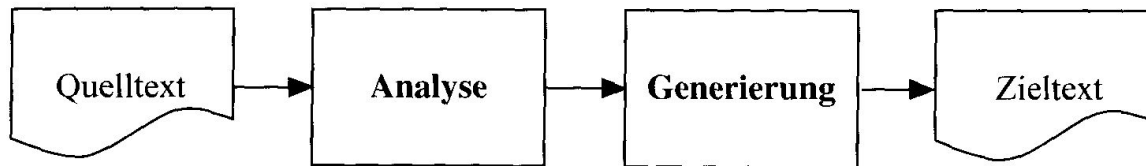


Abbildung 5.6: Interlingua-basierte Übersetzung

Abb. 1: MÜ Vauquois' Dreieck der MÜ nach Hutchins/Somers (1992, 107)

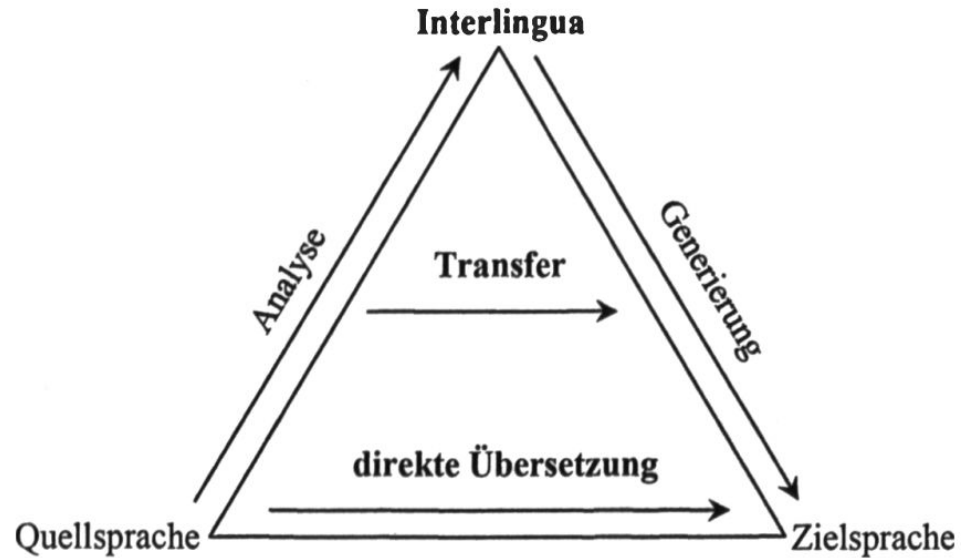


Abbildung von Werthmann und Witt 2014: 87

Vergleich der RBMT-Verfahren

X LIKE Y → Y GUSTAR X	Direkt
gustar (SUBJ (ARG1:NP), OBJ1 (ARG2:PREP)) → like (SUBJ (ARG2:NP), OBJ1 (ARG1:NP))	Transfer
<i>like/gustar</i> : [CAUSE (X, [BE (Y, [PLEASED])])]]	Interlingua

Tabelle 12.3: **Verfahren der MÜ im Vergleich**

Abb. aus Carstensen 2017: 192

Probleme der Regelbasierten MÜ

- in erster Linie weniger die angewendete Technologie, viel eher Komplexität von natürlicher Sprache
- “Je ähnlicher die Sprachen einander sind und je weniger komplex die Sätze der Ausgangssprache sind, desto besser ist meist die Qualität der Übersetzung” (Werthmann und Witt 2014: 81)
- Bedeutung und Kontext eines Textes muss maschinell erfasst werden, um adäquat zu übersetzen (Carstensen et al. 2010: 644)
- kulturelle Anforderungen (Mügge 2002:113ff) zB. Maßangaben, Dezimale, Zeit- und Datumsangaben

- Ambiguitäten

Problematik der Ambiguität

- lexikalische Ambiguität
 - Um welche Wortart handelt es sich?
- syntaktische Ambiguität
 - Anhänge (z.B. PP-attachment) in die syntaktische Struktur --- “I saw the man with a telescope”
- (komplexe) semantische Ambiguität
 - (Homografie, Metonymie) Homografie, Polysemie
- (komplexe) kontextuelle Ambiguität
 - (Rekurrierter) Kontext bei unterschiedlichen Lesarten

- Translation mismatches, Divergenzen

Vor- und Nachteile:

Vorteile:

- einzelne Fehlerquellen sind auch später noch gut auszumachen
- auch kleinere Sprachen sind analysierbar, da kein größtmögliches Korpora für Erfolge gebraucht wird
- Zusammengehörige Sprachbestandteile werden richtig zugeordnet

Nachteile:

- deutlich höherer Zeit- und damit dementsprechender Kostenaufwand
- Ambiguitäten, Bewertungen, Lernerfolge ect.

