

2. Übersicht über die Entwicklung der maschinellen Sprachübersetzung

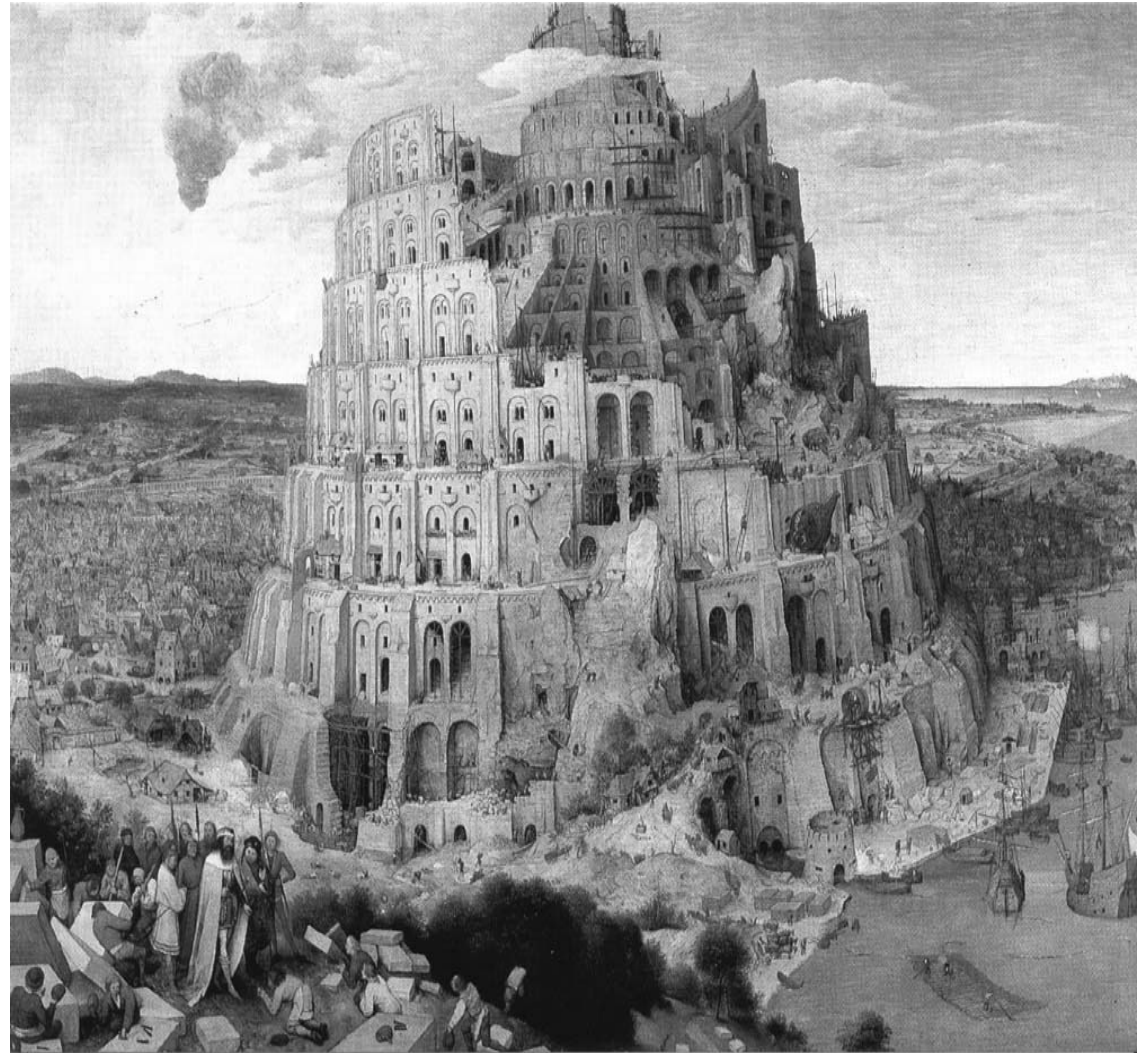
Sehnsucht der Menschen nach Überwindung der Sprachbarrieren:

biblisch/mythologisch: Turmbau zu Babel und Pfingstwunder

wissenschaftlich: Universalsprachen und universale Kunstsprachen wie Esperanto; Einsatz von Technologie

Biblisch/Mythologisch

Die babylonische Sprachverwirrung - *confusio linguarum* - als Strafe Gottes für menschlichen Hochmut:



Pfingstwunder als Ausdruck der Sehnsucht nach Überwindung der Sprachbarrieren.

