

Part I.
Sprachtheorie

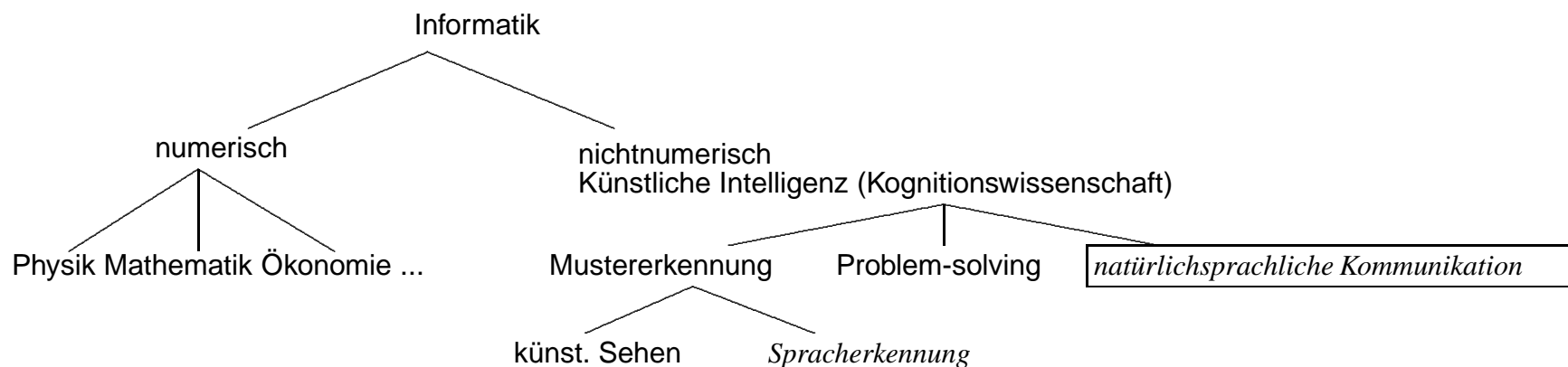
1. Computerbasierte Sprachanalyse

1.1 Mensch-Maschine-Kommunikation

1.1.1 Was ist Computerlinguistik?

Zwei Antworten

- Linguistische Analyse (Philologie) mit Computern
 - Electronische Speicherung großer Mengen realer Daten, z. B. Korpora und Lexika, mit leistungsfähigen Möglichkeiten der Suche und der Reorganisation (verbesserte Datensammlung).
 - Automatisches Testen linguistischer Analysen in Morphologie, Syntax und Semantik an großen Mengen natürlicher Daten (neue Verifikationsmethode)
- Teil der künstlichen Intelligenz, der sich mit natürlichsprachlicher Kommunikation befaßt



Theoretisches Ziel: ein Computermodell der natürlichsprachlichen Kommunikation (wissenschaftliche Erkenntnis).

Praktisches Ziel: Mensch-Maschinenkommunikation in natürlicher Sprachen (maximale Benutzerfreundlichkeit)

1.1.2 Restringsierte vs. nichtrestringierte Kommunikation

1.1.3 Beispiel restringierter Kommunikation: eine Datensatzstruktur

	Nachname	Vorname	Ort	...
A1	Schmidt	Peter	Bamberg	...
A2	Meyer	Susanne	Nürnberg	...
A3	Sanders	Reinhard	Schwabach	...
	:	:	:	

1.1.4 Anfrage an die Datenbank

Query:

```
select A#
where city = 'Schwabach'
```

Result:

A3 Sanders Reinhard

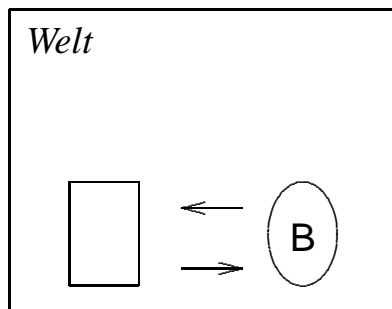
1.1.5 Klassische vs nouvelle KI

Die klassische KI analysiert intelligentes Verhalten als die Manipulation abstrakter Symbole. Ein typisches Beispiel ist ein Schachprogramm. Es arbeitet in Isolation vom Rest der Welt mit fest definierten Figuren auf einem vorgegebenen Brett. Der Suchraum für eine dynamische Gewinnstrategie ist beim Schachspiel zwar astronomisch. Da es sich jedoch um eine abgeschlossene Welt handelt, reichen die technologischen Voraussetzungen eines Standardcomputers aus.

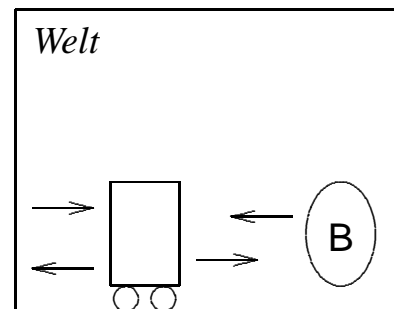
Das Ziel der nouvelle KI ist dagegen die Entwicklung selbständig agierender Roboter (*autonomous agents*). Da sich die Umwelt ständig in unvorhersehbarer Weise ändern kann, muß das System sie mit Hilfe von Sensoren kontinuierlich beobachten. Die Strategie der *task level decomposition* zerlegt den Komplex von Kognition und Verhalten in viele kleinere, problemgerechte Teilaufgaben, wobei die Folgerungsmethoden direkt auf den lokalen Wahrnehmungsdaten aufsetzen.

1.1.6 Drei Typen der Mensch-Maschine-Kommunikation

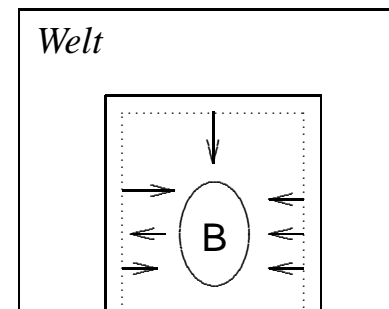
Standardcomputer



autonomer Roboter

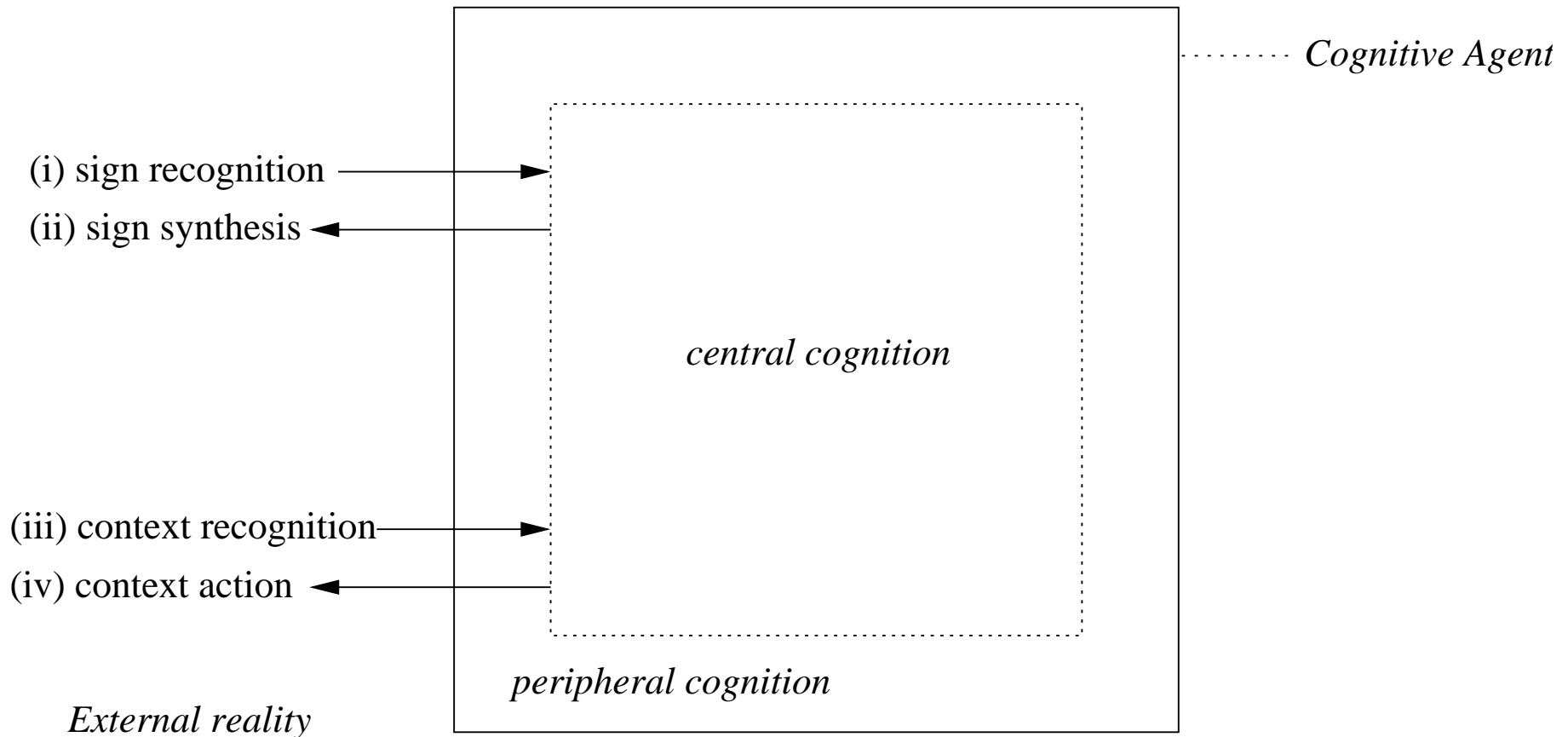


virtuelle Realität

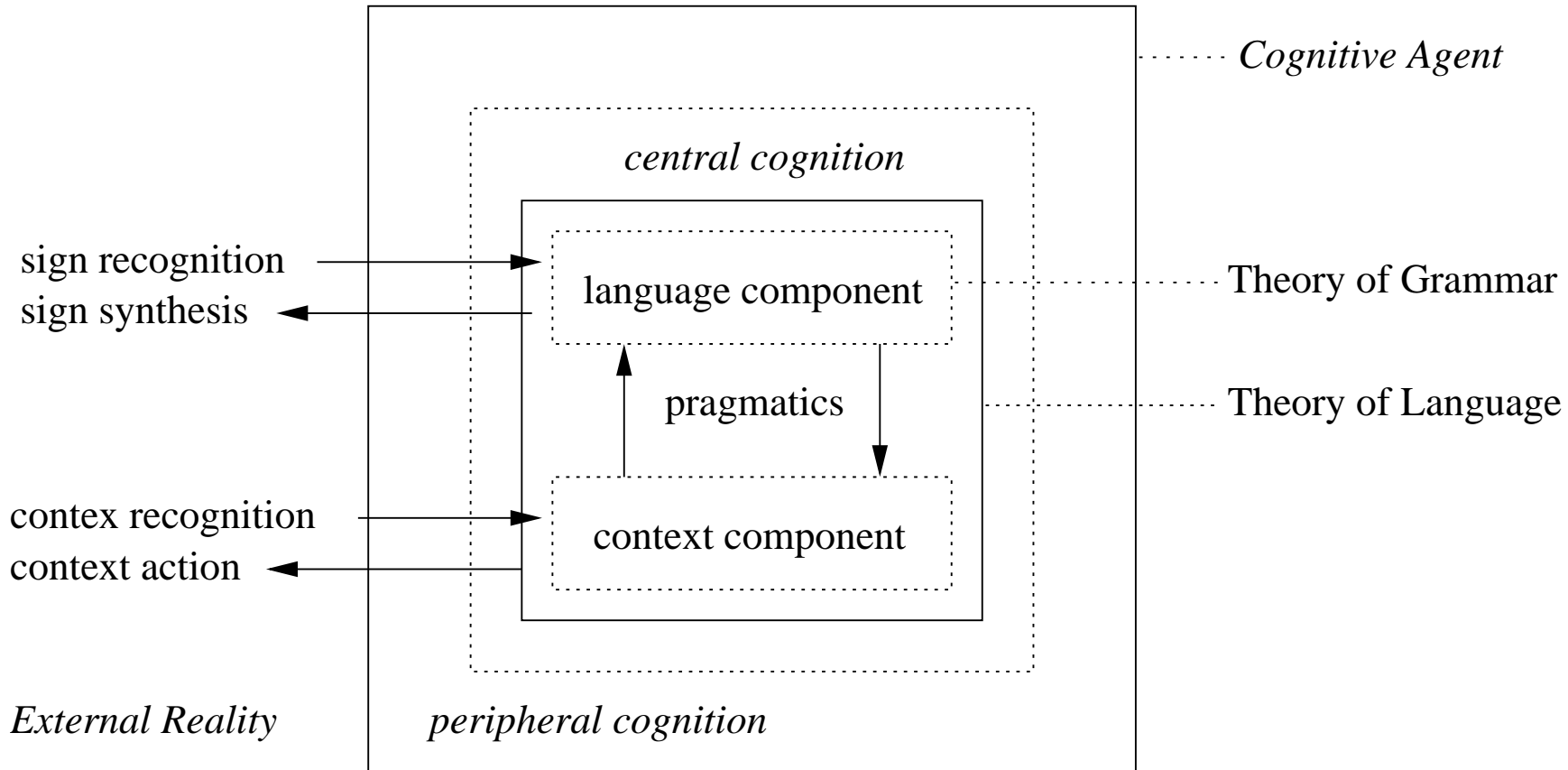


1.2 Schnittstellenbasierte Entwicklung einer Sprachtheorie

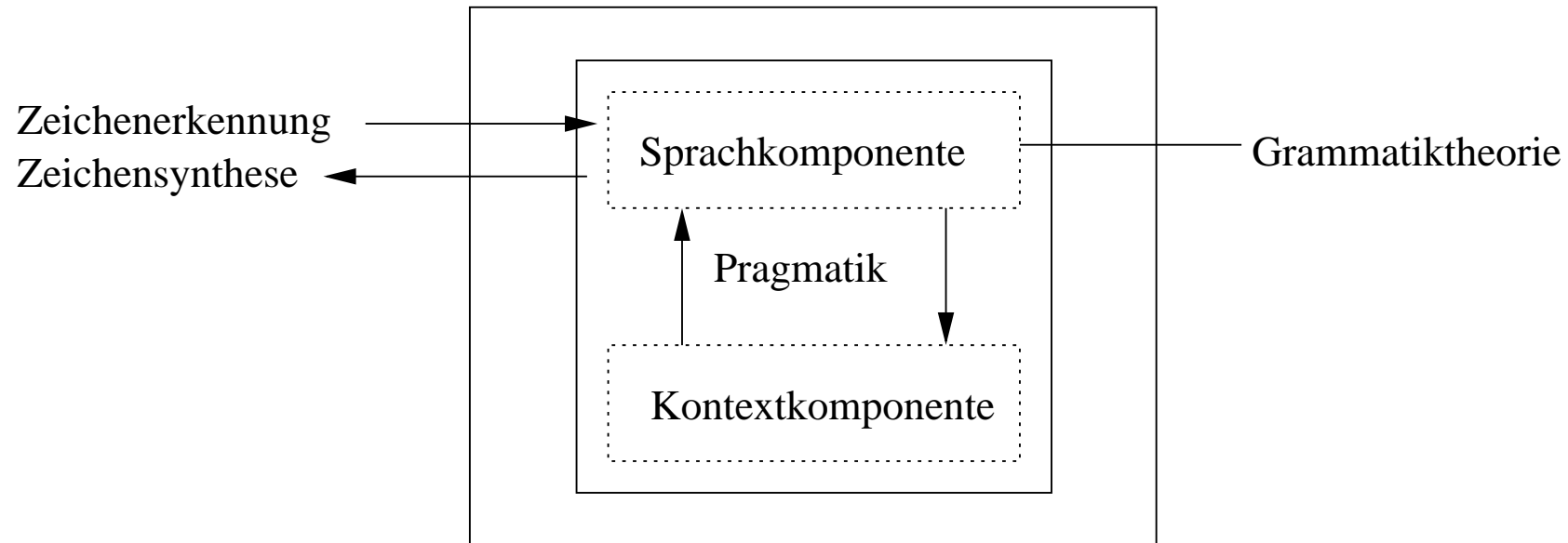
1.2.1 Schnittstellen (interfaces) eines kognitiven agenten



