

Interdisziplinäre  
fachdidaktische Übung:  
Sprachen und  
Modelle

SS 2016: Grossmann, Jenko

# Sprache und Modelle

- Natürliche und formale Sprachen – Vergleich
- Natürliche und formale Sprachen – Computational Thinking
- Natürliche und formale Sprachen – Patterns
- Beispiel für Patterns
- Modelle
- Entwicklung von Modellen
- Modelle und Formale Sprachen

# Natürliche und Formale Sprachen – Vergleich

SPRACHE	
Natürliche Sprache	Formale Sprache
Ähnliche Begriffe in der Beschreibung, unterschiedliche Interpretation	
Alphabet = Zeichen	Alphabet = Zeichen
Wort = ausgezeichnete Zeichenkette	Wort = beliebige Zeichenkette
Wortschatz = alle Wörter Diese werden typisiert Subst., Verb,..., aber auch nach Inhalt	Reservierte Worte, freie Wortbildung Freie Worte können typisiert werden (Objekte)
Auf Basis der Typisierung können wir Sätze bilden. Regeln sind durch eine sehr allgemeine Grammatik bestimmt. Es gibt keine korrekte und vollständige Grammatik (z.B. Kontextproblem). Wir hatten nur ein Sprachmodell.	Es gibt eine Hierarchie von Grammatiken. Im Wesentlichen sind das Ersetzungsregeln, meist deterministisch, ev. Auch stochastisch. Im Vordergrund steht die Frage der Entscheidbarkeit.

# Natürliche und Formale Sprachen – Vergleich

Zentrale Aufgabe von Sprachen sind  
Kommunikation und Selbsta Ausdruck

KOMMUNIKATION UND SELBSTAUSDRUCK	
Natürliche Sprache	Computersprache
Kommunikation zwischen Personen, wobei unterschiedliche Funktionen erfüllt werden müssen.	Unterschiedliche Kommunikations-szenarien: Mensch - Maschine, Maschine - Maschine, Maschine - Mensch, Mensch - Maschine - Mensch
Kommunikation soll Nachrichten übermitteln. Nachricht ist durch Funktion (Information, Kontrolle, soziale Interaktion, Emotion, Gedanken, Identität,...) bestimmt, oft mehrere Funktionen in einer Nachricht (Nachrichtenquadrat)	Die Kommunikation ist immer ein Algorithmus

# Natürliche und Formale Sprachen – Vergleich

Unterscheide verschiedene Betrachtungsweisen  
von Kommunikation für den Menschen  
(R. Burkhart: Kommunikationswissenschaft)

# Natürliche und Formale Sprachen – Computational Thinking

Können wir mit ***Computational Thinking*** die algorithmische Kommunikation näher zur menschliche Kommunikation bringen?

***Computational Thinking*** heißt nicht denken wie ein Computer, sondern Entwicklung Strategien zur Lösung von Problemen

Der Begriff geht auf S. Papert zurück, in neuer Version auf J. Wing

<https://www.cs.cmu.edu/~15110-s13/Wing06-ct.pdf>

# Natürliche und Formale Sprachen – Computational Thinking

Britische Definition:

<http://www.bbc.co.uk/education/guides/zp92mp3/revision>

ACM Definition:

<https://csta.acm.org/Curriculum/sub/CurrFiles/CompThinkingFlyer.pdf>

# Natürliche und Formale Sprachen – Patterns

Wie können Prinzipien des Computational Thinking angewendet werden um die Leistungen natürlicher Sprachen in der Kommunikation zu erfüllen?

Pattern Theory nach U. Grenander ist ein Ansatz bei dem man natürliche Sprachen als **reguläre Strukturen** verstehen kann, die Elemente des Computational Thinking enthalten



# Natürliche und Formale Sprachen – Patterns

## Grundideen:

- Atomistischen Aufbau welche reguläre Strukturen algebraisch aus Grundbausteinen aufbaut
- Probabilistische Komponente kann beim Aufbau der Struktur an verschiedenen Stellen integriert werden
- Empiristisch, d.h. nicht die formale Verifikation steht im Vordergrund sondern der Vergleich von beobachteten Strukturen mit den vom Modell bei Simulationen generierten Strukturen