Neuerer Mythos: **Babelfish** - Internet-Suchmaschine AltaVista

Herkunft: Douglas Adams: The Hitchhiker's guide to the Galaxy (dt.: „Per Anhalter durch die Galaxis“). 1979.

Protagonist des Romans Arthur Dent zu seinem Begleiter Ford Prefect:

„Was macht eigentlich“, „Was macht eigentlich der Fisch in meinem Ohr?“

„Er übersetzt für dich. Er ist ein Babelfisch. Schlag‘s im Buch nach, wenn‘s dir Spaß macht.“

„Der Babelfisch“, ließ der Reiseführer Per Anhalter durch die Galaxis mit ruhiger Stimme vernehmen, „ist klein, gelb und blutegelartig und wahrscheinlich das Eigentümlichste, was es im ganzen Universum gibt. Er lebt von Gehirnströmen, die er nicht seinem jeweiligen Wirt, sondern seiner Umgebung entzieht. Er nimmt alle unbewußten Denkfrequenzen dieser Gehirnströme auf und ernährt sich von ihnen. Dann scheidet er ins Gehirn seines Wirts eine telepathische Matrix aus, die sich aus den bewußten Denkfrequenzen und Nervensignalen der Sprachzentren des Gehirn zusammensetzt. Der praktische Nutzeffekt der Sache ist, dass man mit einem Babelfisch im Ohr augenblicklich alles versteht, was einem in irgendeiner Sprache gesagt wird. Die Sprachmuster, die man hört, werden durch die Gehirnstrommatrix entschlüsselt, die einem der Babelfisch ins Gehirn eingegeben hat.“

Nun ist es aber verdammt unwahrscheinlich, dass sich etwas so wahnsinnig Nützliches rein zufällig entwickelt haben sollte, und so sind ein paar Denker zu dem Schluß gelangt, der Babelfisch sei ein letzter und entscheidender Beweis dafür, dass Gott nicht existiert....." (63/64). Mittlerweise hat der arme Babelfisch dadurch, dass er alle Verständigungsbarrieren zwischen den verschiedenen Völkern und Kulturen niederriß, mehr und blutigere Kriege auf dem Gewissen als sonst jemand in der ganzen Geschichte der Schöpfung".

Wichtiger Aspekt der Übersetzung ist hier angesprochen:

Babelfisch ermöglicht das **Verstehen** der verschiedenen Wesen, die auf dem Weg durch die Galaxis angetroffen werden. ‚Verstehen‘ ist hier nicht das bloß akustische Verstehen, sondern das **inhaltliche** Verstehen.

Wissenschaftlich

In der Menschheitsgeschichte zwei Lösungsvorschläge, beide ebenso faszinierend, wie utopisch:

- Erfindung einer vollkommenen Sprache
- Übersetzungsmaschine - die mechanistisch-technologische Lösung

Erfindung vollkommener Sprachen - Universalsprachen

Ursprung: Mangel an einer gemeinsamen Sprache im Umgang der Menschen zwischen den Nationen und Völkern wurde zu allen Zeiten als großes und lästiges Hindernis empfunden

Überwindung der *confusio linguarum* als **Fascinosa** der Wissenschaft.

In der europäischen Geistesgeschichte Hunderte von Versuchen, ideale Sprachen, universale, internationale Kunstsprachen zu entwickeln.

Übersicht:

Umberto Eco: La ricerca della lingua perfetta nella cultura europea. 1993.
Dt. Übersetzung 1994 u.d.T. „Die Suche nach der vollkommenen Sprache“

Seit den Griechen Entwicklung idealer Sprachen zum Zwecke der mechanischen Lösung der Probleme der linguistischen Analyse.

Dante „De vulgari eloquentia“

Streben nach eindeutiger, einzelsprachenunabhängiger Darstellung in der Philosophiegeschichte, besonders in der

Philosophie des **Rationalismus**:

Lenders: Universalgrammatiken des Rationalismus. In: Grundzüge der Literatur- und Sprachwissenschaft Bd. 2: Sprachwissenschaft. dtv München 1974, 105-117.

Problem der Unabhängigkeit von den Einzelsprachen

Grammatik behandelt Regeln der Sprache, Logik Regeln des Denkens.

„La Logique ou l'art de penser“ - „Grammaire générale ou l'art de parler“ (Grammatik und Logik von Port Royal).

Regeln des Denkens sind für alle Menschen gleich.