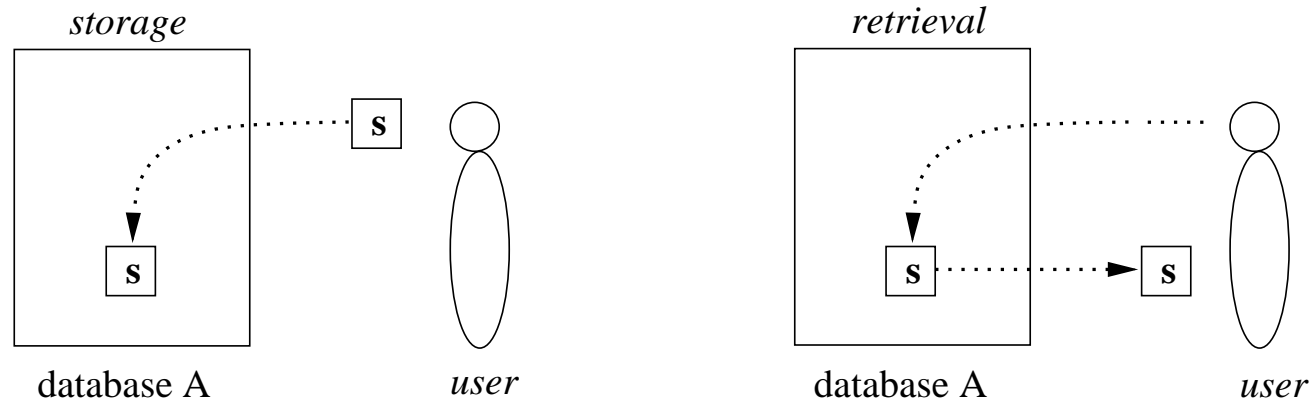
1.2.2 Sprachtheorie als Teil der zentralen Kognition



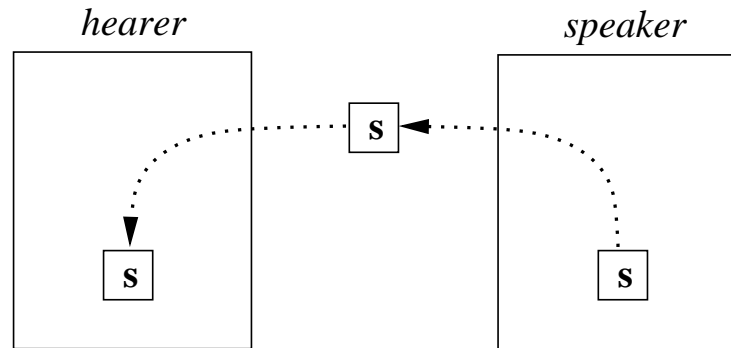
1.2.3 Komponenten einer vereinfachten Sprachtheorie



1.2.4 Interaktion mit einer konventionellen Datenbank



1.2.5 Interaction between speaker and hearer



Datenbankmetapher erfolgreicher Kommunikation

1.3 Sprachwissenschaft und ihre Komponenten

1.3.1 Varianten der Sprachwissenschaft

- *Traditionelle Grammatik*
- *Theoretische Linguistik*
- *Computerlinguistik*

1.3.2 Die Komponenten der Grammatik

- *Phonologie*: Lehre von den Sprachlauten
- *Morphologie*: Lehre von den Wortformen
- *Lexikon*: Auflistung der Wörter
- *Syntax*: Lehre von der Komposition der Wortformen
- *Semantik*: Lehre von den wörtlichen Bedeutungen
- *Pragmatik*: Lehre von den Verwendungen

1.4 Methoden und Anwendungen der Computerlinguistik

1.4.1 Methodologie des Parsing

- *Zerlegung* eines komplexen Zeichens in seine elementaren Bestandteile,
- *Klassifikation* der gefundenen Teile mit Hilfe des Lexikons, und
- *Zusammensetzen* der klassifizierten Teile zur Ableitung einer grammatischen Gesamtanalyse des komplexen Zeichens.

1.4.2 Praktische Aufgaben der Computerlinguistik

- Indizieren von und Abruf aus textuellen Datenbanken
- Maschinelle Übersetzung
- Automatische Textproduktion
- Automatische Textüberprüfung
- Automatische Inhaltsanalyse
- Automatisierter Unterricht
- Dialogsysteme und automatische Auskunft

1.5 Modalitäten und Medium in Spracherkennung und -synthese

1.5.1 Modalitäten der Zeichenerkennung

Zeichenerkennung basiert auf folgenden Eingabegeräten

- Ohren – gesprochene Sprache
- Augen – handgeschriebene, gedruckte, und Gehörlosensprache
- Tastsinn – Blindenschrift (Braille)

1.5.2 Modalitäten der Zeichensynthese

Zeichensynthese basiert auf folgenden Ausgabegeräten

- Stimmbänder in Kombination mit dem Mund – gesprochene Sprache
- Hand – geschriebene Sprache, einschließlich Braille
- Hand-Arm-Gesichtsgesten – Gehörlosensprache

1.5.3 Medien der Sprache

Nichtelektronische Medien:

- *Laute* der gesprochenen Sprache
- *Buchstaben* der handgeschriebenen oder gedruckten Sprache
- *Gebärden* einer Gehörlosensprache

Elektronisches Medium:

- Modalitätsabhängige Repräsentationen:
 - Tonbandaufnahmen gesprochener Sprache
 - Bitmap von geschriebener Sprache
 - Videoaufnahme einer Gebärdensprache
- Modalitätsunabhängige Repräsentation:
 - digital kodierte elektronische Zeichensequenzen, z. B. ASCII

1.5.4 Korrelation von Modalität und Medium

Monomedial

Handschrift, Druck, Gebärdensprache
gesprochene Sprache, Tonband
Braille

Eingabemodalität

Augen
Ohren
Tastsinn

Multimedial

Film

Augen, Ohren

Monomedial

Handschrift, Druck, Braille
gesprochene Sprache, Tonband

Ausgabemodalität

Hand
Stimmbänder mit Mund

Multimedial

Gebärdensprache

Hand-, Arm-, Gesichtsgebärden

1.5.5 Transfer zwischen modalitätsabhängigen und modalitätsunabhängigen Repräsentationen

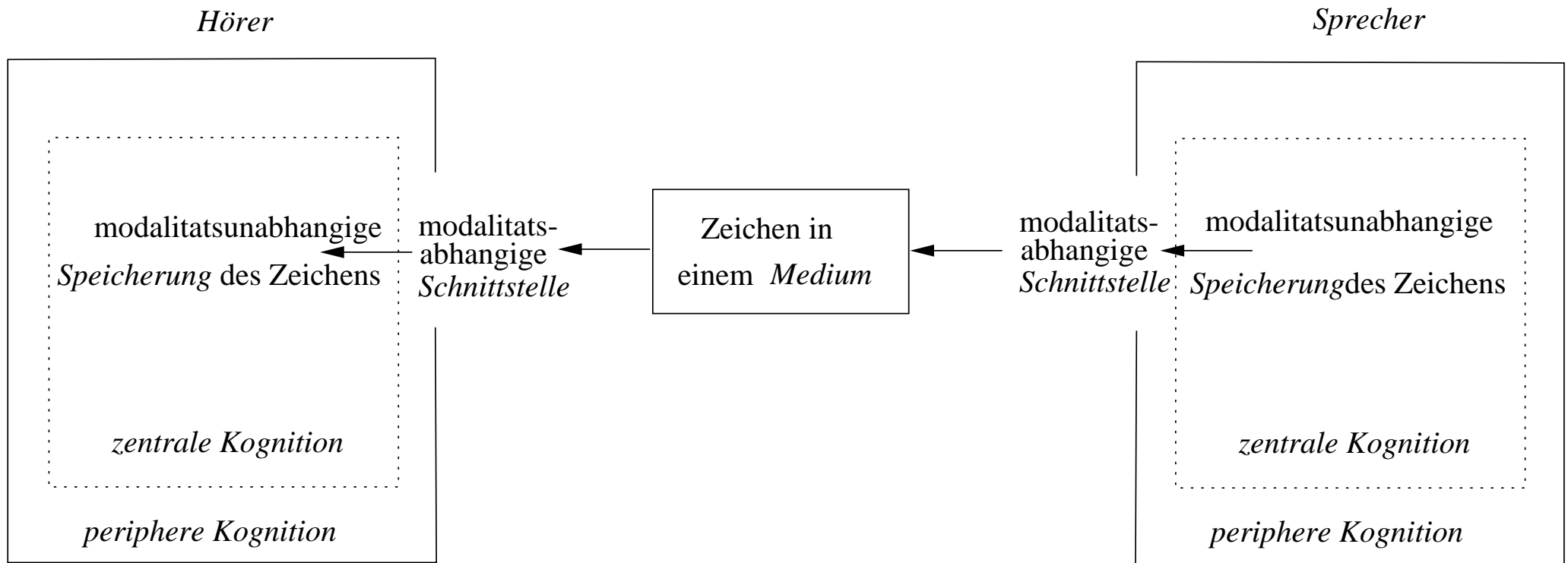
Spracherkennung: $ma \Rightarrow mu$ Transfer

modalitäts-abhängige Repräsentationen werden in modalitäts-unabhängige Repräsentationen abgebildet.

Sprachsynthese: $mu \Rightarrow ma$ transfer

modalitäts-unabhängige Repräsentationen werden in modalitäts-abhängige Repräsentationen abgebildet.

1.5.6 Speicherung, Modalität und Medium in Spracherkennung und Sprachsynthese



Vergleich: eine sprachgesteuerte Schreibmaschine

1.5.7 Methoden des $ma \Rightarrow mu$ Transfer

Nicht-automatischer $ma \Rightarrow mu$ Transfer: überläßt die Spracherkennung dem Menschen, der gesprochene oder geschriebene Sprache in den Computer eintippt.

Automatischer $ma \Rightarrow mu$ Transfer: akustische oder optische Mustererkennung.

1.5.8 Desiderata der automatischen Spracherkennung

- *Sprecherunabhängigkeit*
Das System soll verschiedene Sprecher mit verschiedenen Tonhöhen, Dialekten etc. spontan verstehen, – ohne daß eine anfängliche Lernphase erforderlich ist, in der das System an einen bestimmten Sprecher angepaßt werden muß.
- *Kontinuierlichkeit*
Das System soll kontinuierliche Sprache in unterschiedlicher Geschwindigkeit bewältigen, – ohne daß dabei unnatürliche Pausen zwischen den einzelnen Wörtern erforderlich sind.
- *Domänenunabhängigkeit*
Das System soll in der Lage sein, gesprochene Sprachzeichen unabhängig vom Inhalt zu erkennen – ohne daß ihm vorher eingegeben werden muß, welches Vokabular es zu erwarten bzw. nicht zu erwarten hat.
- *Realistischer Wortschatz*
Das System soll in der Lage sein, mindestens ebenso viele Wortformen zu erkennen wie ein durchschnittlicher Sprecher.
- *Robustheit*
Auch bei Abbrüchen, Kontraktionen und Verschleifungen soll das System in der Lage sein, die intendierten Wortformen zu erschließen.

1.5.9 Wie gut ist die automatische Spracherkennung heute?

It will be many years before unlimited vocabulary, speaker-independent continuous dictation capability is realized.

[Es wird noch viele Jahre dauern bis eine sprecherunabhängige, kontinuierliche Diktatfähigkeit realisiert sein wird.]

Zue, Cole & Ward 1998

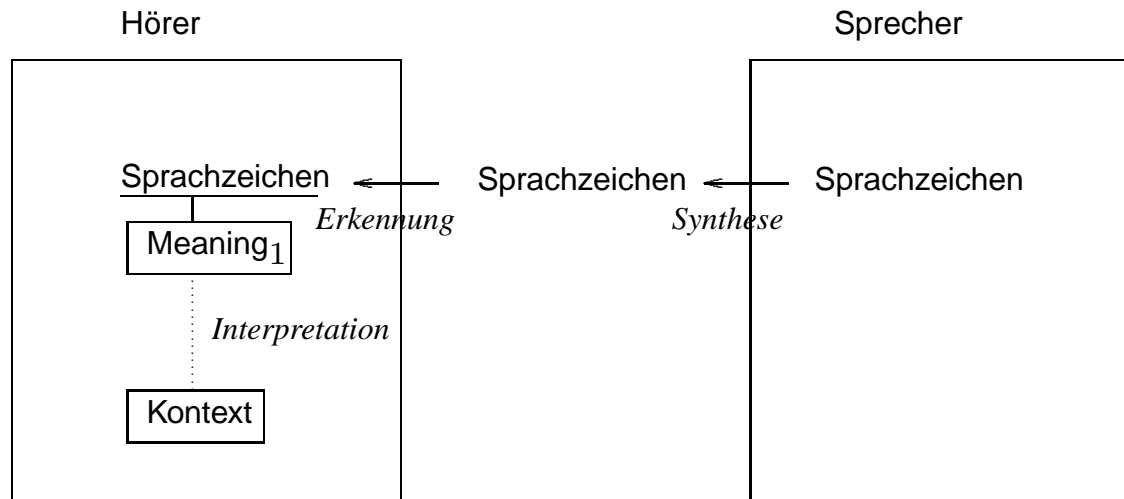
1.5.10 Die entscheidende Frage für die Konstruktion einer wirklich leistungsfähigen Spracherkennung:

Wie sollte das Grammatik- und Domänenwissen am besten organisiert werden?

Die Antwort ist offensichtlich:

Im Rahmen einer funktionalen, mathematisch effizienten und computergerechten Sprachtheorie.

1.5.11 Verkleinerung des Suchraums bei der automatischen Spracherkennung



- automatische Wortformerkennung: entspricht die Hypothese einer möglichen Wortform?
- automatische Syntaxanalyse: entspricht die Hypothese einem wohlgeformten Ausdruck?
- automatische semantisch-pragmatische Analyse: macht die Hypothese einen Sinn in bezug auf den aktuellen Verwendungskontext?

1.6 Zweite Gutenbergsche Revolution

1.6.1 Die Erste Gutenberg Revolution

Basierend auf der technischen Innovation des Druckens mit beweglichen Lettern, stellte sie der breiteren Öffentlichkeit eine Fülle von Wissen zur Verfügung.

1.6.2 Die Zweite Gutenberg Revolution

Basierend auf der automatischen Sprachverarbeitung im elektronischen Medium hat sie zum Ziel, dem Benutzer die gewünschte Information präzise, schnell und komfortabel zu *finden*.

1.6.3 SGML: *standard generalized markup language*.

A family of ISO standards for labeling electronic versions of text, enabling both sender and receiver of the text to identify its structure (e.g. title, author, header, paragraph, etc.)

Dictionary of Computing, p. 416 (ed. Illingworth et al. 1990)

1.6.4 Zeitungstext mit SGML-Steuerzeichen

```
<HTML>
<HEAD>
<TITLE>9/4/95 COVER: Siberia, the Tortured Land</TITLE>
</HEAD>
<BODY>
<!-- #include "header.html" -->
<P>TIME Magazine</P>
<P>September 4, 1995 Volume 146, No. 10</P>
<HR>
Return to <A href="../../../time/magazine/domestic/toc/
950904.toc.html">Contents page</A>
<HR>
<BR>
<!-- end include -->
<H3>COVER STORY</H3>
<H2>THE TORTURED LAND</H2>
<H3>An epic landscape steeped in tragedy, Siberia suffered
grievously under communism. Now the world's capitalists covet
its vast riches </H3>
<P><EM>BY <A href="../../../time/bios/eugenelinden.html">
EUGENE LINDEN</A>/YAKUTSK</EM>
<P>Siberia has come to mean a land of exile, and the place
easily fulfills its reputation as a metaphor for death and
```

1.6.5 Verschiedene Textsorten

- Zeitungsartikel
- Buch
- Theaterstück
- Drehbuch
- Lexikon

1.6.6 TEI

Text encoding initiative: defines a DTD (*document type definition*) for the markup of different types of text in SGML.

1.6.7 Unterschiedliche Ziele der Textauszeichnung

- Funktionsorientiert: SGML und TEI
- Druckbildorientiert: T_EX und L^AT_EX
- Benutzerorientiert: Winword, WordPerfect, etc.

1.6.8 Alphabetische Liste von Wortformen

10	in	STORY
146	in	suffered
1995	in	sun
20	its	than
4	its	that
a	LAND	The
a	land	the
a	landscape	the
a	like	the
a	LINDEN	the
above	Magazine	the
across	markers	the
and	mean	the
and	metaphor	the
and	midnight	the
and	midsummer	the
Arctic	million	through
as	mist	Throughout
as	more	to
barracks	mossy	to
bits	muting	TORTURED

2. Technologie und Grammatik

2.1 Indizieren und Finden in textuellen Datenbanken

2.1.1 Indizieren

Die Indizierung einer textuellen Datenbank ist eine Tabelle, die für jeden Buchstaben sämtliche Positionen (Adressen) im elektronischen Speichermedium der Datenbank auflistet, in denen dieser Buchstabe steht.