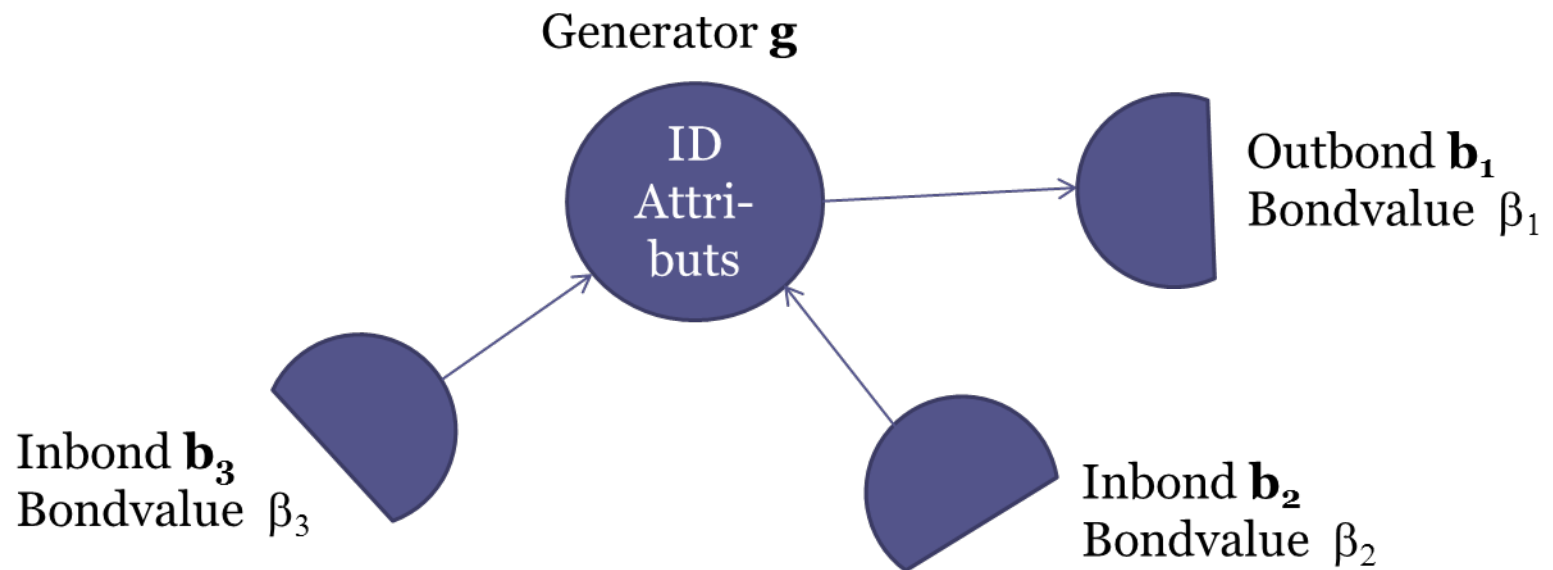
# Natürliche und Formale Sprachen – Patterns

Grundbaustein sind Generatoren

- Jeder Generator wird durch einen Identifier identifiziert und kann durch Attribute beschrieben werden, z.B. Worte mit Attribut Wortarten
- Jeder Generator hat Verbindungselemente, sogenannte Bonds, welche die Verbindung mit anderen Generatoren erlauben
  - Die Bonds haben eine Orientierung und werden durch charakteristische Bondvalues beschrieben