Entwicklung einer künstlichen Sprache, deren Regeln mit denen des Denkens zusammenfallen sollte, von allen Widersprüchen und Mehrdeutigkeiten frei, leicht verstehbar und ein universelles Instrument der Mitteilung vor allem wissenschaftlicher Erkenntnisse sein sollte:

Idee der *lingua universalis, lingua rationalis*

Vorläufer der rationalistischen Philosophie:

Raymundus Lullus (1235-1315):

Gedanke, die Grundlagen des menschlichen Wissens durch Zeichen einer künstlichen Sprache darzustellen und aus diesen Grundlagen durch kombinatorische Verfahren, die sog. Lullische Kunst (*ars magna; ars Lulliana*), komplexere Erkenntnisse abzuleiten.

Idee einer perfekten philosophischen sprache.

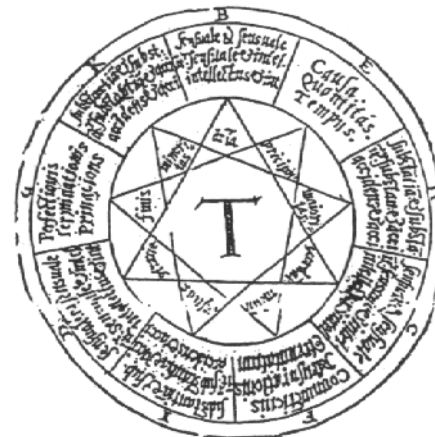
Entwicklung abstrakter Zeichensysteme in der Mathematik und symbolische Logik.

TAFEL DER PRINZIPIEN

	PRINCIPIA ABSOLUTA	PRINCIPIA RELATIVA	QUESTIONES	SUBJECTA	VIRTUTES	VITIA
B	Bonitas	Differentia	Utrum?	Deus	Iustitia	Avaritia
C	Magnitudo	Concordantia	Quid?	Angelus	Prudentia	Gula
D	Aeternitas	Contrarietas	De quo?	Coelum	Fortitudo	Luxuria
E	Potestas	Principium	Quare?	Homo	Temperantia	Superbia
F	Sapientia	Medium	Quantum	Imaginatio	Fides	Acidia
G	Voluntas	Finis	Quale?	Sensitiva	Spes	Invidia
H	Virtus	Majoritas	Quando?	Vegetativa	Charitas	Ira
I	Veritas	Aequalitas	Ubi?	Elementativa	Patientia	Mendacium
K	Gloria	Minoritas	Quomodo? Cum quo?	Instrumentativa	Pietas	Inconstantia



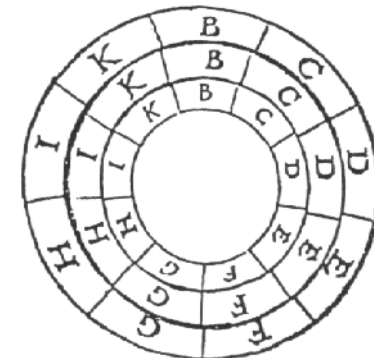
Erste Figur



Zweite Figur



Dritte Figur



Vierte Figur

Figur 1: Die Lullischen Prinzipien und Figuren

George Dalgarno (1626-1687):

Ars signorum: Entwicklung eines Alphabets universeller Charaktere.

John Wilkins:

„An Essay towards a Real Character and a Philosophical Language“,
1668 (auch bei Hutchins genannt).

Gottfried Wilhelm Leibniz (1646-1716):

bedeutendster Verfechter einer *lingua universalis, grammatica rationalis*.

Entwurf einer *ars characteristica universalis*.

De la Mettrie: *L'homme machine*, 1748.

Tendenz zur ‚Mechanisierung‘, Sicht des Menschen als Maschine (materialistische Philosophie) in

Übersetzungsmaschine - die mechanistisch-technologische Lösung

Seit der Antike: zahllose Mythen von Maschinenmenschen, in denen der Mensch mit all seinen Fähigkeiten mechanisch simuliert wurde.

Philosophien, die den materialen, materialistischen Charakter des Menschen verkündeten.

Im Kontext der rationalistischen Philosophie:

materialistische Auffassung vom Menschen als Maschine:

De La Mettrie: „L‘homme machine“

Rationalistische Tradition: Erfindung des Binärsystems durch Leibniz und Forderung nach lingua universalis durch Leibniz.

Erfindung von ‚Automaten‘ und Maschinenmenschen, die Fähigkeiten wie musizieren, tanzen, essen, ja sogar Sprechen simulierten.