2.1.2 Vorteile einer elektronischen Indizierung

- Feinheit der Suche
- Flexibilität
 - Allgemeine Spezifikation von Mustern
 - Kombination von Mustern
- Automatischer Aufbau der Indexstruktur
- Bequemlichkeit, Geschwindigkeit, Zuverlässigkeit
 - Eingabe
 - Ausgabe

2.1.3 Definition von Recall und Precision

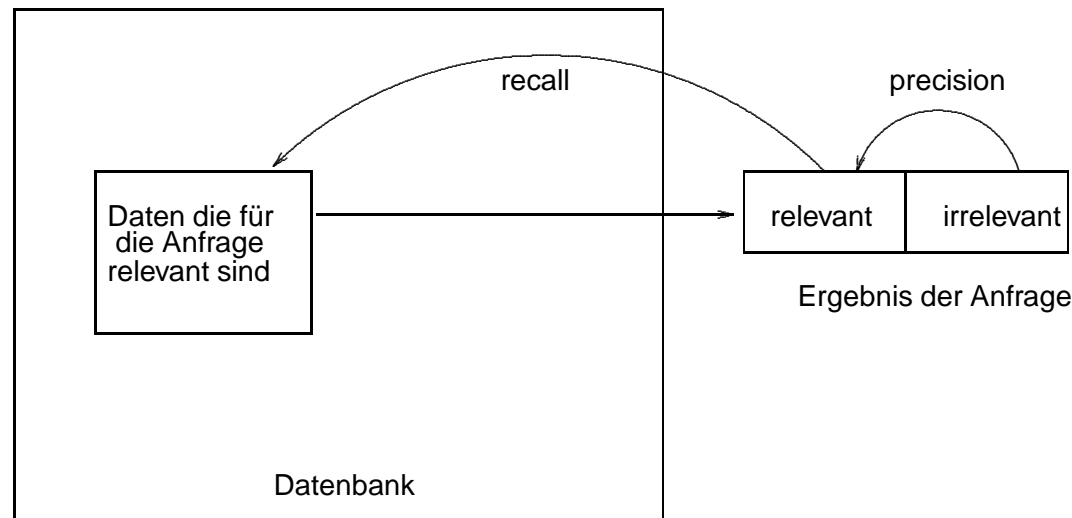
Recall mißt, wieviele der relevanten Texte im Verhältnis zum Gesamtbestand gefunden wurden.

Beispiel: Eine Datenbank aus mehreren Millionen Texten enthält 100 Texte, die für eine gegebene Fragestellung relevant sind. Wenn die Anfrage 75 Texte liefert, von denen 50 relevant und 25 irrelevant sind, dann ergibt sich ein *Recall* von $50 : 100 = 50\%$.

Precision mißt, wieviele relevante Texte im Suchergebnis enthalten sind.

Beispiel: Eine Anfrage hat 75 Texte ergeben, von denen 50 für den Benutzer relevant und 25 irrelevant sind. Dann ergibt sich eine *Precision* von $50 : 75 = 66,7\%$.

2.1.4 Korrelationen zur Bemessung von Recall und Precision



2.2 Einsatz grammatischen Wissens

2.2.1 Linguistische Methoden der Optimierung

A. Verarbeitung der Anfrage

- Automatische *query expansion*
 - (i) Die Suchwörter der Anfrage werden in ihre möglichen Flexionsformen ‘explodiert’. Dann wird die Datenbank mit allen Formen der jeweiligen Paradigmen durchsucht.
 - (ii) Über einen Thesaurus werden den Suchwörtern sämtliche Synonyme, Hypernyme und Hyponyme zugeordnet, die dann bei der Suche berücksichtigt werden – möglicherweise ebenfalls in explodierter Form.
 - (iii) Die syntaktischen Strukturen der Anfrage, z. B. **betrunkenen Fahrer**, werden automatisch in äquivalente Versionen *transformiert*. Hier z. B. in: **Fahrer, der betrunken war**.
- Interaktive Anfrageoptimierung

Vor Beginn der eigentlichen Suchprozedur wird dem Benutzer die bearbeitete Anfrage vorgelegt, um unsinnige Ergebnisse der automatischen Expansion zu eliminieren und die Fragestellung neu einzugrenzen.

B. Differenzierung der Indexierung

- **Buchstabenbasierte Indexierung:**
Diese technologische Standardmethode erlaubt es, sämtliche Vorkommen einer Buchstabenfolge in der Datenbank zu finden.
- **Morphologiebasierte Indexierung:**
Beim Einlesen wird jede Wortform im Text analysiert und ihrer Grundform zugeordnet. Mit dieser Information wird ein Index aufgebaut, der es erlaubt, zu einer bestimmten Grundform sämtliche entsprechenden Wortformen in der Datenbank zu finden.
- **Syntaxbasierte Indexierung:**
Beim Einlesen wird der Text syntaktisch geparkt (d.h. automatisch bezüglich seiner grammatikalischen Struktur analysiert), wobei viele morphologische Ambiguitäten eliminiert werden. Aus den grammatischen Analysen wird ein Index aufgebaut, mit dem man sämtliche Vorkommen einer syntaktischen Konstruktion sowie syntaktische Variationen einer Konstruktion finden kann.
- **Konzeptbasierte Indexierung:**
Beim Einlesen wird der Text semantisch und pragmatisch analysiert. Dabei werden nicht nur semantische Ambiguitäten eliminiert und domänenabhängige Spezialverwendungen erschlossen, sondern es wird zudem über den Inhalt des Textes ein Index aufgebaut, mit dem man sämtliche Vorkommen eines bestimmten Konzepts in der Datenbank finden kann.

C. Nachbearbeitung der gefundenen Daten

- Der niedrige Precision-Wert bei einer allgemein gehaltenen Fragestellung kann durch eine automatische Nachbearbeitung der gefundenen Daten ausgeglichen werden. Da die Rohergebnisse der Suche im Vergleich zur Gesamtmenge der Texte verhältnismäßig klein sind, können sie nach dem Anfragevorgang geparkt und auf ihren Inhalt untersucht werden. Anschließend werden nur die Texte an den Benutzer ausgegeben, die sich in der Nachbearbeitung als relevant erwiesen haben.

2.3 *Smart oder Solid Solutions?*

2.3.1 Smart solutions

vermeiden schwierige, kostspielige oder theoretisch ungelöste Aspekte der anstehenden Aufgabe, wie z. B.

- Weizenbaum's Eliza Programm, das natürliche Sprache zu verstehen scheint, es aber nicht tut.
- Statistisch-basiertes Tagging, das Wortarten erraten kann.
- Direkter und Transfer-Ansatz bei der maschinellen Übersetzung, welche ein Verstehen des Quelltextes vermeiden.

2.3.2 Solid solutions

zielen auf ein vollständiges theoretisches und praktisches Verstehen der Phänomene. Bei Anwendungen wird auf fertige *off-the-shelf* Komponenten zurückgegriffen, wie z. B.

- eine automatische Wortformanalyse, die auf einem online Lexikon und einem regelbasierten Morphologieparser basiert, der alle Aspekte der Flexion, Derivation und Komposition der Sprache behandelt
- ein syntaktischer Parser für freie Texte, der die automatische Wortformanalyse verwendet
- eine semantische Interpretation der syntaktischen Analyse, die die wörtlichen Bedeutungen der Ausdrücke ableitet
- eine pragmatische Interpretation in bezug auf einen Verwendungskontext, die die Sprecherbedeutung der Äußerung ableitet.

2.3.3 Vergleich

- Solange fertige *off-the-shelf* Komponenten nicht zur Verfügung stehen, scheinen *smart solutions* zunächst billiger und schneller zu sein. *Smart solutions* sind jedoch teuer in der Wartung und ihre Leistung kann nicht wesentlich verbessert werden.
- Die Entwicklung von Grammatikkomponenten im Rahmen einer *solid solution* sind eine langfristige Investition. Die Komponenten können in einer Vielzahl von Anwendungen wiederverwendet werden, wobei Verbesserungen der Grammatik direkt zu Verbesserungen bei den Anwendungen führen.

2.3.4 Wahl zwischen *smart* oder *solid solution* hängt auch von der Anwendung ab

- Eine *smart solution*, die bei einer großen Datenbank einen 70% Recall bringt, liefert wesentlich mehr als das, was die Benutzer mit Hilfe menschlicher Arbeitskraft erreichen könnten. Außerdem wissen die Benutzer nicht, welche Texte das System übersehen hat.
- Dagegen sind die Mängel einer 70% akkuraten *smart solution* in der maschinellen Übersetzung für die Benutzer von schmerzlicher Offensichtlichkeit. Zudem gibt es die Alternative menschlicher Übersetzer, die ein wesentlich besseres Ergebnis erzielen können.

2.4 Anfänge der maschinellen Übersetzung

2.4.1 Sprachpaare

Französisch → *Englisch* und *Französisch* ← *Englisch* sind zwei verschiedene Sprachpaare.

2.4.2 Formel zur Berechnung der Zahl von Sprachpaaren

$n \cdot (n - 1)$, wobei $n = \text{Zahl der verschiedenen Einzelsprachen}$

Zum Beispiel muß eine EU mit derzeit $11 \cdot 10 = 110$ Sprachpaaren fertig werden.

2.4.3 Sprachpaare bei der Übersetzung eines französischen Dokuments in der EU

Französisch → Englisch

Französisch → Deutsch

Französisch → Italienisch

Französisch → Holländisch

Französisch → Finnisch

Französisch → Spanisch

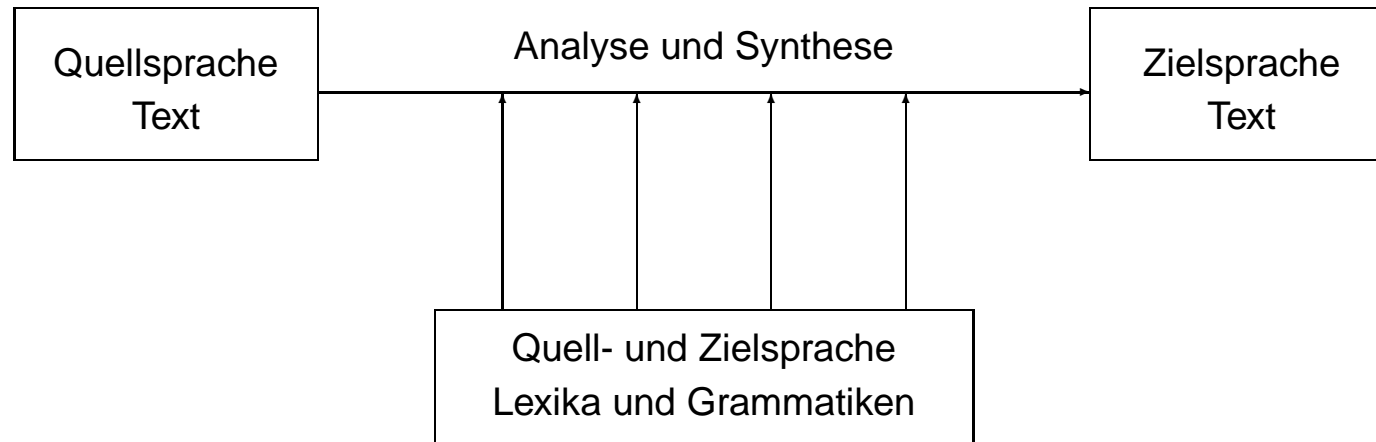
Französisch → Portugiesisch

Französisch → Griechisch

Französisch → Dänisch

Französisch → Schwedisch

2.4.4 Schema der direkten Übersetzung



2.4.5 Was ist FAHQT?

FULLY AUTOMATIC HIGH QUALITY TRANSLATION

2.4.6 Beispiele von automatischen Fehlübersetzungen

Out of sight, out of mind. \Rightarrow *Invisible idiot.*

The spirit is willing, but the flesh is weak. \Rightarrow *The whiskey is alright, but the meat is rotten.*

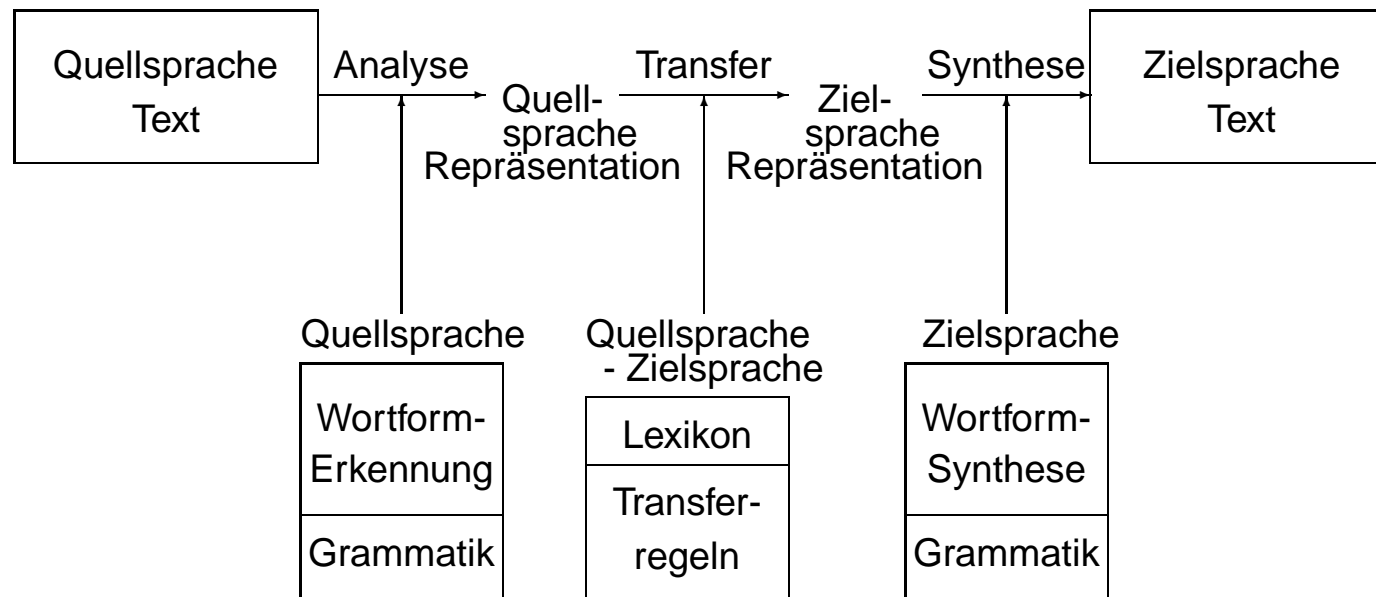
La Cour de Justice considère la création d'un sixième poste d'avocat général. \Rightarrow *The Court of Justice is considering the creation of a sixth avocado station.*

2.4.7 Transfer-Ansatz

versucht eine modulare Trennung

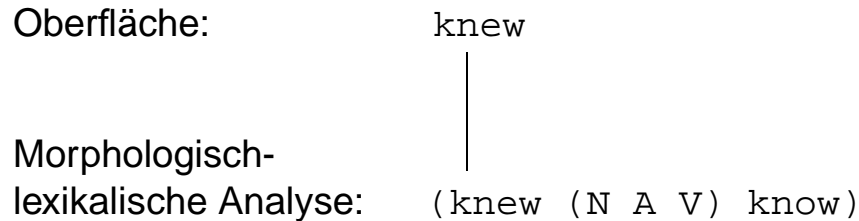
- von Quellsprach-Analyse und Zielsprach-Synthese,
- von linguistischen Daten und Verarbeitungsprozeduren und
- von Lexika für die Quellsprach-Analyse, den Quellsprach-Zielsprach-Transfer und die Zielsprach-Synthese.

2.4.8 Schema des Transfer-Ansatzes



2.4.9 Drei Phasen eines Wortform-Transfers *Englisch-Deutsch*

1. Quellsprach-Analyse der englischen Wortform knew:



Die Quellsprach-Analyse der unanalysierten Oberfläche **knew** ist ein geordnetes Tripel, bestehend aus der Oberfläche, der syntaktischen Kategorie und der Grundform **know**.