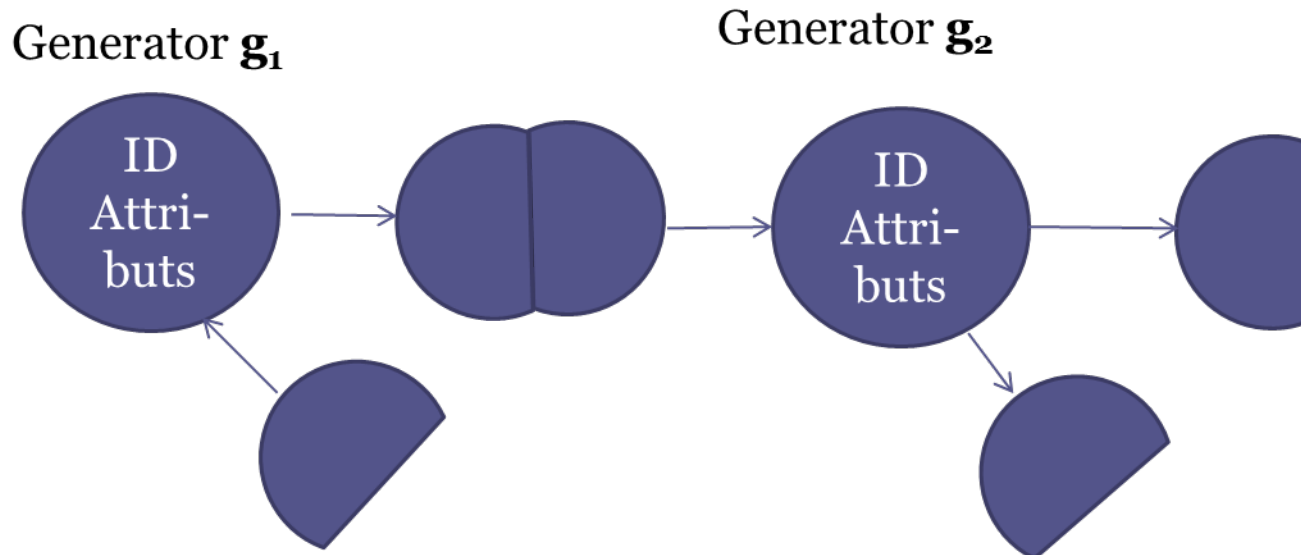
# Natürliche und Formale Sprachen – Patterns

## Symbolische Darstellung



# Natürliche und Formale Sprachen – Patterns

Eine Konfiguration ist eine Menge von Generatoren, die durch die Bonds miteinander verknüpft sind (können auch nach stochastischen Regeln verbunden sein)



# Natürliche und Formale Sprachen – Pattern Theory

Im allgemeinen lassen sich die Konfigurationen nicht beobachten, sondern durch Deformationen gestörte oder unvollständige Äquivalenzklassen von solchen Konfigurationen

Beispiel: Unterschiedliche Formulierungen einer Aussage bilden eine Äquivalenzklasse

Die Nachrichtenübermittlung kann durch Mehrdeutigkeit der Wörter gestört sein, oder es fehlt ein Bestandteil des Nachrichtequadrats

# Natürliche und Formale Sprachen – Patterns

Wichtige Aufgaben der Pattern Theory:

- Pattern Synthesis (Simulation)
  - Definiere einen Konfigurationsraum
  - Simuliere die sich daraus ergebenden Images
  - Vergleiche die simulierten Images (Konfigurationen) mit der Wirklichkeit
- Pattern Recognition
  - Bestimme die Klasse zu der ein Image gehört (Supervised Learning)

# Natürliche und Formale Sprachen – Patterns

Wichtige Aufgaben der Pattern Theory:

- Image Segmentation
  - Definiere Klassen von Patterns (Unsupervised Learning)
- Image Restoration
  - Stelle aus dem deformierten Image das „richtige“ Image her

# Beispiele für Patterns

Patterns des Übermittlungsmediums: Grundlage ist die klassische Signalverarbeitung (Wie verstehen Computer natürliche Sprachen?)

Patterns der Semantik: Entwicklung einer Ontologie für Worte

- [Wordnet](#) als Ontologie basierend auf linguistischen Begriffen
- [ConceptNet](#) als Ontologie des Alltagswissens

Patterns der Emotion: <http://sentic.net/computing/>



# Beispiele für Patterns

Patterns der Linguistik: Automatisches Erkennen von Wortarten und Satzstrukturen (Grammatik)

Patterns der Häufigkeit von Termen in Texten: Text Mining, automatische Keyword-Generierung

Gemeinsamkeit der verschiedene Patterns:

- Modell für bestimmten Teilaspekte von natürlichen Sprachen
- Formale Analyse im Rahmen dieses Modells

# Modelle

Referenz: Models in Science

<http://plato.stanford.edu/entries/models-science/>

Modell soll einen Ausschnitt der Realität derart repräsentieren, dass bestimmte Aspekte deutlich werden und Fragen über die Realität beantwortet werden können

# Modelle

Wesentliche Fragen:

Repräsentation:

Was repräsentiert ein Modell?

Präsentation:

Wie repräsentiert ein Modell?

Analyse:

Wie werden die Fragen analysiert?

# Modelle - Repräsentation

Was repräsentiert ein Modell?