Fazit

→ Durch die Zunahme der elektronischen Verfügbarkeit und den spürbar steigenden Bedarf an Übersetzungen in der globalisierten Welt nimmt die Bedeutung solcher integrierender Verfahren in der Zukunft ganz zweifellos weiter zu.

Beispiele

- Langenscheidts T1
- linguattec Personal Translator <https://www.linguatec.de/personal-translator-demo/>
- Systran
- Apertium

Literatur und Quellen

- Carstensen, Kai-Uwe, Ebert, Christian, Ebert, Cornelia, Jekat, Susanne, Klabunde, Ralf, Langer, Hagen. 2010. Computerlinguistik und Sprachtechnologie: Eine Einführung. Springer Spektrum Akademischer Verlag: Heidelberg.
https://books.google.de/books?hl=de&lr=&id=OioqBAAAQBAJ&oi=fnd&pg=PA1&dq=carstensen+et+al+2009&ots=xo_-LC788H&sig=isc6rOko1YUEVfyl-I93QBcamnk#v=onepage&q=carstensen%20et%20al%202009&f=false
- Carstensen, Kai-Uwe. 2017. Sprachtechnologie. Ein Überblick. <http://kai-uwe-carstensen.de/Publikationen/Sprachtechnologie.pdf>
- Krenz, Michael, Ramlow, Markus. 2008. Maschinelle Übersetzung und XML im Übersetzungsprozess: Prozesse der Translation und Lokalisierung im Wandel. Frank & Timme Verlag: Berlin.
https://books.google.de/books?hl=de&lr=&id=2Lhr6QyYcqlC&oi=fnd&pg=PA15&dq=regelbasierte+maschinelle+übersetzung&ots=HI7WVuZjgr&sig=hq388YeqYtr1BW1ZBYh_JCFbpWg#v=onepage&q=regelbasierte%20maschinelle%20übersetzung&f=false
- Muegge, Uwe. 2002. Lokalisierung und Maschinelle Übersetzungssysteme. *Lokalisierung von Technischer Dokumentation*. Stuttgart.
http://works.bepress.com/uwe_muegge/27/
- Seewald-Heeg, Uta, Stein, Daniel. 2009. Maschinelle Übersetzung – von der Theorie zur Anwendung. JLCL – Volume 24 – Number 3 – 2009.
<https://jcl.org/content/2-allissues/15-Heft3-2009/ldv-forum-24-3.pdf#page=8>
- Stampf, Judith (2012) Maschinelle Übersetzung. Masterarbeit. <https://othes.univie.ac.at/18315/>
- Trujillo, Arturo. 1999. Translation Engines: Techniques for Machine Translation. Springer Verlag: Berlin Heidelberg.
https://books.google.de/books?hl=de&lr=&id=VpgRBwAAQBAJ&oi=fnd&pg=PA3&dq=arnold+1994+machine+translation&ots=NRHMkWDsGh&sig=NabFTJ_pSAa6LJScKik_9lpwA8k#v=onepage&q=arnold%201994%20machine%20translation&f=false
- Werthmann, Antonina, Witt, Andreas : Maschinelle Übersetzung - Gegenwart und Perspektiven Erschienen in: Stickel, Gerhard (Hrsg.): Translation and Interpretation in Europe. Contributions to the Annual Conference 2013 of EFNIL in Vilnius. - Frankfurt am Main/Berlin/Bern/Bruxelles/New York/Oxford/Wien: Lang, 2014. S. 79-103. (Duisburger Arbeiten zur Sprach- und Kulturwissenschaft 102)
https://ids-pub.bsz-bw.de/frontdoor/deliver/index/docId/4487/file/Werthmann_Witt_Maschinelle_Uebersetzung_Gegenwart_und_Perspektiven_2014.pdf
- <https://files.ifi.uzh.ch/cl/sicemat/lehre/hs10/ec1/script/html/scripthse43.html#x63-4000010.2.3> [letzter Zugriff: 30.06.2020]
- <https://www.dicits.com/de/technologie/faq/wie-funktioniert-mt.html> [letzter Zugriff: 30.06.2020]
- https://en.wikipedia.org/wiki/Machine_translation#:~:text=Machine%20translation%2C%20sometimes%20referred%20to,speech%20from%20one%20language%20to [letzter Zugriff: 30.06.2020]
- <https://www.linguattec.de/personal-translator-demo/> [letzter Zugriff: 30.06.2020]
- https://en.wikipedia.org/wiki/Machine_translation#:~:text=Machine%20translation%2C%20sometimes%20referred%20to,speech%20from%20one%20language%20to [letzter Zugriff: 30.06.2020]
- https://de.wikipedia.org/wiki/Maschinelle_Übersetzung#cite_ref-19 [letzter Zugriff: 30.06.2020]
- <http://kurt-eberle.de/downloads/transkoethenCompact.pdf> [letzter Zugriff: 30.06.2020]

Vielen Dank für Eure Aufmerksamkeit!