2. Quell-Zielsprach-Transfer:

Aus der Grundform der Quellsprach-Analyse werden über ein Lexikon die Grundformen korrespondierender Zielsprach-Wörter erzeugt:

know \implies wissen
kennen

3. Zielsprach-Synthese:

Über die Kategorie der Quellsprach-Form und die Grundformen der korrespondierenden Zielsprach-Wörter werden mittels Zielsprach-Morphologie und -Lexikon die gewünschten Zielsprach-Wortformen erzeugt:

wußte	wüßte	kannte	kennte
wußtest	wüßtest	kanntest	kenntest
wußten	wüßten	kannten	kennten
wußtet	wüßtet	kanntet	kenntet

2.4.10 Nachteile des direkten und des Transfer-Ansatzes

- Jedes Sprachpaar erfordert eine eigene Quell-Zielsprachkomponente.
- Analyse und Synthese beschränken sich auf einzelne Sätze.
- Die Ansätze vermeiden eine semantisch-pragmatische Analyse und versuchen ohne ein Verstehen der Quellsprach-Texte auszukommen.

2.5 Maschinelle Übersetzung heute

2.5.1 Warum Sprachverstehen für die Übersetzung unverzichtbar ist

- Syntaktische Ambiguität der Quellsprache
 1. Nicht kopierfähige Folien schmelzen und verkleben die Maschine.
Folien (schmelzen und verkleben die Maschine)
(Folien schmelzen) und (verkleben die Maschine)
 2. Susanne beobachtete die Yacht mit dem Fernglas.
Susanne beobachtete den Mann mit dem Bart.
 3. The mixture gives off dangerous cyanide and chlorine fumes.
(dangerous cyanide) and chlorine fumes
dangerous (cyanide and chlorine) fumes
- Lexikalische Quell-Zielsprach-Unterschiede
 1. Die Männer ermordeten die Frauen. Sie wurden drei Tage später gefaßt.
Die Männer ermordeten die Frauen. Sie wurden drei Tage später begraben.
 2. know – wissen – savoir
kennen – connaître
 3. The watch included two new recruits that night.

- Syntaktische Quell-Zielsprach-Unterschiede

- Deutsch:

- Auf dem Hof sahen wir einen kleinen Jungen, der einem Ferkel nachlief.
Dem Jungen folgte ein großer Hund.

- Englisch:

- In the yard we saw a small boy running after a piglet.
A large dog followed the boy.
The boy was followed by a large dog.

- Kollokation und Idiom

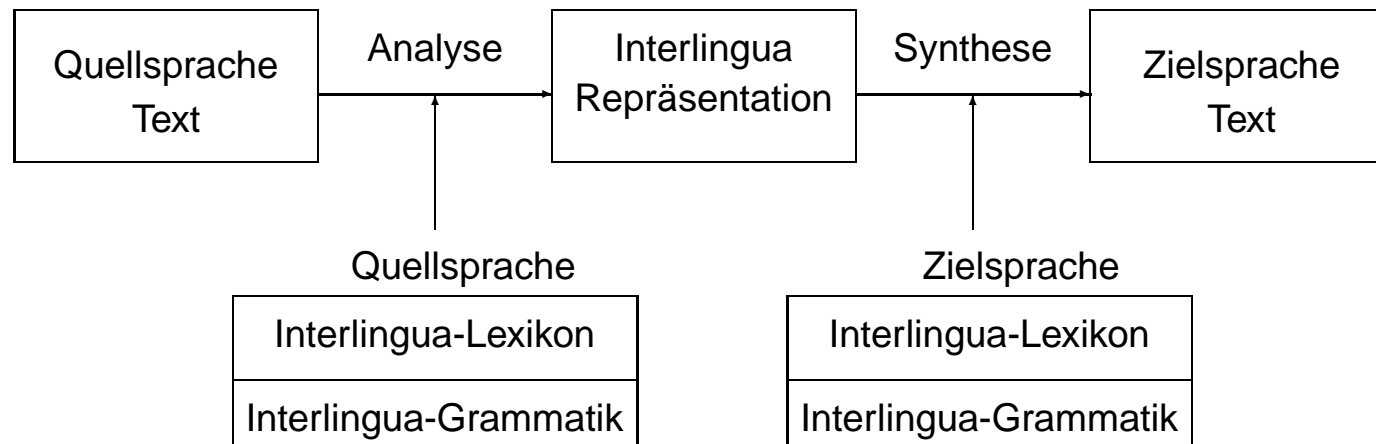
strong current | high voltage (but: *high current | *strong voltage)

bite the dust | ins Gras beißen (but: *bite the grass | *in den Staub beißen)

2.5.2 Praktische Teillösungen in der MÜ

1. **MAT** (*machine aided translation*), also menschliche Übersetzung, aber mit Hilfe von elektronischen Hilfsmitteln wie Online-Lexika, Textverarbeitung, morphologischer Analyse etc.
2. **Rohübersetzungen**, etwa im Rahmen des Transfer-Ansatzes, die stark von Übersetzern nachbearbeitet werden müssen.
3. **Kontrollierte Sprache**, d.h. zwar vollautomatische Übersetzung, aber nur von Texten, deren lexikalische und syntaktische Struktur kanonisch beschränkt ist.

2.5.3 Schema des Interlingua-Ansatzes



2.5.4 Vorteil des Interlingua-Ansatzes

Die Grundstruktur des Interlingua-Ansatzes hat zur Folge, daß für $n(n - 1)$ Sprachpaare nur $2n$ interlinguale Teilsysteme benötigt werden (nämlich n Analyse- und n Synthese-Module).

2.5.5 Kandidaten, die als Interlingua vorgeschlagen worden sind

- eine künstliche Logiksprache,
- eine Mischsprache wie Esperanto, die zwar konstruiert ist, aber wie eine natürliche Sprache funktioniert,
- eine Menge semantischer Urausdrücke, die allen natürlichen Sprachen gemeinsam ist, eine Art universales Vokabular.

3. Kognitive Grundlage der Semantik

3.1 Prototyp der Kommunikation

3.1.1 Varianten der sprachlichen Kommunikation

- Zwei Gesprächspartner stehen sich gegenüber (*face to face communication*) und sprechen über konkrete Gegenstände in ihrer unmittelbaren Umgebung.
- Zwei Gesprächspartner unterhalten sich am Telefon über gemeinsame Erlebnisse in der Vergangenheit.
- Ein Kaufmann schreibt an eine Firma und bestellt Waren in bestimmter Anzahl, Größe, Farbe etc. Die Firma reagiert, indem sie die Bestellung ausführt.
- Eine Tageszeitung informiert über den Ausbau der öffentlichen Verkehrsmittel.
- Ein Übersetzer überträgt eine deutsche Kurzgeschichte ins Englische.
- Eine Physiklehrerin erklärt die Fallgesetze.
- Ein Standesbeamter stellt einen Trauschein aus.
- Eine Richterin verkündet ein Urteil.
- Ein Schaffner sagt: Endstation, alles aussteigen!
- Auf einem Schild steht: Betreten des Rasens verboten!
- Ein Autor schreibt eine Science-Fiction-Geschichte.
- Eine Schauspielerin spricht eine Rolle.

3.1.2 Kommunikationsprototyp

Der Prototyp der natürlichen Kommunikation ist die direkte (*face to face*) Unterhaltung zweier Gesprächspartner über konkrete Gegenstände in ihrer direkten Umgebung.

Mögliche Alternativen: geschlossene Texte oder natürliche Indizien, wie die Anzeige von Feuer durch Rauch.

3.1.3 Drei Komponenten der Kommunikationsprototyps

- Spezifikation der Handlungsumgebung (*task environment*)
- Struktur des Sprechers-Hörers (*cognitive agent*)
- Spezifikation der Sprache

3.1.4 Objekte in der Welt von CURIOUS

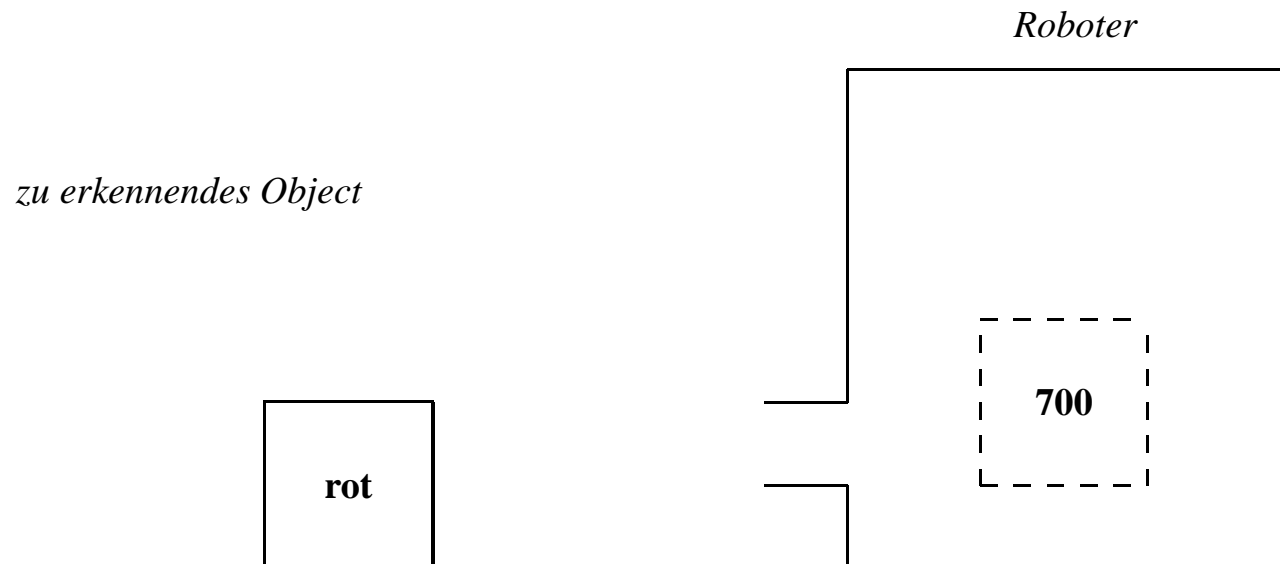
- Dreiecke (gleichseitig, rechtwinklig etc.)
- Vierecke (quadratisch, nicht quadratisch)
- Kreise und Ellipsen

3.2 Vom Wahrnehmen zum Erkennen

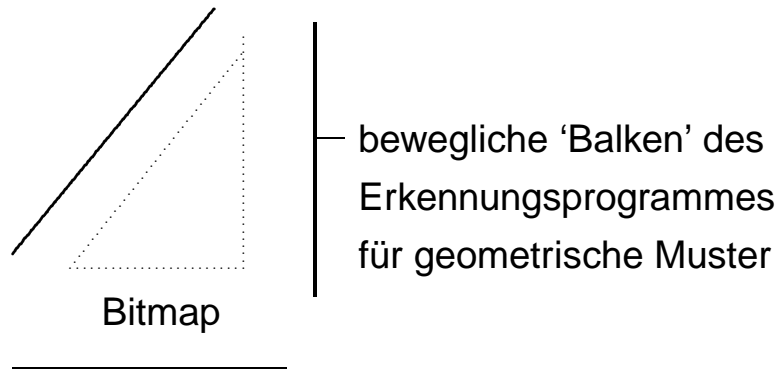
3.2.1 Kriterien zur Bewertung von CURIOUS

- Beobachtung des aktiven und reaktiven Verhaltens (Verhaltenstest)
- direkte Beurteilung der kognitiven Verarbeitung (Kognitionstest)

3.2.2 Interne Bitmap-Repräsentation der Umwelt



3.2.3 Analyse eines internen Abbilds



3.2.4 Definition des Kontexts

Der Kontext eines kognitiven Agenten KAG zu einem gegebenen Zeitpunkt t umfaßt

1. die Gesamtheit aller aktuellen kognitiven Parameterwerte KAG_t ,
2. die logische Analyse der Parameterwerte (rekonstruierte Muster) und ihre Kombination sowie
3. die Klassifikation der rekonstruierten Muster und ihrer Kombinationen mit Hilfe von Konzepten.

3.3 Ikonizität formaler Konzepte

3.3.1 Definition: M-Konzept

Ein M-Konzept ist die strukturelle Darstellung einer charakteristischen Parameterkonstellation, wobei bestimmte Parameterwerte als Variablen definiert sind.

3.3.2 Beispiel eines M-Konzepts: Quadrat (Type)

Kante 1: α cm
Winkel 1/2: 90^0
Kante 2: α cm
Winkel 2/3: 90^0
Kante 3: α cm
Winkel 3/4: 90^0
Kante 4: α cm
Winkel 4/1: 90^0

3.3.3 Definition: I-Konzept_{loc}

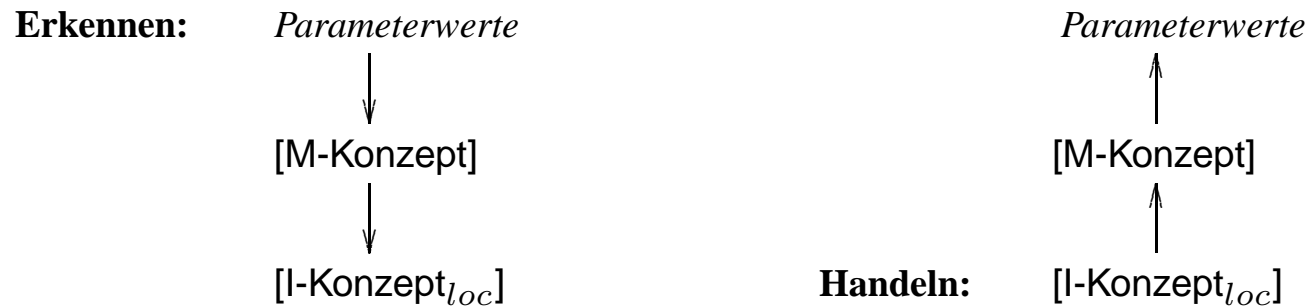
Ein I-Konzept_{loc} entsteht über die erfolgreiche Abgleichung eines M-Konzeptes auf eine entsprechende Konstellation von Parameterwerten an einem raumzeitlichen Ort.

3.3.4 Beispiel eines I-Konzept_{loc}: ein Quadrat (Token)

Kante 1: 2cm
Winkel 1/2: 90 ⁰
Kante 2: 2cm
Winkel 2/3: 90 ⁰
Kante 3: 2cm
Winkel 3/4: 90 ⁰
Kante 4: 2cm
Winkel 4/1: 90 ⁰

loc

3.3.5 Kontextuelles Erkennen und Handeln



3.3.6 Aspekte der Ikonizität

- Die Parameterwerte im internen Kontext sind insofern Abbilder, als sie korrespondierende Strukturen der realen Welt wiedergeben.
- Die rekonstruierten Muster (I-Konzept_{loc}) sind Abbilder der Parameterwerte, weil sie eine logische Analyse der Parameterwerte sind.
- Die M-Konzepte des internen Kontexts sind insofern Abbilder, als sie (i) als Abstraktion über ähnlichen Parameterkonstellationen entstanden und (ii) entsprechende Klassen von rekonstruierten Mustern charakterisieren.