Übersetzungsmaschine aus früheren Jahrhunderten nicht bekannt, erst nach Entwicklung der programmierbaren Computer.

Übersetzungsmaschine - Aufsehen in den Medien - unrealistische Versprechungen

Beispiele:

Pressemeldung 1993 mit Bezug auf den Einsatz der MÜ bei der Kommission der Europäischen Gemeinschaft „Per Drehknopf durch die Sprachen wandern“

dpa 1993 „Der Computer machts möglich: jetzt können auch Menschen miteinander telefonieren, die verschiedene Sprachen sprechen und sich eigentlich gar nicht verstehen. Forscher aus Deutschland, Japan und den USA haben gestern zum ersten Mal mit Hilfe einer solchen automatischen Übersetzung telefoniert. Dabei wurden die Worte des jeweiligen Sprechers von einem Computer simultan übersetzt. Das System soll nach der Jahrtausendwende für internationale Konferenzen oder die Buchung von Hotels und Flügen zur Verfügung stehen.“ (Dpa, 29.1.1993).

Entwicklung seit 1945

Wichtige Einflüsse zur statistischen Analyse von Sprache kamen von Seiten der Telegraphie und der Informationstheorie (Shannon, Weaver)

Entwicklung der MÜ aus linguistischer Sicht

Erste Phase 1945 - 1967: Die euphorische Phase

a) 1945-1954 Wort-für-Wort-Übersetzung

1946 in Gesprächen zwischen Andres D. Booth (engl.) und Warren Weaver (Präs. der Rockefeller Foundation)

1947: Booth und Britten (Institute for Advances Studies, Princeton):
Wörterbuchansatz

1948: Erste Überlegungen von Booth und Richens zur Konstruktion eines Wörterbuch, das auf einem Computer anwendbar sein sollte und sich von gewöhnlichen Wörterbüchern unterschied.

Lit.: Booth, A.D. and Richens, R.H.: Machine Translation of Languages. New York 1955, 24-26: in das Wörterbuch nicht ganze Wörter, sondern nur Wurzeln und Endungen aufnehmen.

1949: Memorandum von Warren Weaver: "Translation". New York, July 1949.

beeinflusste die öffentliche Meinung über MÜ stark

1950: Beginn des Erscheinens des ersten Fachorgans, "MT", mit dem Artikel von E. Reifler "Studies in Mechanical Translation", No. 1, M.T., Mimeographed, Washington 1950.

Forderung nach Pre- und Post-editor.

Erster richtiger Zeitschriftenartikel:

Victor Oswald und Stuart L. Fletcher (MÜ-Gruppe an UCLA):
"Proposals for the Mechanical Resolution of German Syntax Patterns.
In: Modern Language Forum 36 (1951), 1-24.

Hier (nach Hutchins): Erstmals die Forderung, syntaktische Aspekte zu untersuchen.

Vermutung, dass Wort-für-Wort-Übersetzung aus dem Engl. ins Deutsche z.B. nicht gehe, wegen der anderen Stellung der Wörter im Satz als im Engl.

1951: Forschungsbericht von Bar-Hillel

Keine semantisch gestützte Übersetzung in absehbarer Zeit, daher keine Vollautomat. Übers.

Wohl "mixed MT" ---> postediting

Morphologische Analyse machbar, auch die Übersetzung eines Satzes in seinen anderen, logisch äquivalenten.

Empfehlung, MÜ auf Gebiete mit "restricted vocabulary" zu begrenzen.

1952: Konferenz am MIT. Ergebnis dieser Konferenz war u.a. Notwendigkeit von

- Untersuchungen zur Worthäufigkeit und Wortbedeutung und
- Untersuchungen zur operationalen Analyse und Syntax, so dass man auf Maschinen anwendbare Regeln erhält.

Harper und Oettinger beschäftigten sich mit dem Russischen
Oswald und Lawson: Untersuchungen zur Worthäufigkeit
Yngve arbeitete über die Nützlichkeit von teilweisen Übersetzungen
Booth forderte mehr Zeit zur Prüfung der Möglichkeiten von
Wörterbüchern.

Gruppe am MIT, unter der Leitung von Bar-Hillel, seit 1953 Yngve.
Gruppe unter Leitung von Leon Dostert seit 1952 an der Georgetown
University, Washington.

