

# **Berührungspunkte natürlicher und formaler Sprachen**

12.06.2017

Susanne Bednarik, Yvonne Friedrich, Katja Schirmer, Lisa Türk

050305 PR Interdisziplinäre Fachdidaktik Informatik: Natürliche Sprachen und formale Sprachen, Sommersemester 2017

# Agenda

## Natürliche Sprachen

- Definition und Forschungsbereiche
- Theorien von Noam Chomsky
- Peter Norvig, statistische Sprachmodelle

## Formale Sprachen

- Logische Semantik vs. Linguistische Semantik
- Die gemeinsamen Eigenschaften
- Merkmale einer formalen Sprache
- Fallbeispiel Human Ressource Maschine



# Implizites Sprachproduktionsmodell



Pragmatik

Semantik

Syntax

Lexik

Morphologie

Phonetik

PRA	„Innerhalb eines bestimmten Kontextes, und unter Berücksichtigung sozialer Konventionen, hat der Sprecher
SEM	vor, einen bestimmten Inhalt zum Ausdruck zu bringen. Diese `Bedeutung´ wird innerhalb einer ausgewählten
SYN	sprachlichen Struktur und durch Auswahl angemessener
LEX	lexikalischer Begriffe, die unter Berücksichtigung ihrer
MOR	syntaktischen Funktion u.U. inflektiert werden müssen,
PHO	als Sequenz von Lauten (mit entsprechender Intonation) von seinen Sprechorganen produziert.“

(Edmondson/House 2006, S. 72)

# Noam Chomsky (1)



<https://www.youtube.com/watch?v=jB54XxbgI0E> (ca. 4:00 bis 5:00)

# Noam Chomsky (2)

- seit 1957 linguistische Strukturanalyse auf der Syntaxebene bereits unter Berücksichtigung der Semantik
- Ziel: Definition der tieferen Eigenschaften von menschlicher Sprache im Allgemeinen
- Fokus auf der Kompetenz /*competence* des idealen Sprechers  
(= Kreativität, Fähigkeit zur Verknüpfung von Lauten und Bedeutung nach den Regeln einer Sprache),  
weniger auf der Performanz /*performance* (= tatsächlicher Sprachgebrauch)  
→ endliche Mittel und unendlicher Gebrauch einer Sprache
- Universalgrammatik → generative Grammatik

# Noam Chomsky (3)

## **Universalgrammatik**

gültig für alle natürlichen Sprachen



syntaktische Komponente

(begrenztes System von Regeln zur unendlichen Kombination von)

=

phonetische + semantische Komponente

(Laute /*surface structure* + Bedeutung /*deep structure*)



**generative Grammatik**

Anwendung auf Einzelsprache

# Noam Chomsky (4)

- Universalgrammatik → generative Grammatik
- Begründung: Spracherwerb
- *Language Acquisition Device* als Sprachorgan  
→ (angeborene) Universalgrammatik als Grundlage für die Sprachproduktion von Einzelsprachen

# Noam Chomsky (5)

- Aktuelle, enorme Bedeutung des generativen Ansatzes für
  - Syntaxforschung
  - Sprachtheorie
  - maschinelle Übersetzung
  - Spracherwerbsforschung/Kreolistik
  - ...
- Kritik: Konstrukt des idealisierten Sprechers ohne Berücksichtigung der gesellschaftlichen Bedingtheit von Sprache, ...

# Peter Norvig

- Google's Director of Research
- einer der einflussreichsten Linguisten derzeit
- Kritiker Chomsky's
  - ist der Meinung, statistische Sprachmodelle können die natürliche Sprache besser abbilden als Chomsky's Theorie



[https://de.wikipedia.org/wiki/Peter\\_Norvig](https://de.wikipedia.org/wiki/Peter_Norvig)

# Computerlinguistik

- Sprachverarbeitung früher (1960er Jahre): Basierend auf vorhandenen linguistischen Beschreibungsebenen wurde versucht natürliche Sprachen zu formalisieren.
- seit 1980er Jahre: Sprachverarbeitung mit statistischen Verfahren
- “trainierte” Wahrscheinlichkeitsmodelle
- heute: Kombination beider – D.h. zu den statistischen Verfahren wird die klassische Strukturanalyse hinzugefügt, um bessere Ergebnisse zu erzielen.  
→ hybride Systeme