3.3.7 Verfehlte Argument gegen Ikonizität

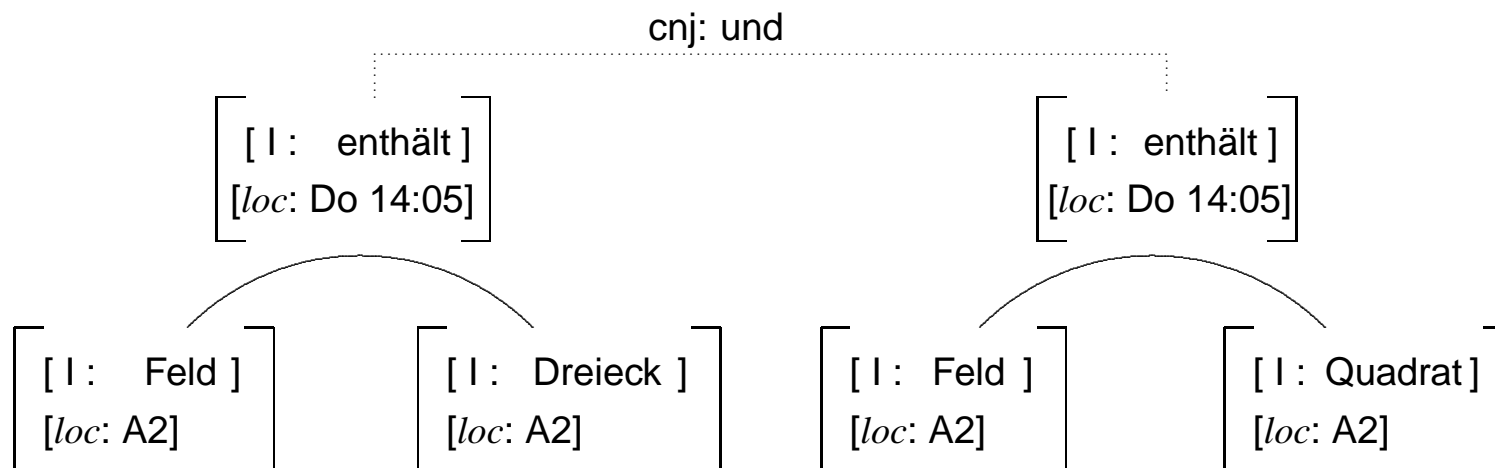
- Die chirurgische Untersuchung des Gehirns einer Person, die einen Baum gesehen hat, würde ein solches Bild nicht finden.
- Was für eine *Art* von z. B. Dreieck sollte das zugehörige Konzept *genau* sein: ein gleichseitiges, ein rechtwinkliges, ein unregelmäßiges? (BERKELEY 1685–1753)
- Wenn es im Kopf Bilder gibt, dann muß es jemanden geben, der sie sieht. Es nützt aber nichts, einen kleinen Mann (homunculus) im Kopf zu postulieren, denn der hätte dann ja auch Bilder im Kopf, für die er wiederum einen noch kleineren Homunculus benötigen würde und so weiter. Da also ein Homunculus bei der Interpretation von Bildern im Kopf offenbar keine Hilfe ist, sind die Bilder selbst als überflüssig zu betrachten (HUME 1711–1776).

3.4 Kontextuelle I-Propositionen

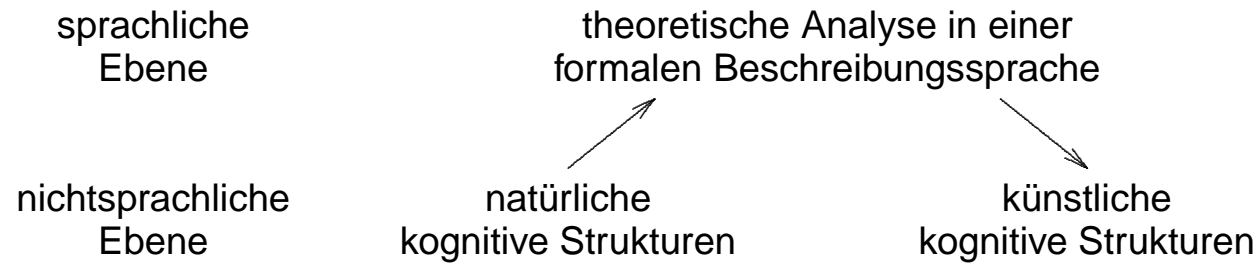
3.4.1 Die Bausteine elementarer Propositionen

<i>Logik</i>	<i>Welt</i>	<i>Sprache</i>
1. Argument	Objekt	Substantiv
2. Funktor	Relation	Verb
3. Modifikator	Eigenschaft	Adjektiv-Adverbial

3.4.2 Verkettung zweier I-Propositionen

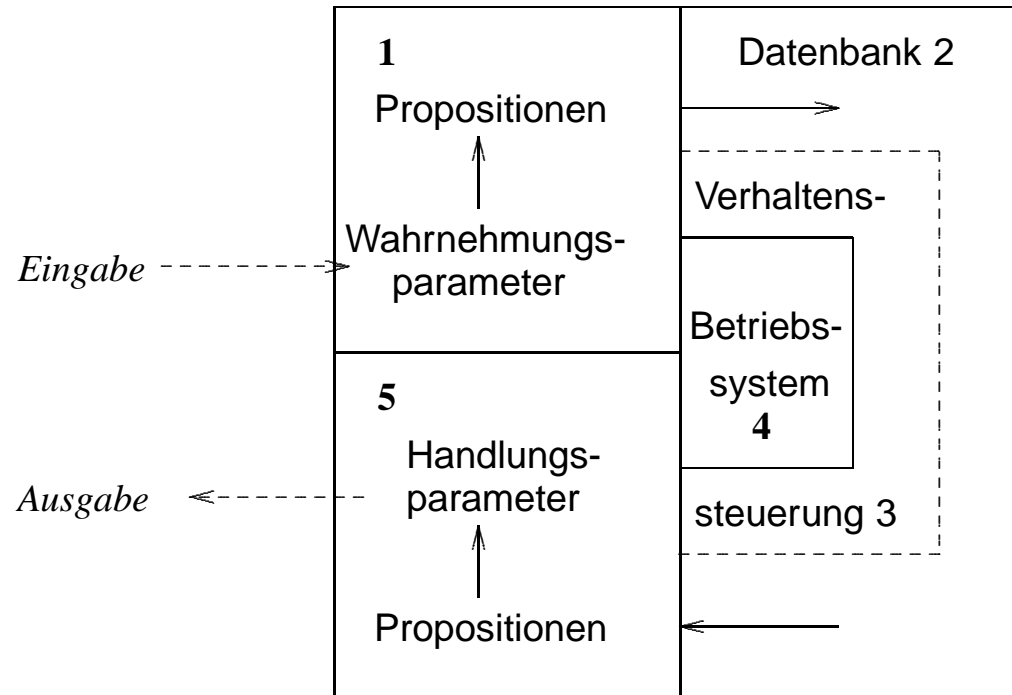


3.4.3 Künstliche Modellierung natürlicher Kognition



3.5 Erkennen und Handeln

3.5.1 Strukturschema der kontextuellen Kognition



3.5.2 Beispiel eines Programms zur Verhaltenssteuerung

1. Primäre Analyse der aktuellen Handlungsumgebung:
 - (a) Bewege Dich über das Startfeld A1.
 - (b) Analysiere aktuelles Feld:
 - i. führe Kanten an den Bitmap-Umriß heran,
 - ii. messe Farbwert innerhalb des Umrisses,
 - iii. bilde I-Proposition und
 - iv. schreibe I-Proposition in den Bereich 'aktuelle Primäranalyse' der Datenbank.
 - (c) Wenn aktuelles Feld nicht D4, bewege Dich ins nächste Feld und gehe in Zustand b. Sonst gehe nach 2.
2. Sekundäre Analyse der aktuellen Handlungsumgebung (Inferenzen):
 - (a) Zähle alle Dreiecke, Vierecke, Quadrate, roten Dreiecke etc. in der aktuellen Primäranalyse und schreibe das Ergebnis in den Bereich aktuelle Sekundäranalyse der Datenbank.
 - (b) Vergleiche die aktuelle Sekundäranalyse mit der vorherigen Sekundäranalyse und schreibe das Ergebnis (z. B. 'Zahl der roten Dreiecke um 2 höher') in den Bereich 'aktuelle Sekundäranalyse' der Datenbank.
3. Warte 10 Minuten.
4. Gehe in Zustand 1.

3.5.3 CURIOUS als autonomer Agent (nouvelle AI)

Without a carefully built physical grounding any symbolic representation will be mismatched to its sensors and actuators. These groundings provide the constraints on symbols necessary for them to be truly useful.

[Ohne eine sorgfältig gebaute physische Grundlage wird eine symbolische Repräsentation niemals mit ihren Sensoren und Aktivationsmodulen harmonieren. Diese physischen Grundlagen fungieren als Beschränkungen für die Symbole, ohne die diese nicht wirklich nützlich sein können.]

R.A. Brooks, 1990, S. 6.

3.5.4 CURIOUS als ein physisches Symbolsystem (classic AI)

The total concept [of a physical symbol system] is the join of computability, physical realizability (and by multiple technologies), universality, the symbolic representation of processes (i.e. interpretability), and finally, symbolic structure and designation.

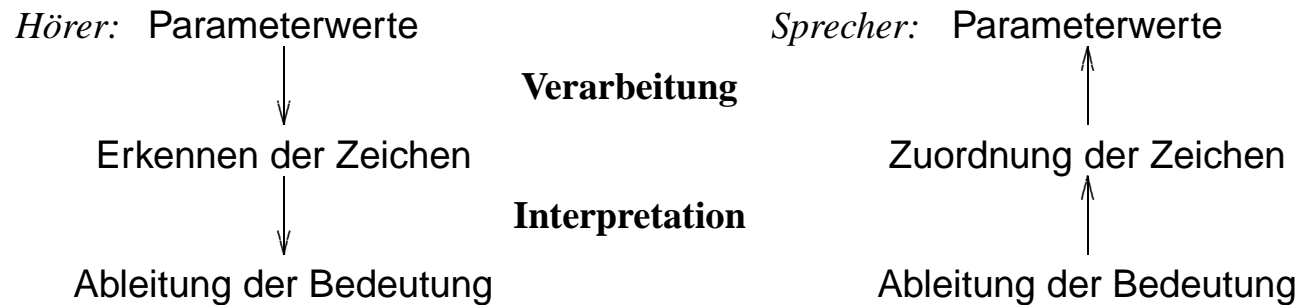
[Das Gesamtkonzept eines physischen Zeichensystems ist die Verbindung von Berechenbarkeit, physischer Realisierbarkeit (und zwar mit Hilfe unterschiedlicher Technologien), Allgemeingültigkeit, die zeichenbasierte Repräsentation von Prozessen (d.h. Interpretierbarkeit) und schließlich die Struktur des Zeichens und das, was sie bezeichnet.]

A. Newell & H. Simon 1975, p. 46

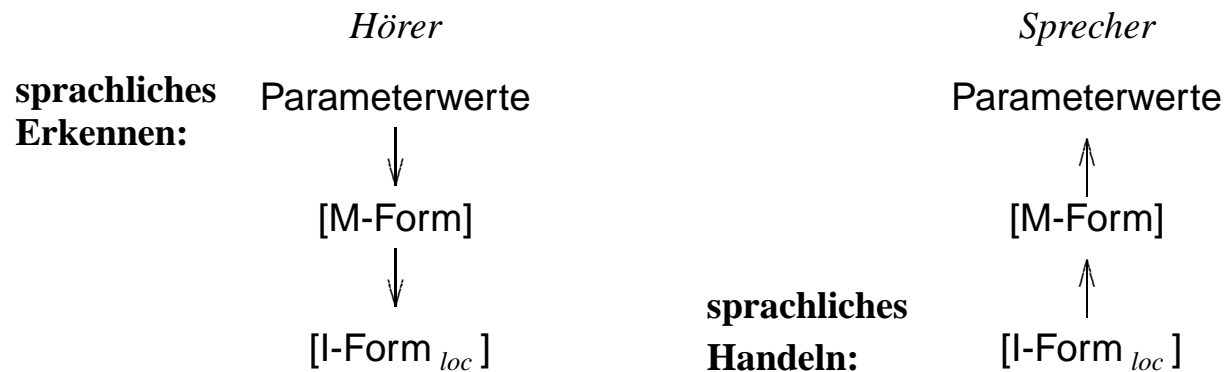
4. Sprachliche Kommunikation

4.1 Erweiterung auf Sprache

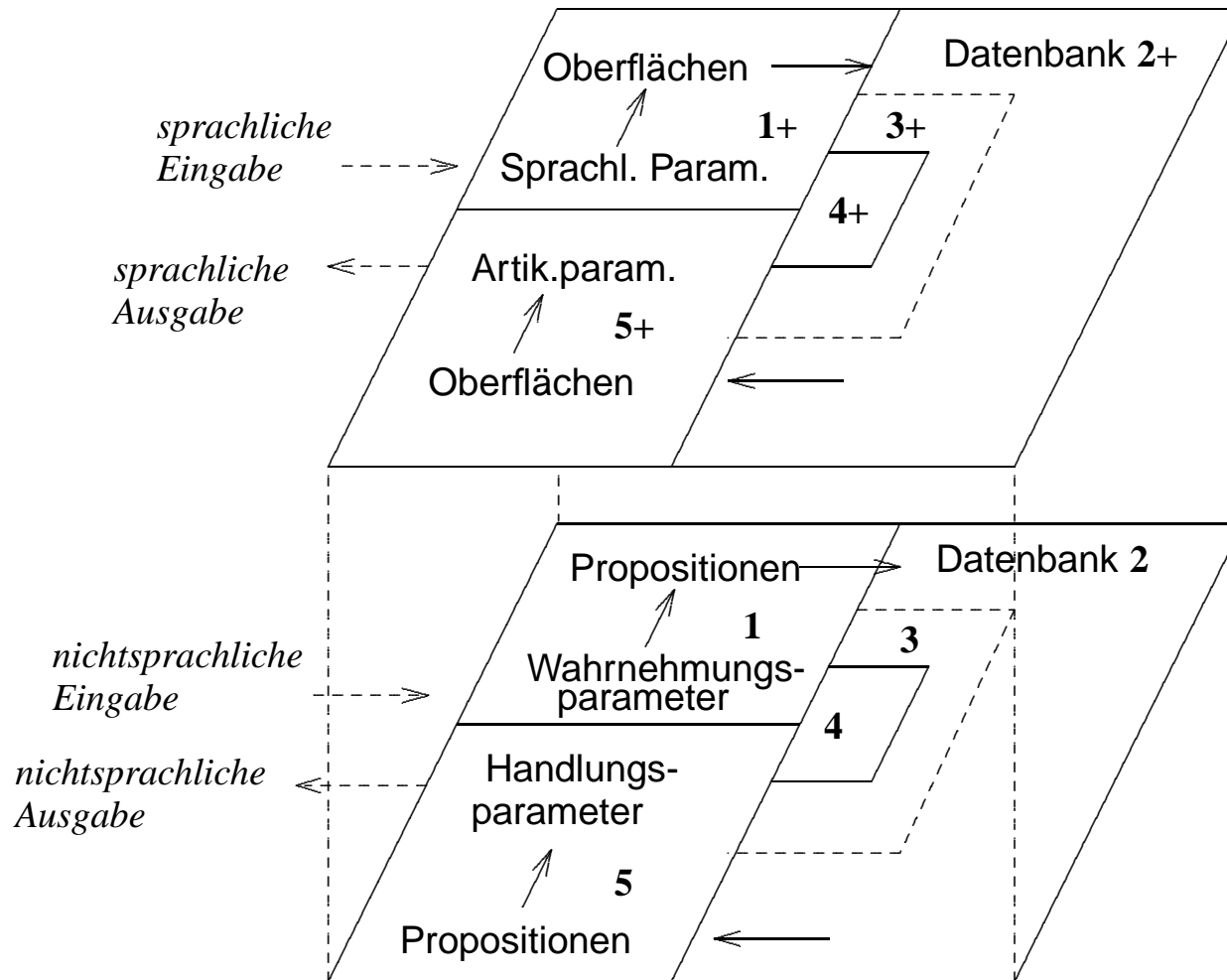
4.1.1 Zwei Teilprozesse der Sprachverwendung