1954: Georgetown-IBM-Experiment (Russisch-Englisch)
Vorführung ausgewählter 49 Sätze aus dem Gebiet der Chemie.
Wörterbuch und 6 Regeln. Linguistische Ausführung: Garvin.
Einzelheiten bei Hutchins, 71ff.

Das Georgetown-Projekt wurde in 4 Teilprojekte geteilt, eines davon GAT (SERNA), das später in Systran aufging.

1963 endete das Georgetown-Projekt. Ein Ableger wurde in Ispra bei EURATOM installiert.

Nach dem Georgetown-Projekt überall umfangreiche Projektförderung der MÜ in USA .

b) 1954-1967 Einführung **syntaktischer** Mechanismen

Ergänzung des Lexikons durch syntaktische Verfahren

- **Finite-State-Grammar** (Hockett)

Grammatik besteht aus endlichen Menge von Strukturen; zu Beginn eines Satzes wird mit einer gewissen Wahrscheinlichkeit ein bestimmtes Element erwartet. Ist dieses aufgetreten, so werden mit gewisser Wahrscheinlichkeit andere Elemente erwartet.

- Verwendung von **Phrasenstrukturgrammatiken** (Harris 1951; Nida/Jane Robinson): Ersetzungsregeln; Einführung kontextfreier (cf) und kontextsensitiver (cs) Regeln.

- Generative Grammatik, Chomsky 1957
- Generative Transformationsgrammatik 1965
- *Stratificational Grammar* (S. Lamb): Einführung des Strata-Prinzips
- *Dependency Grammar* (Hays 1963):
- *'fulcrum'-approach* von Garvin (Georgetown): IC-Analyse der wichtigsten Träger syntaktischer Information (fulcrum = Angelpunkt)

Gründung zahlreicher MÜ-Projekte, vor allem in USA

Zweite Phase 1968 - 1974 Die depressive Phase

Seit 1964 Beginn der Arbeiten am ALPAC-Report, der die Ergebnisse der MÜ zusammenfassen und bewerten sollte.

1966/67: ALPAC-Report (Automatic Language Processing Advisory Committee):

- *high quality fully automatic translation* (HQFAT) ist nicht möglich
- einseitig pragmatische Beurteilungsgrundlage: Preis-Leistungs-Relation
- beruht auf altem hardware-Stand (vor der IBM 360er Generation)
- keine Aussagen über die Tauglichkeit der linguistischen Konzepte
- Forderung nach Grundlagenforschung

Folgen des ALPAC-Reports:

- in USA: Einstellung der Projekte
- in Europa: Kenntnisnahme. Kauf des Systems 'Computer Concepts' durch DFG; LIMAS-Gruppe
- Abkehr vom Aspekt der Machbarkeit von MÜ
- MÜ wurde als linguistisches, theoretisches Problem erkannt
- Übernahme neuer linguistischer Theorien in die MÜ
- Besinnung auf die semantischen Grundlagen der Sprache
- Zunehmende Einflüsse der Künstlichen Intelligenz auf die MÜ

Dritte Phase 1974 - 1994: Die realistische Phase

Neuanstöße aus kommerziellem Bereich:

SYSTRAN (durchgehend seit ca. 1963 in USA am Markt) wird bei EG vorstellig

1974/5 verschiedene Vorführungen von SYSTRAN (Bonn, Luxemburg, Zürich u.a.). Ausführliche Tests von Systran. EG setzt SYSTRAN ein

seit 1974 Entwicklung des EUROTRA-Konzepts (Group Leibniz)

1982 Beginn von EUROTRA

1982 Beginn der MÜ-Förderung durch die japan. Regierung

Neue Einsichten in die Natur des Sprachverstehensprozesses;
Notwendigkeit, Semantik einzubeziehen

Motive für MÜ: Weltweiter Handel, Bedarf der EG; kommerzieller
Bedarf;

zahlreiche MÜ-Projekte in Japan, unter dem Zwang, die Sprachbarrieren
angesichts der wirtschaftlichen Expansion zu überwinden.

ab 1988 Einbeziehung der gesprochenen Sprache:

ALR in Japan (seit 1988)

Verbmobil in Deutschland (seit 1992)

Vierte Phase seit 1994: Miniaturisierung, Einbeziehung der gesprochenen Sprache und breite Vermarktung

mehrsprachige elektronische Wörterbücher als
Übersetzungshilfssysteme

Idiom-Übersetzer

Übersetzungsspeicher, Translation Manager

Integrierbarkeit in Schreibsysteme

preiswerte PC-Übersetzungssysteme (unter 1000 DM)

z.B. Systran Professional for Windows (Englisch, Spanisch, Französisch, Deutsch, Italienisch, Japanisch)

Globalink: Power Translator, viele Sprachenpaare

IBM-Personal-Translator Ptplus
(Engl./Dt.; Dt./Engl.)