- Phänomene = interessierende Eigenschaften erlauben eine Erklärung der Phänomene
  - Idealisierte Repräsentation: z.B. Atom-modelle
  - Analoge Repräsentation: z.B. Herz ist eine Pumpe
  - Phänomenologische Repräsentation: erkläre das Verhalten aus beobachtbaren Eigenschaften unter Berücksichtigung von statistischer Ungenauigkeit z.B. Verhalten von SchülerInnen

# Modelle - Repräsentation

Was repräsentiert ein Modell?

– Daten:

Wir lernen das Modell aus vielen Beispielen, die gespeichert wurden und Plausibilitäten für Interpretationen ergeben

Prinzip des Data Mining (Big Data)

# Modelle - Repräsentation

Anwendung im Automatischen Übersetzen

*„Die Bank steht auf einem freien Platz“*



60%



30%



10%

# Modelle - Repräsentation

Was repräsentiert ein Modell?

- Theorien: Ein Kalkül wird in einem bestimmten Kontext interpretiert, z.B. Repräsentation einer generativen Grammatiktheorie

# Modelle - Präsentation

Wie repräsentiert ein Modell?

- Ikonische Modelle (Bilder, Infographik), wesentlich für die Ergebnisdarstellung



Sprache und Modelle



# Modelle - Präsentation

Wie repräsentiert ein Modell?

- Sprachlich orientierte Modelle
  - Begriffssystem
  - Spezielle Notation
  - Strukturierte Darstellung, orientiert an:
    - Logisch-algebraischen Strukturen
    - Strukturen der Graphentheorie
    - Analytischen Strukturen (Analysis, Wahrscheinlichkeit, Statistik)

# Modelle – Analyse

Wie werden Modelle analysiert?

- Jede Repräsentation kennt entsprechend der Semantik der Repräsentation unterschiedliche Methoden der Analyse
- Dabei werden die Analysemethoden im Kontext der Modellrepräsentation interpretiert

# Modelle – Analyse

## Analyse in logisch algebraischen Modellen

- Das Semantic Web strebt eine logisch orientierte Formalisierung solcher Darstellungen der Semantik an

- Jede Aussage im System wird durch einen einfachen Satz der Form

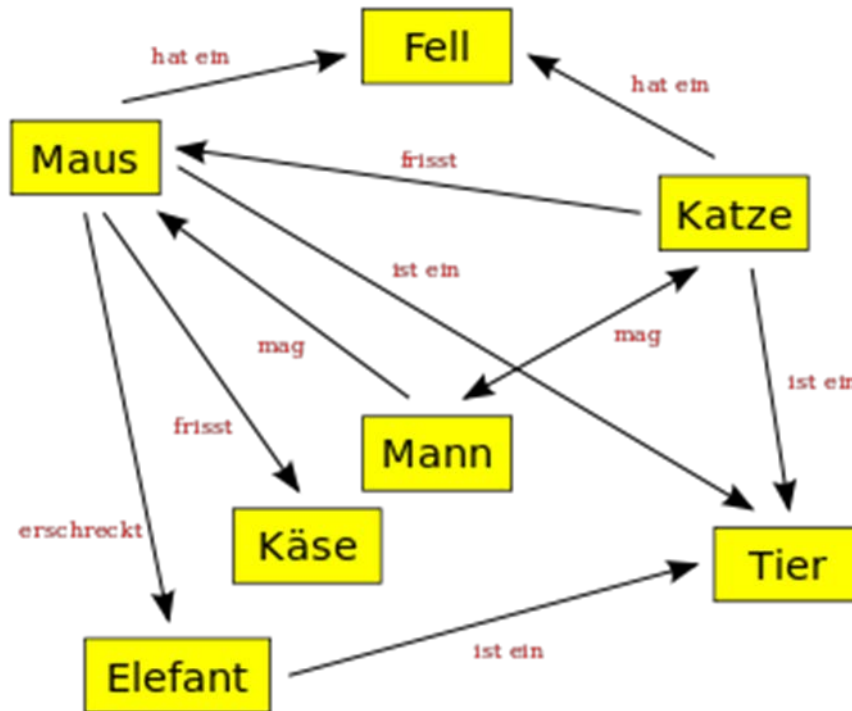
Subjekt – Prädikat – Objekt

dargestellt

# Modelle – Analyse

Analyse in logisch- algebraischen Modellen

Beispiel Mausontologie



# Modelle – Analyse

- Aussagen können automatisch mit logischen Regeln (Datenbankabfragen) überprüft werden
- Beispiel:

`"John is a married bachelor"`

Der Satz wird als falsch erkannt, wenn es im System eine Regel der Form gibt:

`"A bachelor is never married"`

# Modelle – Analyse

- Zwei mögliche Betrachtungen:
  - Closed World: Die vorhandenen Daten (Aussagen) repräsentieren die Welt. Alle Aussagen, die nicht dieser Welt widersprechen können dem System hinzugefügt werden.
    - "A bachelor is never married"wird hinzugefügt, wenn
    - "John is a married bachelor"nicht im System ist.