4.1.2 Verarbeitung eines sprachlichen Zeichens

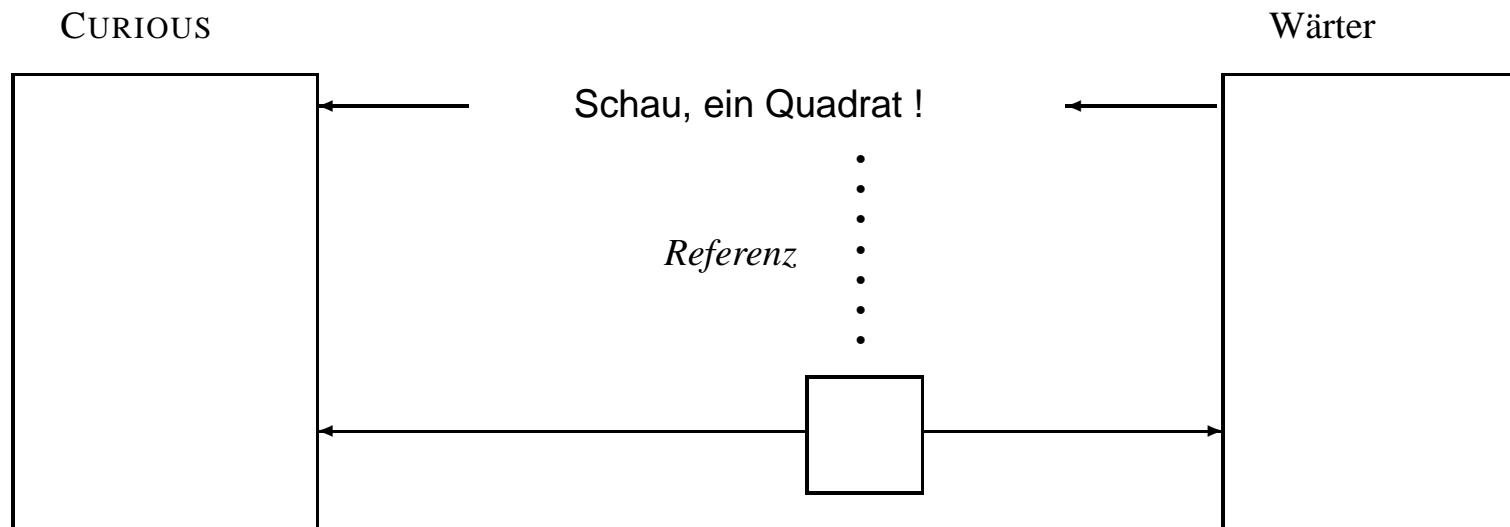


4.1.3 Erweitertes Strukturschema von CURIOUS



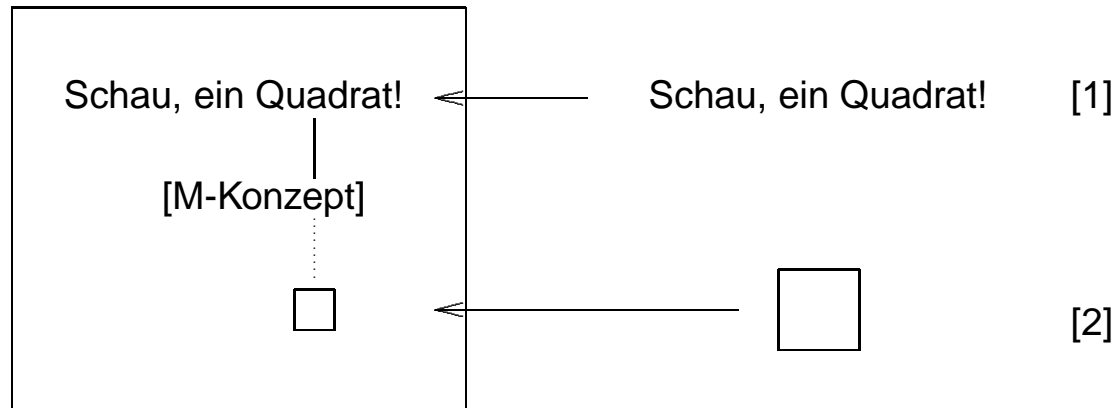
4.2 Modellierung der Referenz

4.2.1 Referenz aus externer Sicht

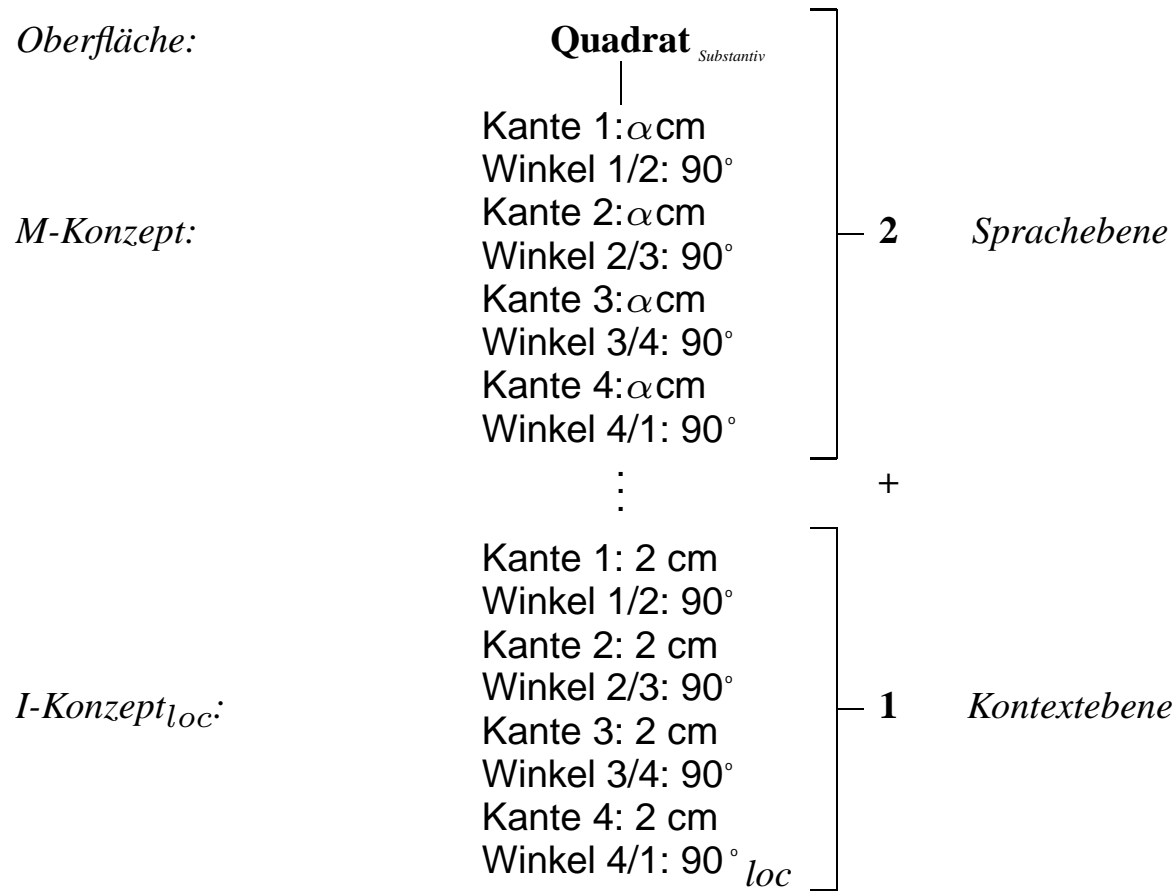


4.2.2 Interner und externer Referenzaspekt

CURIOUS



4.2.3 Kognitive 2+1 Ebenen-Analyse der Referenz



4.3 Using literal meaning

4.3.1 Unmittelbare und mittelbare Referenz

- *Unmittelbare Referenz* ist der Bezug des Sprechers oder Hörers auf Objekte in deren aktueller Handlungsumgebung.
- *Mittelbare Referenz* ist der Bezug des Sprechers oder Hörers auf Objekte, die sich nicht in deren aktueller Handlungsumgebung befinden.

4.3.2 Zwei Bedeutungsbegriffe

- Meaning_1 = Eigenschaft des Ausdrucks, auch wörtliche Bedeutung genannt.
- Meaning_2 = Eigenschaft der Äußerung, auch Sprecherbedeutung genannt.

4.3.3 Erster Hauptsatz der Pragmatik (PoP-1)

Die Sprecherbedeutung der Äußerung (Meaning_2) besteht in der Verwendung der Sprachbedeutung (Meaning_1) des Zeichens relativ zum Kontext.

4.4 Freges Prinzip

4.4.1 Das Fregesche Prinzip

Die Bedeutung eines komplexen Ausdrucks ist eine Funktion der Bedeutung der Teile und der Art ihrer Zusammensetzung.

4.4.2 Verschiedene Teile

- a.* Der Hund beißt den Mann.
- b.* Der Hund beißt den Knochen.

4.4.3 Verschiedene Zusammensetzungen

- a.* Der Hund beißt den Mann.
- a'*. Der Mann beißt den Hund.

4.4.4 Standard-Interpretation des Fregeschen Prinzips

Oberfläche:	a	=	a		a	≠	b
	⋮		⋮		⋮		⋮
Meaning ₁ :	A	=	A		A	≠	B

4.4.5 Syntaktische Ambiguität

Das Kind füttert die Mutter.

4.4.6 Paraphrase

Der Hund beißt den Mann. (Aktiv)

Der Mann wird von dem Hund gebissen. (Passiv)

4.4.7 Scheinbare Ausnahmen (Inkorrekte Grob-Analysen)

	Ambiguität	Paraphrase
Oberfläche:	$a = a$	$a \neq b$
	⋮ ⋮	⋮ ⋮
Meaning ₁ :	$A \neq A'$	$A = B$

4.4.8 Syntaktische Ambiguität (korrekte Analyse)

unanalysierte Oberfläche:	$a = a$	
analysierte Oberfläche:	$a \neq a'$	Anwendungsbereich des Fregeschen Prinzipes
	⋮ ⋮	
Meaning ₁ :	$A \neq A'$	

4.4.9 Syntactische Paraphrase

	<i>Falsch</i>	<i>Korrekt</i>
Oberfläche:	$2 + 4 \neq 3 + 3$	$2 + 4 \neq 3 + 3$
	⋮ ⋮	⋮ ⋮
Meaning ₁ :	$6 = 6$	$2' + 4' \sim 3' + 3'$
	Identität	Äquivalenz

4.5 Oberflächenkompositionalität

Die Standard-Interpretation 4.4.4 des Fregeschen Prinzips entspricht dem Prinzip Oberflächenkompositionalität.

4.5.1 Oberflächenkompositionalität I (OK-Prinzip I)

Eine Analyse natürlicher Sprache ist oberflächenkompositional, wenn sie nur die konkreten, grammatikalisch analysierten Wortformen als Bausteine der syntaktischen Analyse verwendet, so daß alle syntaktischen und semantischen Eigenschaften eines komplexen Ausdrucks kompositional aus der Kategorie und der Meaning_1 der Bausteine abgeleitet werden.

4.5.2 Konsequenzen der Oberflächenkompositionalität

- Methodologisch:
Die syntaktischen Analysen sind maximal *konkret* weil keinerlei Nullelemente oder ‘*underlying forms*’ verwendet werden dürfen,
- Mathematisch:
Die syntaktischen und semantischen Analysen können von *niedriger Komplexität* sein.
- Funktional:
Die interne Abgleichung zwischen Meaning_1 und Kontext kann von einzelnen Wörtern auf die systematische syntaktisch-semantische *Kombination* der Ausdrücke ausgedehnt werden.

Verletzung der Oberflächenkompositionalität: BEISPIEL I

4.5.3 Linguistische Generalisierungen in der Transformationsgrammatik

Transformationen werden für angeboren gehalten, haben aber keine Funktion in der Kommunikation.

4.5.4 Beispiele ‘klassischer’ Transformationen

TIEFENSTRUKTUR:

OBERFLÄCHENSTRUKTUR:

Passiv-Transformation:

Julia las ein Buch.

⇒ Ein Buch wurde von Julia gelesen.

Reflexivierungs-Transformation

Peter_i rasiert Peter_i.

⇒ Peter rasiert sich.

Pronominalisierungs-Transformation

Julia_i sagte, daß Julia_i schlief.

⇒ Julia sagte, daß sie schlief.

Relativsatz-Transformation

Der Mann [der Mann las ein Buch]

⇒ Der Mann, der ein Buch las,

Hauptsatz-Wortstellungs-Transformation

Julia ein Buch gelesen hat ⇒ Julia hat ein Buch gelesen.

Object-raising-Transformation

Peter befahl Maria [Maria schläft] ⇒ Peter befahl Maria zu schlafen.

Subject-raising-Transformation

Peter versprach Maria [Peter schläft] ⇒ Peter versprach Maria zu schlafen.

4.5.5 Transformationen und die standard Interpretation des Fregeschen Prinzips

Eine Zeit lang wurden in der Transformationsgrammatik Aktiv und Passiv für äquivalent gehalten. Aufgrund des folgenden Gegenbeispiels musste diese Annahme wieder aufgegeben werden.

Aktiv Jeder in diesem Raum spricht mindestens zwei Sprachen.

Passiv Mindestens zwei Sprachen werden von jedem in diesem Raum gesprochen.

4.5.6 Transformationen und Darwins Gesetz: Form folgt Funktion

Die Struktur z. B. eines Entenfußes ist angeboren. Gute Wissenschaft sollte diese Struktur über ihre Funktion erklären.

Das Gleiche gilt für angeborene kognitive Strukturen, z. B. die Sprachfähigkeit.

4.5.7 Kognitive Variante von Ockhams Rasiermesser

Entitäten oder Komponenten der Grammatik sollen nicht als angeboren postuliert werden, wenn sie im Rahmen der natürlichsprachlichen Kommunikationsmechanik keine klare Funktion haben.

4.5.8 Anwendungen des kognitive Rasiermessers

Das kognitive Rasiermesser trifft auf die Transformationsgrammatik ebenso zu wie auf alle späteren Varianten des Nativismus, einschließlich LFG, GPSG und HPSG. Wie die Transformationsgrammatik sind ihre linguistischen Generalisierungen dysfunktional in bezug auf die Kommunikation und inhärent im Widerspruch zum Prinzip der Oberflächenkompositionalität

Verletzung der Oberflächenkompositionalität: BEISPIEL II

4.5.9 Bedeutungsdefinition von Grice

Definiendum: U meant something by uttering x.

Definiens: For some audience A, U intends his utterance of x to produce in A some effect (response) E, by means of A's recognition of the intention.

[Definiendum: U meinte etwas mit dem Äußern von x.

Definiens: Für eine Zuhörerschaft A intendiert U, daß seine Äußerung von x in A einen Effekt E hervorruft (Reaktion), und zwar über A's Erkennen dieser Absicht.]

4.5.10 Grices Problem, die Sprachevolution zu erklären

Grice definiert die Satzbedeutung als Äußerungstyp und Äußerungsbedeutung als Token dieses Äußerungstyp.

Problem:

Wie kann sich ein Typ *entwickeln* wenn er bei der ersten Äußerungsbedeutung bereits vorausgesetzt wird?

4.5.11 Widersprüchliche Verwendung von Konvention

Wörtliche Verwendung: Vermittlung von Absichten durch Einhalten der Konventionen

Metaphorische Verwendung: Vermittlung von Absichten durch Verletzung der Konventionen

4.5.12 Grundbegriffe für eine programmiertechnische Implementierung geeignet?

Erkennen einer Absicht, Hervorrufen eines Effekts, Intendieren für eine Zuhörerschaft...

4.5.13 Erfolgreiche Mensch-Maschine-Kommunikation

L = eine natürliche Sprache, SH = ein menschlicher Sprecher-Hörer von L, KAG = ein kognitiver Agent.

- *Erfolgreiches Verstehen*

Als Hörer kommuniziert KAG erfolgreich in L, wenn KAG die L-Äußerung von SH so versteht, wie SH sie gemeint hat. Das heißt technisch, daß KAG in seiner Datenbank die Sprecherbedeutung der L-Äußerung korrekt aufbaut. Ob dies tatsächlich geschehen ist, können die Entwickler überprüfen, weil sie selbst (i) die Äußerung in L korrekt verstehen (als natürliche Sprecher von L) und (ii) die Interpretation von KAG anhand der internen Datenbankstruktur direkt überprüfen können.

- *Erfolgreiches Äußern*

Als Sprecher kommuniziert KAG erfolgreich in L, wenn KAG sich in L so äußert, daß SH das von KAG Gemeinte versteht. Das heißt technisch, daß KAG einen gegebenen Informationskomplex in seiner Datenbank so in eine L-Äußerung abbildet, daß sie für SH korrekt verständlich ist. Ob diese L-Äußerung die in Frage stehende Information tatsächlich adäquat darstellt, können die Entwickler überprüfen, weil sie (i) direkten Zugang zu KAGs Sprecherbedeutung in dessen interner Datenbankstruktur haben und (ii) die Adäquatheit der L-Äußerung als natürliche Sprecher von L überprüfen können.

5. Anwendung von Zeichen auf den Kontext

5.1 Bühlers Organon-Modell

5.1.1 Theorie der Pragmatik

Analysiert die allgemeinen Prinzipien sinnvollen Handelns.

Sie analysiert, wie ein kognitiver Agent – unter Verwendung bestimmter Mittel in bestimmten Situationen – bestimmte Ziele erreichen kann.

5.1.2 Beispiele einer pragmatischen Aufgabenstellung

- Die Verwendung eines Schraubenziehers, um eine Schraube festzuziehen.
- Die Verwendung der Beine, um von a nach b zu gehen.
- Die Durchsuchung des Kühlschranks, um ein Mayonnaisesandwich zuzubereiten und damit seinen Hunger zu stillen.
- Die Bitte an jemanden, ein solches Sandwich zuzubereiten und zu bringen.

5.1.3 Linguistische und nichtlinguistische Pragmatik

Je nachdem, ob die eingesetzten Mittel sprachlicher Natur sind oder nicht, sprechen wir von Sprachpragmatik oder Kontextpragmatik.