T1 von Langenscheidt
(Engl./Dt.; Dt./Engl.)

versch. Japanisch-Engl. Systeme

Fünfte Phase: seit ca. 1998: Übersetzung im Internet; Erwartungen für das Internet 2 (Semantic Web)

Babelfish Übersetzungstool der Suchmaschine AltaVista

wurzelt tief in der 50-jährigen Geschichte der maschinellen Übersetzung (Systran)

Beispiel für Babelfish - Übersetzung

Literatur:

John Hutchins: Compendium of Translation Software. Commercial machine translation systems and computer-aided translation support tools. First edition: Published by the European Association for Machine Translation, April 2000. Neuauflage 2002.

Hutchins führt 290 auf dem Weltmarkt erhältlich Systeme auf, viele ausschließlich für Internet konzipiert, andere für das Internet eingerichtete Varianten umfassender Systeme

unter den 290 Systemen sind auch reine MÜ-Hilfssysteme wie Wörterbuch-Systeme

Große Zahl an Firmen, die maschinelle Übersetzungswerkzeuge anbieten: weltweit 127 mehrere ganz auf MÜ spezialisiert (SYSTRAN)

andere bieten MÜ-Tools neben anderen sprachverarbeitenden Werkzeugen an (LINGUATEC Sprachtechnologien GmbH (IBM-Ausgründung); Lernout & Houspie Speech Products.

Beeindruckendes Angebot an Sprachpaaren: überwiegend Englisch als Ausgangs- und Zielsprache (Englisch-Französisch, Englisch-Deutsch, Englisch-Spanisch, Englisch-Italienisch und Englisch-Japanisch, jeweils auch in die andere Richtung)

Aber auch erwerbbar Systeme z.B. für Arabisch, Chinesisch, Finnisch, Hebräisch, Ungarisch, Koreanisch, Russisch und Vietnamesisch.

Entwicklung der MÜ aus Sicht der Computerwissenschaft/Informatik

Bátori (in Bátori/Weber 1986, S. 5) unterscheidet drei Phasen, und zwar unter besonderer Betonung des Anteils der Computerwissenschaft:

- Computerphase
- CL- Phase
- KI-Phase

- Das Computermodell der Computerphase (nach Bátori S. 10ff):
bis ca. 1967: Dominanz von Computerspezialisten
MÜ-Systeme dieser Phase bestehen im wesentlichen aus einem Lexikon. Zum Teil Versuche einer linguistischen Fundierung, aber viele ad-hoc Lösungen.

- Die CL-Phase: anschließend an den ALPAC-Report (1967):
Ausdehnung des Syntaxbereichs: lückenlose Ermittlung der syntaktischen Struktur
Benutzung von formalen Grammatiken und Parsern; Trennung von Grammatik und Parser
Gliederung der Verarbeitung in zwei große Schritte, Analyse und Synthese; später drei Schritte: Analyse, Synthese, Transfer.
Einführung einer semantischen Repräsentationsebene
Defizite: Überschätzung der Syntax (TG-Einfluss); Vernachlässigung satzübergreifender Strukturen

- ab ca. 1974: KI-Modelle der MÜ (Bátori S. 17ff):
Allgemeine Grundzüge:

- Verlangen nach professionellen Informatikgrundlagen
- Übersetzung ist nur möglich auf der Grundlage des Verstehens eines Textes. Also: Simulation von Verstehen nötig.
- Ganzheitlicher Ansatz, d.h. alle Komponenten sind gleichrangig und stehen im Dienste eines Ziels. Keine Komponente muss um ihrer selbst willen besonders detailliert ausgearbeitet werden (z.B. Syntax)

Besondere Merkmale des KI-Paradigmas:

- keine Trennung zw. sprachl. und nicht-sprachl. Wissen. Hinein-
nahme von Alltagswissen
- keine Separierung von Syntax und Semantik
- Ergebnisse der Textlinguistik und Diskursanalyse werden ein-
bezogen
- Transferkomponente als eigentlicher Kern des Übersetzungspro-
zesses erkannt
- Synthese wird nicht mehr als trivial angenommen

Übersetzungsvorgang ist von vorneherein Teil eines Mensch-Maschine-Kommunikationssystems (Systemverbundes)