# ein statistisches Sprachmodell

- ist eine Wahrscheinlichkeitsverteilung über die Abfolge von Wörtern
- basierend auf Daten
- heute viele Anwendungsfelder
- laut Chomsky unwissenschaftlich

# Wie erfolgreich sind statistische Sprachmodelle?

## aus Usersicht:

- Suchmaschinen: 100% trainiert und probabilistisch
- Spracherkennung: 100% trainiert und probabilistisch
- Übersetzungsprogramme: hybrider Ansatz
- Fragen beantworten: Watson trainiert und probabilistisch, andere z.T. hybride Ansätze

# Wie erfolgreich sind statistische Sprachmodelle?

## aus linguistischer Sicht:

- Word sense Disambiguation (WSD): Mehrdeutigkeit einzelner Wörter  
*Beispiel: Auflösen des Wortes "Bank"*
- Auflösung v. Koreferenz: verschiedene Ausdrücke mit gleicher Bedeutung  
*Beispiel: Klaus sagte, er schreibe für Wikipedia.*
- Part-of-speech Tagging  
*Beispiel: Petra/NE liest/VVFIN einen/ART langen/ADJA Roman/NN ./\$.*
- Parser

# Chomsky vs Norvig

Chomsky meint, statistische Modelle haben nur mäßigen Erfolg.

Beispiel (1955):

- a) Colorless green ideas sleep furiously.
- b) Furiously sleep ideas colorless green.

beides gibts nicht,  
aber b ist grammatikalisch falsch

Laut Chomsky müsste Wahrscheinlichkeit für beide Sätze 0 sein.

Heut. stat. Modelle würden sagen, dass a) viel wahrscheinlicher ist als b), aber trotzdem beide sehr unwahrscheinlich.

# Chomsky vs Norvig

Chomsky:

statistische Vorhersagen simulieren etwas in der realen Welt aber fragen nicht nach dem WARUM → keine Wissenschaft

Wenn wir Kepler's Gesetze über das Sonnensystem nicht wüssten, könnten wir die Bewegungen vielleicht auch aus Daten gewinnen (Kevin Gold, 2011)

Norvig: Wissenschaft basiert auf Experimenten & diese erzeugen Daten

# Chomsky vs Norvig

- Norvig: Alle erfolgreichen AIs nutzen statistische Methoden.  
→ Es wird nicht auf Grammatik geachtet & funktioniert besser.
- bei schlechten Ergebnissen  
→ Ansatz von Google = mehr Daten
- Bei natürlichen Sprachen gibt es auch Phrasen, die der Theorie nach falsch sein müssten, aber richtig sind. Solche können mit statistischen Methoden erfasst werden.

# Chomsky vs Norvig

- Ch. meint, ein Kind kann nicht  $10^9$  Parameter (1960) lernen, während einer Kindheit, die nur  $10^8$  Sekunden dauert.
- heutige Modelle noch viel umfangreicher
- Besitzt unser Gehirn tatsächlich eine vorprogrammierte Schablone zum Erlernen von Grammatik? → eher nein
- Sprache im Gehirn noch nicht genügend erforscht.
- Kleinkinder nutzen verschiedene Denkweisen, Intuition spielt eine zentrale Rolle

# Fazit statistischer Sprachmodelle

- Statistische Modelle haben nicht (wie Chomsky behauptet) einen mäßigen Erfolg, sondern eine dominante Position.
- Nach Jahren der Versuche die Sprache mit logischen Modellen zu erfassen, funktionieren statistische Modelle jetzt besser.
- Auch die Wissenschaft beschäftigt sich damit.
- statistische Verfahren werden um strukturelle erweitert, um bessere Ergebnisse zu erzielen