5.1.4 Einbettung der sprachlichen in die nichtsprachliche Pragmatik

So wie die Spracherkennung als eine phylo- und ontogenetische Spezialisierung des nichtsprachlichen Wahrnehmens und die Sprachäußerung als eine phylo- und ontogenetische Spezialisierung des nichtsprachlichen Handelns analysiert werden kann (siehe Sektion 4.1), so ist auch die Sprachpragmatik eine phylo- und ontogenetische Spezialisierung der Kontextpragmatik.

5.1.5 Sprache als ein Organon

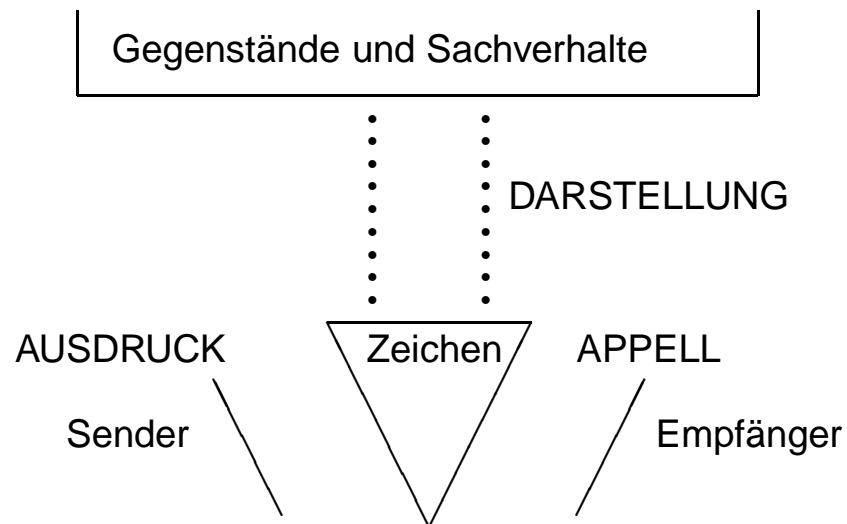
PLATO (427(?)–347 BC)

KARL BÜHLER (1879–1963 AD)

5.1.6 Der Werkzeugcharakter der Sprache

Die Sprache ist dem Werkzeug verwandt; auch sie gehört zu den Geräten des Lebens, ist ein Organon wie das dingliche Gerät, das leibesfremde Zwischending; die Sprache ist wie das Werkzeug ein *geformter Mittler*. Nur sind es nicht die materiellen Dinge, die auf den sprachlichen Mittler reagieren, sondern es sind die lebenden Wesen, mit denen wir verkehren.

5.1.7 Schema von Böhlers Organon-Modell



Die *Darstellung* bezieht sich auf die sprachliche Übertragung von Informationen.

Der *Ausdruck* bezeichnet die Art, in der der Sender das Zeichen hervorbringt.

Der *Appell* bezeichnet die Art, in der das Zeichen den Empfänger über die Darstellung hinaus bewegt.

5.1.8 Informationstheorie von Shannon & Weaver 1949

Zentrale Begriffe sind neben Sender und Empfänger die Bandbreite des Kanals, die Redundanz und relative Entropie des Codes und der Einfluß des Rauschens auf die Signalübertragung.

5.1.9 Das Organon-Modell und CURIOS (4.1.3) im Vergleich

Das Organon-Modell beschreibt die Beziehung zwischen dem ‘Sender’ und dem ‘Empfänger’ von Außen und deshalb auf unmittelbare Referenz beschränkt.

Das SLIM-Modell von CURIOS beschreibt die interne Struktur des Sprecher-Hörers und kann deshalb neben der unmittelbaren auch die mittelbare Referenz behandeln.

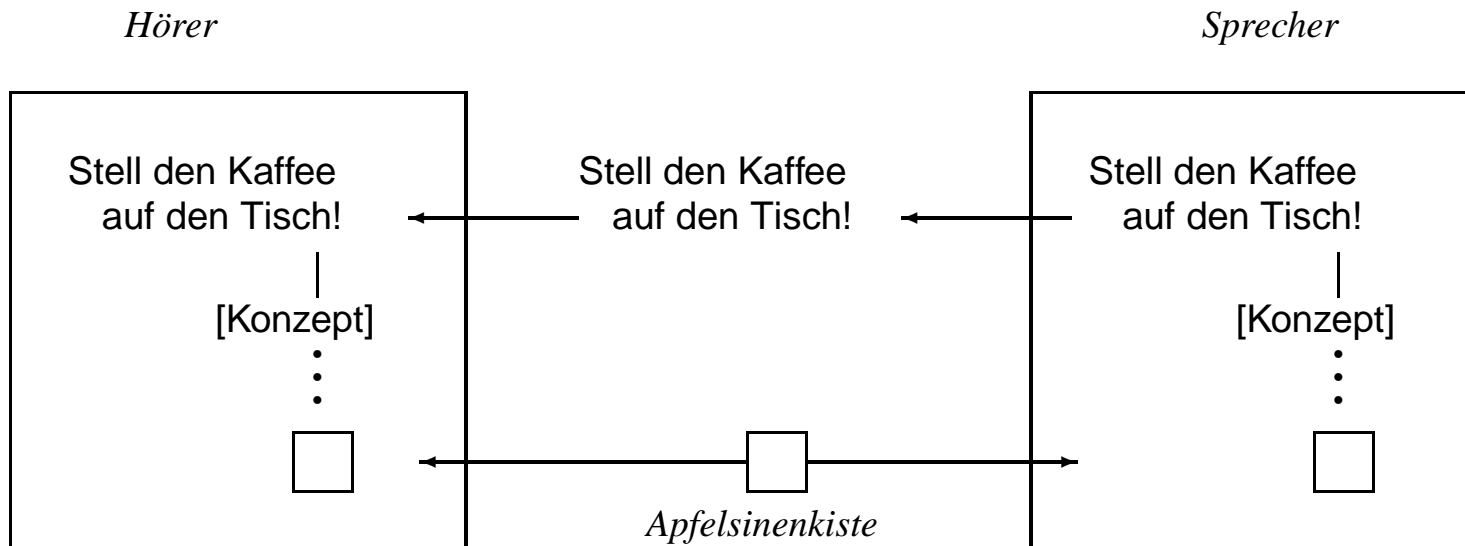
Die Organon-Funktion des ‘Ausdrucks’ ist der Komponente 5+ (Sprachsynthese) von CURIOS zuzuordnen.

Die Organon-Funktion des ‘Apells’ ist der Komponente 1+ (Spracherkennung) von CURIOS zuzuordnen.

Die Organon-Funktion der ‘Darstellung’ wird von CURIOS mit Hilfe einer Lexikon-, Syntax- und Semantik-Komponente in dem sprachlichen Datenbankkomplex 2+ geleistet und in Verbindung mit der nichtsprachlichen Datenbankstruktur 2 interpretiert.

5.2 Pragmatik der Werkzeuge und Pragmatik der Wörter

5.2.1 Eine nichtwörtliche Verwendung von Tisch



5.2.2 Entscheidende Frage der Sprachpragmatik

Wie kodiert der Sprecher die genauen Koordinaten seines begrenzten Verwendungskontexts in das sprachliche Zeichen, und wie kann der Hörer diese Abgrenzungen korrekt erschließen?

5.3 Finden des korrekten Interpretationskontexts

5.3.1 Postkartenbeispiel

New York, 1. Dezember 1999

Liebe Heidi,

Deinem Hund geht es gut. Das Wetter ist sehr kalt. Am Morgen spielte er im Schnee. Dann fraß er einen Knochen. Im Moment sitze ich gerade in der Küche. Fido ist auch hier. Das Wollknäuel hat ihn wieder angefaucht. Wir vermissen Dich.

Alles Liebe,

Siegfried

5.3.2 Ursprungsparameter eines Zeichens (STAR-Punkt)

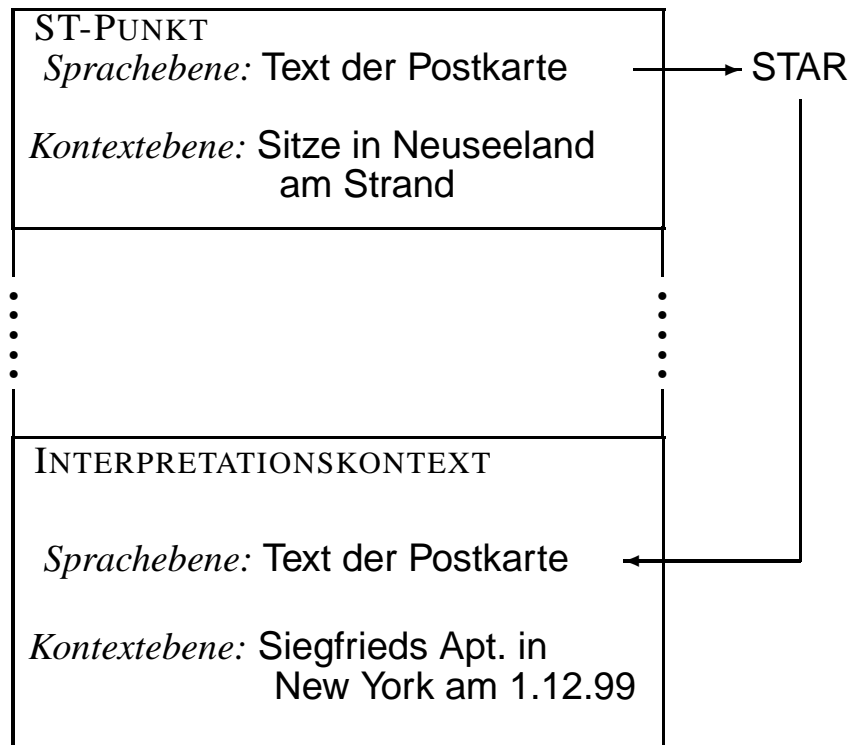
1. S = der räumliche Ort der Entstehung (S für *space*)
2. T = der Zeitpunkt der Entstehung (T für *time*)
3. A = der Autor
4. R = der intendierte Rezipient

5.3.3 Zweiter Hauptsatz der Pragmatik (PoP-2)

Der STAR-Punkt des Zeichens leistet seine Primärpositionierung in der kontextuellen Datenbank, indem er den *Einstiegskontext* bestimmt.

5.3.4 Primärpositionierung über den STAR-Punkt

Heidis kognitive Repräsentation:

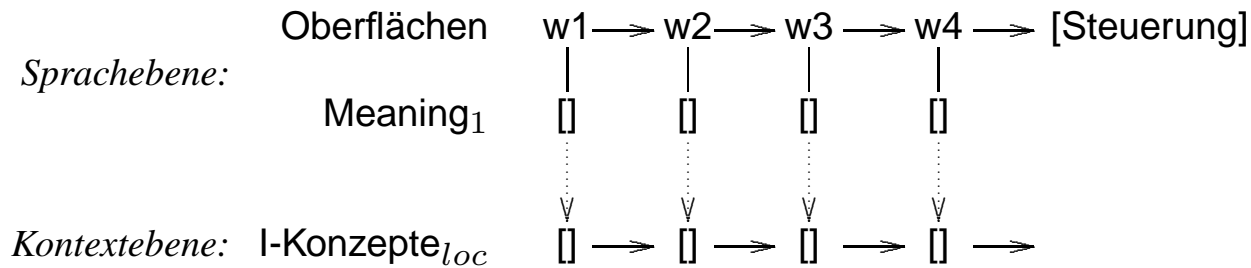


5.3.5 Fiktiver STAR-Punkt: Anfang von ‘Felix Krull’

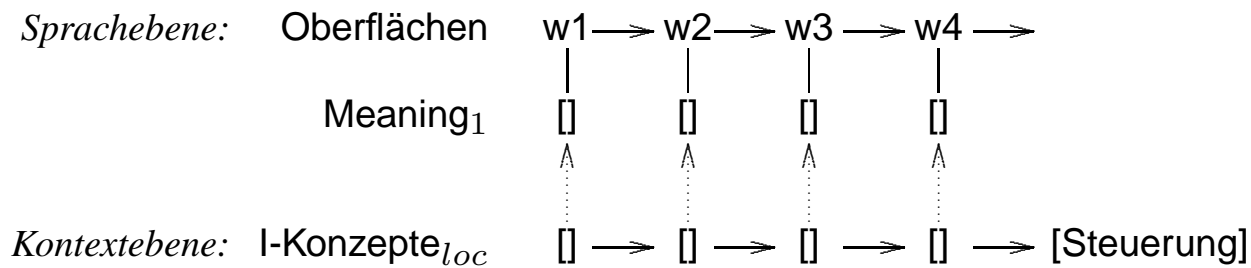
Indem ich die Feder ergreife, um in völliger Muße und Zurückgezogenheit – gesund übrigens, wenn auch müde, sehr müde ...

5.4 Sprachproduktion und -interpretation

5.4.1 Zeitlineare Interpretation



5.4.2 Zeitlineare Produktion



5.4.3 Zeitlineare Grundstruktur sprachlicher Zeichen

Die elementare Grundstruktur natürlicher Zeichen ist ihre *zeitlineare Reihenfolge*. Dies gilt für die einzelnen Sätze im Text, die einzelnen Wortformen im Satz und die einzelnen Morpheme in der Wortform.

Zeitlinear heißt:

linear wie die Zeit und in der Richtung der Zeit.

5.4.4 De Saussure's zweites Gesetz: *Linearer Charakter des Bezeichnenden*

SECOND PRINCIPE; CARACTÈRE LINÉAIRE DU SIGNIFIANT.

Le signifiant, étant de nature auditive, se déroule dans le temps seul et a les caractères qu'il emprunte au temps: a) *représente une étendue*, et b) *cette étendue est mesurable dans une seule dimension*: c'est une ligne.

Ce principe est évident, mais il semble qu'on ait toujours négligé de l'énoncer, sans doute parce qu'on l'a trouvé trop simple; cependant il est fondamental et les conséquences en sont incalculables; son importance est égale à celle de la première loi. Tout le mécanisme de la langue en dépend.

[Das Bezeichnende als etwas von Natur aus Auditives verläuft ausschließlich in der Zeit und hat Eigenschaften, die der Zeit entlehnt sind: a) es stellt eine Ausdehnung dar, und b) diese Ausdehnung ist meßbar in einer einzigen Dimension: sie ist eine Linie.

Dieser Grundsatz ist offensichtlich, aber es scheint, daß man bisher immer versäumt hat, ihn auszusprechen, zweifellos weil man ihn für zu einfach hielt; er ist jedoch grundlegender Art, und seine Konsequenzen sind unabsehbar; er ist ebenso wichtig wie das erste Gesetz. Der ganze Mechanismus der Sprache hängt davon ab.]

F. de Saussure 1913/1972, p. 103

5.4.5 Dritter Hauptsatz der Pragmatik (PoP-3)

Das Abgleichen der Wortformen auf ihre jeweiligen Teilkontextkonzepte erfolgt inkrementell, wobei die Elementarzeichen bei der Produktion der zeitlinearen Reihenfolge des zugrunde liegenden Gedankenpfades folgen, während bei der Interpretation der Gedankenpfad der zeitlinearen Reihenfolge der einlaufenden Elementarzeichen folgt.

5.5 Denken als Motor spontaner Sprachproduktion

5.5.1 Das ehemals berühmte Motto des Behaviorismus

THOUGHT IS NONVERBAL SPEECH

5.5.2 Das Motto der SLIM-Sprachtheorie

SPEECH IS VERBALIZED THOUGHT.

Das Denken wird als zeitlineare Navigation eines Fokuspunkts durch die verketteten Propositionen einer internen Datenbank definiert.

5.5.3 Die Rolle der zeitlinearen Ordnung für die semantische Interpretation

Ursprüngliche Ordnung:

Am Morgen spielte er im Schnee. Dann fraß er einen Knochen.

Umgekehrte Order (inkohärent):

Dann fraß er einen Knochen. Am Morgen spielte er im Schnee.

5.5.4 Alternative Navigation durch den propositionalen Inhalt (anti-temporale Sequenzierung)

Am Morgen fraß Fido einen Knochen. Vorher spielte er im Schnee.

5.5.5 Veränderung der Interpretation durch Änderung der Sequenzierung

(a) 1. Im Februar war ich am Äquator. 2. Dort war es sehr heiß. 3. Im März war ich in Alaska. 4. Dort war es sehr kalt.

(b) 3. Im März war ich in Alaska. 2. Dort war es sehr heiß. 1. Im Februar war ich am Äquator. 4. Dort war es sehr kalt.