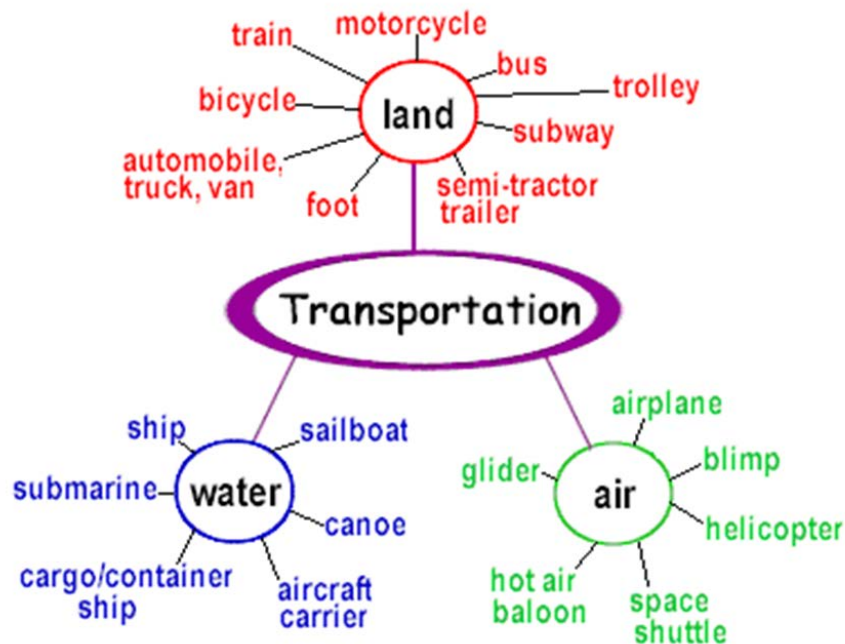
# Modelle – Analyse

- Open World: Auch wenn  
"John is a married bachelor"  
nicht zum Wissen der Welt gehört, wird  
anstelle von  
"A bachelor is never married"  
der folgende Satz eingefügt:  
„Married bachelors may be  
possible“

# Modelle – Analyse

Analyse in Modellen der Graphentheorie

- Erreichbarkeit, „Entfernung“, „Hierrarchie“





# Modelle – Analyse

## Analyse in analytischen Modellen

- Analysis:

Beispiel Erkennung von Audiosignalen

Methoden der Signalverarbeitung (Analyse der Frequenzen, Filterung)

Unterscheide Sprache - Musik

# Modelle – Analyse

## Analyse in analytischen Modellen

- Wahrscheinlichkeit

### Beispiel Part of Speech Tagger

Ein Satz wird als sogenannte Hidden Markov Chain interpretiert, d.h. im Sinne der Patterns sind die Generatoren die Wortarten, beobachtet werden aber die Wörter

# Modelle – Analyse

## Analyse in analytischen Modellen

- Statistik

### Beispiel Textmining

Die Häufigkeit der Terme (Worte) von Dokumenten werden in einer Term-Document Matrix dargestellt und die Häufigkeiten werden statistisch interpretiert  
Einfachste Analyse: Word Cloud