# Semantik von natürlichen und formalen Sprachen

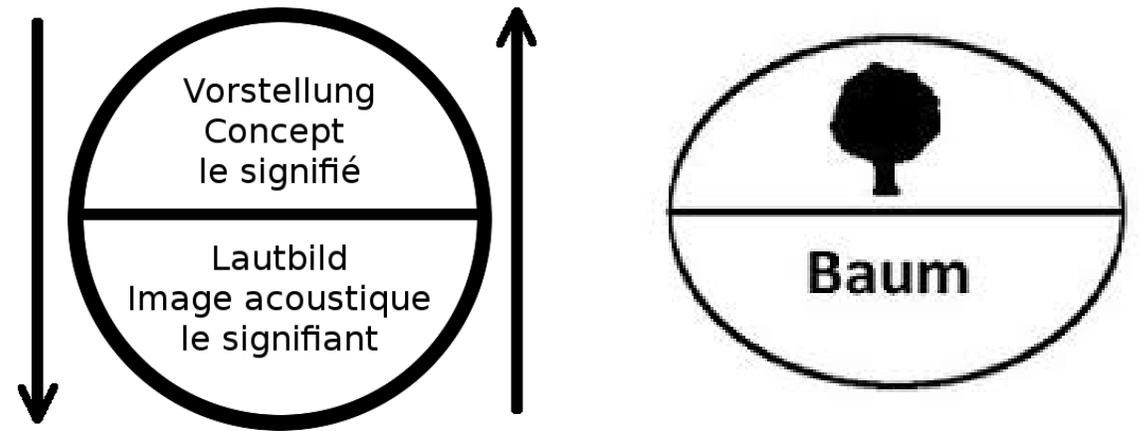
Linguistische Semantik	Logische Semantik
<ul style="list-style-type: none"><li>• Semantische Theorien natürlicher Sprachen</li><li>• Empirische Wissenschaft → Zeichen und Zeichenkombinationen in normalen Sprechakten</li></ul>	<ul style="list-style-type: none"><li>• Semantische Theorien formaler Sprachen</li><li>• Formales Paradigma für jede spezielle Semantik (auch für natürliche Sprachen!)</li></ul>

# Linguistische Semantik – Beispiele

## 1. Semiotik:

### Ferdinand de Saussure:

- Sprachliche Zeichen als bilaterale Einheit von Signifikat (Inhalt) und Signifikant (Ausdruck)