5.5.6 Die Zeitlinearität des Gesprochenen

Speech is irreversible. That is its fatality. What has been said cannot be unsaid, except by adding to it: to correct here is, oddly enough, to continue.

[Das Gesprochene ist irreversibel. Das ist seine Fatalität. Was gesagt worden ist, kann nicht ungesagt gemacht werden, außer indem hinzugefügt wird: Korrigieren bedeutet hier – merkwürdigerweise – fortsetzen.]

R. Barthes, 1986, p. 76

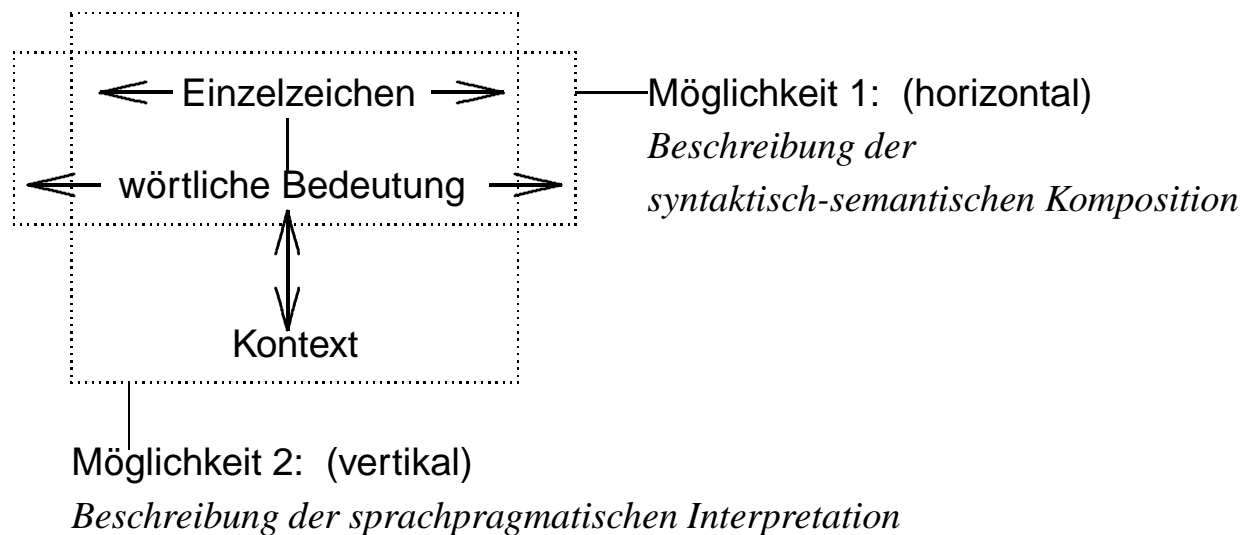
6. Struktur und Funktionsweise der Zeichen

6.1 Referenzmechanismen verschiedener Zeichentypen

6.1.1 Die bisherigen Prinzipien der natürlichen Kommunikationsmechanik

1. Zeichenverwendung als interne Abgleichung zwischen der Meaning_1 und dem Kontext (PoP-1, Sektion 4.3).
2. Bestimmung des Einstiegskontexts über den STAR-Punkt der Äußerung (PoP-2, Sektion 5.3).
3. Festlegung der Ableitungsordnung über die zeitlineare Abfolge der Wortformen (PoP-3, Sektion 5.4).

6.1.2 Alternativen der Beschreibung



6.1.3 Beispiel mit minimaler Syntax

Ich erwachen. Wetter schön. Schnell aufstehen. Laune gut. Küche gehen. Kaffee machen. Wo Brötchen? Wollen essen. Aber frisch!

6.1.4 Vierter Hauptsatz der Pragmatik (PoP-4)

Der Referenzmechanismus des Zeichentyps **Symbol** beruht auf einer Meaning_1 , die als M-Konzept definiert ist. Symbole referieren von ihrem Platz im positionierten Satz, indem sie ihr M-Konzept auf geeignete kontextuelle Referenten (I-Konzepte_{loc}) abgleichen.

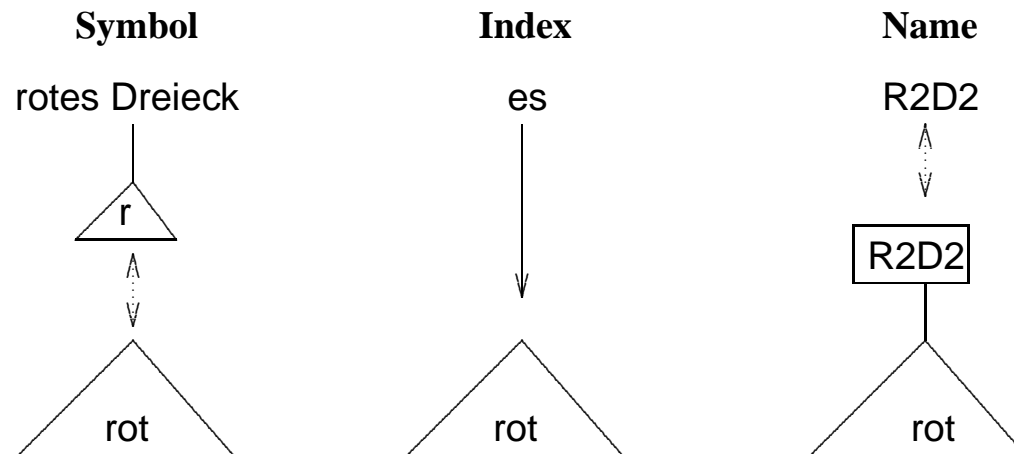
6.1.5 Fünfter Hauptsatz der Pragmatik (PoP-5)

Der Referenzmechanismus des Zeichentyps **Index** beruht auf einer Meaning_1 , die als Zeiger definiert ist. Indizes referieren von ihrem Platz im positionierten Satz, indem sie auf für sie jeweils charakteristische Parameterbereiche des Sprechers zeigen.

6.1.6 Sechster Hauptsatz der Pragmatik (PoP-6)

Der Referenzmechanismus des Zeichentyps **Name** basiert auf einem Akt der Namensgebung, der darin besteht, in die interne Repräsentation des entsprechenden Individuums oder Objekts einen *Marker* einzufügen. Die Referenz mit einem Namen funktioniert über die interne Abgleichung der Namensoberfläche mit dem entsprechenden Marker.

6.1.7 Referenz mit Nominalphrase, Pronomen und Namen



6.1.8 Referenz in nicht- und vorsprachlicher Kommunikation

1. *ikonische* Referenz aus dem spontanen *Nachahmen* des Referenten mit Hilfe von Gesten und Lauten,
2. *indexikalische* Referenz aus dem gestischen *Zeigen* auf den Referenten und
3. *namensbasierte* Referenz aus dem gestischen *Zeigen* auf den Referenten und dem *gleichzeitigen Aussprechen* eines als Namen geeigneten Lautes.

6.1.9 Siebter Hauptsatz der Pragmatik (PoP-7)

Der Zeichentyp *Symbol* tritt als Substantiv, Verb und Adjektiv-Adverb auf. Der Zeichentyp *Index* tritt als Substantiv und Adjektiv-Adverb auf. Der Zeichentyp *Name* tritt nur als Substantiv auf.

6.1.10 Verhältnis von Zeichentypen und Grundwortarten

Name	Peter		
Index	er	hier	
Symbol	Mann	alt	sehen
	Substantiv	Adj-Adv	Verb

6.2 Innere Struktur von Symbol und Index

6.2.1 Struktureller Aufbau von Symbolen und Indizes

- Die *Oberfläche* des Sprachzeichen folgt strukturell primär den Gesetzen der akustischen Artikulation.
- Die *Kategorie* folgt den kombinatorischen Eigenschaften der Sprachklassen und der Flexionsunterklassen, der das Sprachzeichen angehört.
- Die *Meaning*₁ folgt strukturell den Konzeptstrukturen des internen Kontexts oder enthält Zeiger auf bestimmte kontextuelle Parameterwerte.
- Die *Verknüpfung* zwischen Oberflächen, Kategorien und wörtlichen Bedeutungen basiert in der Regel auf Konventionen, die von jedem Sprecher-Hörer erlernt werden müssen.

6.2.2 De Saussures erstes Gesetz

PREMIER PRINCIPE; L'ARBITRAIRE DU SIGNE.

Le lien unissant signifiant au signifié est arbitraire, ou encore, puisque nous entendons par signe le total résultant de l'association d'un signifiant à un signifié, nous pouvons dire plus simplement: *le signe linguistique est arbitraire*. [[ERSTES PRINZIP: DIE BELIEBIGKEIT DES ZEICHENS

Das Band, das das Bezeichnete mit der Bezeichnung verknüpft, ist beliebig; und da wir das Zeichen in seiner Gesamtheit als eine assoziative Verbindung einer Bezeichnung mit einem Bezeichneten verstehen, so können wir dafür auch einfacher sagen: *Das sprachliche Zeichen ist beliebig*.]

F. de Saussure 1913/1972, S. 100

6.2.3 Die möglichen Funktionen des Zeichentyps Symbol

1. *Erstreferenz*: Bezug auf ein Objekt, das bisher noch nicht erwähnt wurde (Beispiel 4.2.2).
2. *Referenzwiederholung*: Bezug auf ein Objekt, das bereits sprachlich eingeführt wurde (Beispiele 6.3.8 und 6.3.12), mit der Möglichkeit, den Referenten zusätzlich zu charakterisieren.
3. *Metaphorische Referenz*: Bezug auf einen partiell kompatiblen Referenten (Beispiel 5.2.1), sowohl bei Erstreferenz als auch bei Referenzwiederholung.

6.3 Referenzwiederholung

6.3.1 Reine Indizes

Indexwörter, die keine zusätzlichen grammatischen oder symbolischen Bedeutungskomponenten enthalten, z. B. hier und jetzt.

6.3.2 Nicht-reine Indizes

Indexwörter, die symbolisch-grammatischen Unterscheidungen enthalten, z. B. zwischen Singular (ich, Du) und Plural (wir, ihr) oder zwischen den verschiedenen Kasus (ich, mir, mich, Du, Dir, Dich etc.).

6.3.3 Zeigebereich der Pronomina der dritten Person

Er liegt außerhalb der STAR-Punkt-Parameter und umfaßt alle Personen und Objekte, die bisher aktiviert wurden und nicht Sprecher oder Hörer sind.

6.3.4 Referenzwiederholung

Diese spezielle Aufgabe ergibt sich in längeren Sätzen oder Texten, wo ein Referent bereits mit Hilfe einer Nominalphrase oder eines Namens eingeführt wurde, und wo erneut auf diesen Referenten Bezug genommen werden soll.

6.3.5 Referenzwiederholung mit Pronomina der dritten Person

Aufgrund ihrer symbolisch-grammatischen Differenzierung und ihres allgemeinen Zeigebereichs sind Pronomina der dritten Person für eine knappe, präzise und vielseitige Realisierung der Referenzwiederholung ideal geeignet.

6.3.6 Indexikalische Referenzwiederholung

Nachdem der kleine Hund nach Hause kam, fraß er einen Knochen.

The diagram shows the sentence "Nachdem der kleine Hund nach Hause kam, fraß er einen Knochen." with underlines under "der kleine Hund" and "er". A vertical arrow points down from "der kleine Hund", and a curved arrow points from "er" back to "der kleine Hund", indicating that "er" refers back to the noun phrase.

6.3.7 Indexikalische Referenzwiederholung ohne Koreferenz

Nachdem der kleine Hund nach Hause kam, fraß er einen Knochen.

The diagram shows the sentence "Nachdem der kleine Hund nach Hause kam, fraß er einen Knochen." with underlines under "der kleine Hund" and "er". Vertical arrows point down from both "der kleine Hund" and "er", indicating that they are treated as separate, independent referents without coreference.

6.3.8 Symbolische Referenzwiederholung I

Nachdem er nach Hause kam, fraß der kleine Hund einen Knochen.

The diagram shows the sentence "Nachdem er nach Hause kam, fraß der kleine Hund einen Knochen." with underlines under "er" and "der kleine Hund". A vertical arrow points down from "er", and a curved arrow points from "der kleine Hund" back to "er", indicating that "er" refers back to the noun phrase.

6.3.9 Symbolische Referenz ohne Koreferenz

Nachdem er nach Hause kam, fraß der kleine Hund einen Knochen.

6.3.10 Durch Satzstruktur geblockte Referenzwiederholung

1. Anaphorisch:

Als Peter_i nach Hause kommt, ist er_i erfreut.

Peter_i ist erfreut, als er_i nach Hause kommt.

% Neben Peter_i sieht er_i eine Schlange.

2. Kataphorisch:

Als er_i nach Hause kommt, ist Peter_i erfreut.

% Er_i ist erfreut, als Peter_i nach Hause kommt.

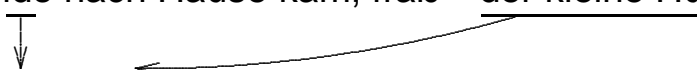
Neben sich_i sieht Peter_i eine Schlange.

6.3.11 Intersententiale Referenzwiederholung

Peter_k wollte ins Grüne fahren. Er_k wartete auf Fido_i. Als der kleine Hund_i nach Hause kam, freute er_k sich_k.

6.3.12 Symbolische Referenzwiederholung II

Nachdem Fido nach Hause kam, fraß der kleine Hund einen Knochen.



6.3.13 Etablierung initialer Referenz mit Symbol, Index und name

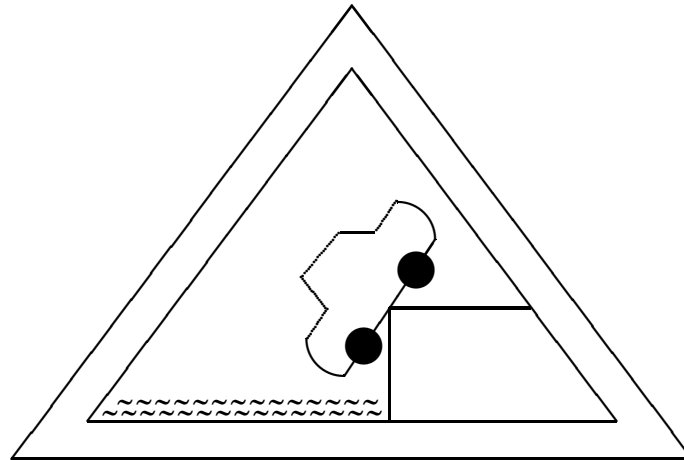
	der kleine Hund	
Nachdem	er	nach Hause kam, schlief er.
	Fido	

6.3.14 Referenzwiederholung mit Symbol, Index und Name

		der kleine Hund.
Nachdem er nach Hause kam,	schlief	er.
		Fido.

6.4 Sonderrolle von Ikons und Namen

6.4.1 Beispiel eines Ikons (Straßenschild)



6.4.2 Der Streit zwischen den *Naturalisten* und den *Konventionalisten*

[The naturalists] maintained that all words were indeed ‘naturally’ appropriate to the things they signified. Although this might not always be evident to the layman, they would say, it could be demonstrated by the philosopher able to discern the ‘reality’ that lay behind the appearance of things. Thus was born the practice of conscious and deliberate etymology. The term itself (being formed from the Greek stem *etymo-* signifying ‘true’ or ‘real’) betrays its philosophical origin. To lay bare the origin of a word and thereby its ‘true’ meaning was to reveal one of the truths of ‘nature’.

[Die Naturalisten behaupteten, alle Wörter und die Dinge, die sie bezeichneten, paßten tatsächlich ‘natürlich’ zusammen. Zwar sei dies für den Laien vielleicht nicht immer zu erkennen, so pflegten sie zu argumentieren, es könne aber vom Philosophen nachgewiesen werden, der fähig sei, die ‘Wirklichkeit’ hinter der Erscheinung der Dinge zu sehen. So entstand das bewußte und überlegte Etymologisieren. Der Ausdruck selbst (aus dem griechischen Stamm *etymo-* ‘wahr’ oder ‘wirklich’ gebildet) verrät seine philosophische Herkunft. Den Ursprung eines Wortes und damit seine ‘wahre’ Bedeutung aufzudecken, hieß, eine der Wahrheiten der ‘Natur’ aufzuspüren.]

J. Lyons 1968, p. 4 f.

6.4.3 Strukturvergleich von Symbol und Ikon

Symbol: Oberfläche
|
unmotiviert
|
M-Konzept

Ikon: Oberfläche
|
motiviert
|
M-Konzept

6.5 Bilder, Piktogramme und Buchstaben

6.5.1 Übergang vom Bild über Ikon zum Symbol