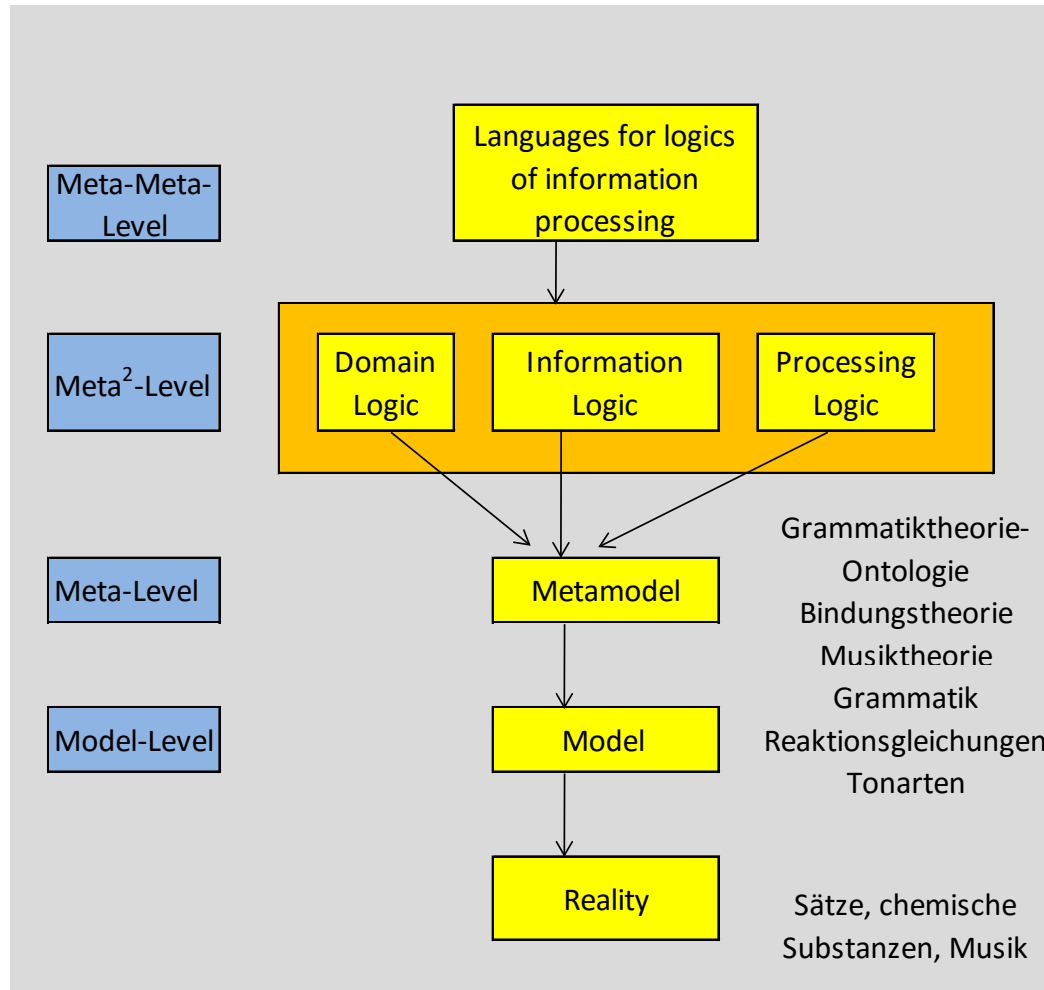
# Entwicklung von Modellen

- Eine häufig verwendete Möglichkeit ist die Definition eines Meta-modells zur Erzeugung von Modellen
- Metamodelle vereinigen:
  - Domainlogik: bestimmt die Semantik des Modells
  - Informationslogik: bestimmt die Syntax der Modellstruktur (Logische Regeln, Mathematik)
  - Prozesslogik: bestimmt Algorithmen und Prozeduren der Modellstruktur, die angewendet werden können (z.B. Musik Filter, Semantic Web Logik)

# Entwicklung von Modellen

- Die Frage wie entstehen Metamodell führt zu Modellen für Metamodelle, also  
Meta-Metamodelle
- Dies impliziert einen infiniten Regress
- Auflösung durch eine Trick:
  - Wir brauchen nur mehr eine Schicht darüber, die in eine Sprache zur Beschreibung von logischen Strukturen verwendet

# Entwicklung von Modellen



# Modelle und Formale Sprachen

- Formale Sprachen (Programmiersprachen) erlauben die Realisierung eines logisch-mathematischen Kalküls auf einer Maschine
- Die (Semantik) Bedeutung wird auf die Frage ob eine mit den im System zulässigen Methoden (Informationslogik, Prozesslogik) gebildete Aussage logisch wahr ist reduziert

# Modelle und Formale Sprachen

- Über diese Logik können wir ein Modell mit einer formalen Sprache verbinden
- Das Modell wird dadurch die Realisierung einer formalen Theorie



# Modelle und Formale Sprachen

- Schematische Darstellung