# Linguistische Semantik – Beispiele

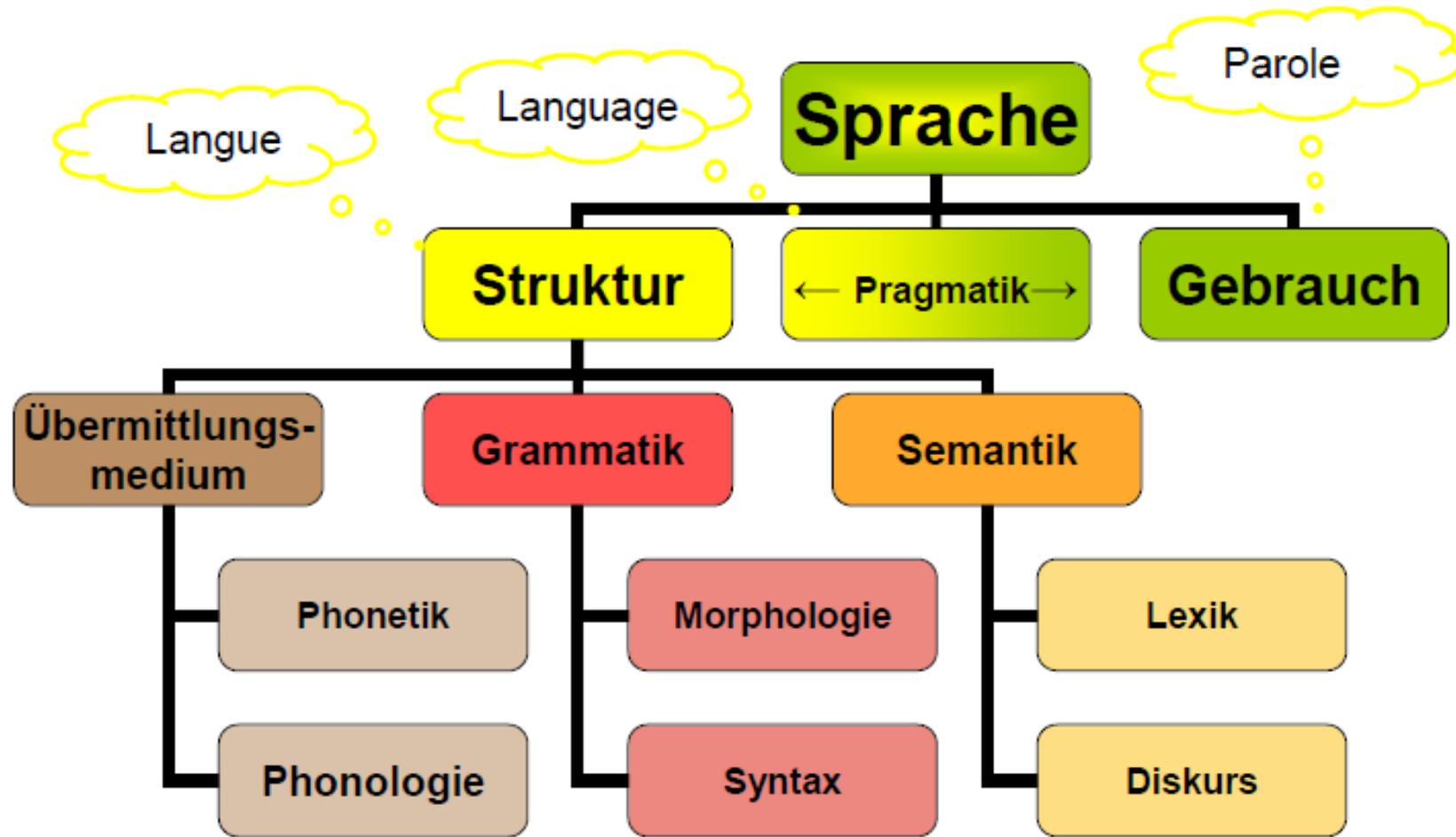
## 2. Pragmatische Ansätze

### Ludwig Wittgenstein: Philosophische Untersuchungen

§43: Man kann für eine *große* Klasse von Fällen der Benützung des Wortes »Bedeutung« - wenn auch nicht für *alle* Fälle seiner Benützung - dieses Wort so erklären: Die Bedeutung eines Wortes ist sein Gebrauch in der Sprache.

§6: Die Kinder werden dazu erzogen, *diese* Tätigkeiten zu verrichten, *diese* Wörter dabei zu gebrauchen, und *so* auf die Worte des Anderen zu reagieren.

# Modell: Langue & Parole



# Übung: Koch und Lehrling

Stellt euch vor, die folgende Szene würde sich nicht zwischen einem Koch und seinem Lehrling, sondern zwischen einem Koch und einem (Lehrlings-)Roboter begeben. Dieser Roboter hätte alle motorischen Fähigkeiten, die für das Kochen nötig sind. Er hätte auch das nötige Wissen über relevante Begriffe (Kochlöffel, Speisen) und Rezepte, nach denen er bereits gekocht hat. Welche nachfolgend dargestellten Handlungen könnten ihm trotzdem so einige Schwierigkeiten bereiten?

# Semantische Eigenschaften natürlicher Sprachen

- Keine Eindeutigkeit von Ausdrücken (Zeichenkombinationen) in natürlichen Sprachen
  - a) Mehrdeutigkeiten: „Flying planes can be dangerous.“
    - **Synonymie**: Aubergine – Melanzani, Tomaten – Paradeiser
    - **Homonymie**: Schloss, Bank
    - **Polysemie**: scharfer Wind, scharfes Messer, scharfe Braut

# Semantische Eigenschaften natürlicher Sprachen

## b) Kontextabhängigkeit:

- **Zeigewörter:** dieses, jenes, ich, du
- Lokale und temporale **Adverbien:** hier, dort, morgen

## c) Vage Ausdrücke:

- **Bedeutung einzelner Wörter:** Berg – Hügel, einige Bücher – viele Bücher
- Verschiedene **Sprachschichten von Ausdrücken:** Berufssprachen, sozialbedingte Gruppensprachen, emotionale Färbungen, verschiedene Funktionen

# Die gemeinsamen Eigenschaften

- Beide Sprachen operieren mit sprachlichen Zeichen und Zeichenkombinationen.
- Formale Sprachen orientieren sich an natürlichen Sprachen. (Formale Sprachen folgen natürlichen Sprachen nach, auch zeitlich gesehen.)
- Wie bei jeder Sprache existieren sowohl für natürliche als auch für formale Sprachen Syntax und Semantik.

Sie teilen also die grundlegenden Eigenschaften, die alle Sprachen teilen!

# Merkmale formaler Sprachen

- In einer formalen Sprache spielt die Präzision eine wesentliche Rolle:
  - Die Kommunikation findet entweder vollständig oder gar nicht statt.
  - Ohne Präzision findet keine Kommunikation statt.
- Es gibt keine Mehrdeutigkeit:
  - Der Aufbau der formalen Sprachen schließt Mehrdeutigkeit aus.
- Keine vielfältigen Kommunikationsziele
- Kommunikation auf einer einfachen Ebene:
  - Die Kommunikation in der natürlichen Sprache kann gezielt mehrdeutig ausgedrückt werden, dies kann die Kommunikation verkomplizieren. Im Gegensatz dazu wird in der formalen Sprache keine Interpretation benötigt.

# Merkmale formaler Sprachen

- Übertragung der emotionalen Beimengungen:
  - Durch den Gebrauch einer künstlichen Sprache wird es möglich von den im jeweiligen Zusammenhang unwesentlichen Bestimmungen und Konnotationen abzusehen.
- Die Abhängigkeit der einzelnen Teile:
  - Ein Fehler – Error – in einer formalen Sprache hält die Interaktion ab.
- Wohlgeformte Ausdrücke:
  - Explizite und eindeutige Formationsregeln müssen für die in einer solchen Sprache erlaubten Ausdrücke gegeben sein. Beispiel: „ $2 + 2 = 4$ “ korrekt, „ $2 + = 4$ “ nicht korrekt.
  - Bei einer künstlichen Sprache deutet man sog. Formationsregel als Regeln, die die SYNTAX der betreffenden Sprache eindeutig bestimmen.

# Merkmale formaler Sprachen

- Denken durch ein algorithmisches bzw. rechnendes Operieren mit Formeln:
  - als Merkmal logischer Kalkül-Sprachen
  - Es ist nicht möglich mit Ausdrücken unserer Umgangssprache rein algorithmische Operationen vorzunehmen.
  - **Vieldeutige Elemente einer natürlichen Sprache dürfen deshalb in einer formalen Sprache nicht auftreten. Es ist eine eindeutige Definition notwendig.**
- **Beispiel:** Kopula „sein“:
  - London ist die Hauptstadt Englands → symbolisch  $x = y$
  - London ist eine Millionenstadt → symbolisch  $x \in A$
  - Der Wal ist ein Säugetier → symbolisch  $A \subset B$

# HUMAN RESOURCE MACHINE



# Literaturverzeichnis

- Brekle, Herbert E. (1991): Semantik. Eine Einführung in die sprachwissenschaftliche Bedeutungslehre. 3. Aufl., unveränd. Nachdr. der 2., verb. Aufl. München: Fink (UTB für Wissenschaft Uni-Taschenbücher Linguistik, 102).
- Carstensen, K.-U., Ebert, C., Ebert, C., Jekat, S., Langer, H., Klabunde, R. (Hrsg.) (2010): Computerlinguistik und Sprachtechnologie: Eine Einführung. Springer Spektrum
- Chomsky, Noam (1972): Die formale Natur der Sprache. In: Lenneberg, Eric H.: Biologische Grundlagen der Sprache. Frankfurt am Main: Suhrkamp.
- Chomsky, Noam (1974): Thesen zur Theorie der generativen Grammatik. Frankfurt am Main: Athenäum-Fischer-Taschenbuch-Verl.
- Edmondson, Willis J., House, Juliane (2006): Einführung in die Sprachlehrforschung. Tübingen: Francke.
- Fischer, Gottfried (1997): Was ist Sprache? In: Ernst, Peter (Hrsg.): Einführung in die synchrone Sprachwissenschaft. Wien: Ed. Praesens.
- Gold, Kevin (2011): Norvig vs. Chomsky and the Fight for the Future of AI. <http://www.tor.com/2011/06/21/norvig-vs-chomsky-and-the-fight-for-the-future-of-ai>
- Ibbotson, Paul, Tomasello, Michael (2017): Ein neues Bild der Sprache. <http://www.spektrum.de/news/kritik-an-der-universalgrammatik-von-chomsky/1439388>
- Norvig, Peter: On Chomsky and two cultures of statistical learning. <http://daselab.cs.wright.edu/nesy/NeSy13/norvig.pdf>
- Norvig, Peter (2012): On Chomsky and the Two Cultures of Statistical Learning. <http://onlinelibrary.wiley.com/doi/10.1111/j.1740-9713.2012.00590.x/epdf>
- Pustka, Elissa (2016): Einführung in die Phonetik und Phonologie des Französischen. Berlin: ESV.
- Stein, Achim (2010): Einführung in die französische Sprachwissenschaft. Weimar: Metzler.
- Wittgenstein, Ludwig; Schulte, Joachim (2014): Tractatus logico-philosophicus. 21. Aufl. Frankfurt am Main: Suhrkamp (Suhrkamp-Taschenbuch Wissenschaft, 501).