

# Grammatik und Stochastik

---

Gerhard Jäger  
*ZAS/Uni Potsdam*

# Der Generativist:

---

„Evidently, one's ability to produce and recognize grammatical utterances is not based on notions of statistical approximation and the like“

*Chomsky, Syntactic Structures*

„It would, incidentally, not be particularly surprising if statistical models turn out to be of little relevance to grammar.“

*Chomsky 1957a*

---

# Der Stochastiker:

---

„Every time I fire a linguist, my performance goes up“

*F. Jelinek, IBM*

---

# Chomskys Argument

---

- (1) Colorless green ideas sleep furiously.
- (2) Furiously sleep ideas green colorless.

„It is fair to assume that neither sentence (1) nor (2) [...] has ever occurred in an English discourse. Yet (1), though nonsensical, is grammatical, while (2) is not. Hence, in any statistical model for grammaticality, these sentences will be ruled out on identical grounds as equally ‚remote‘ from English.“

# Häufigkeit vs. Wahrscheinlichkeit

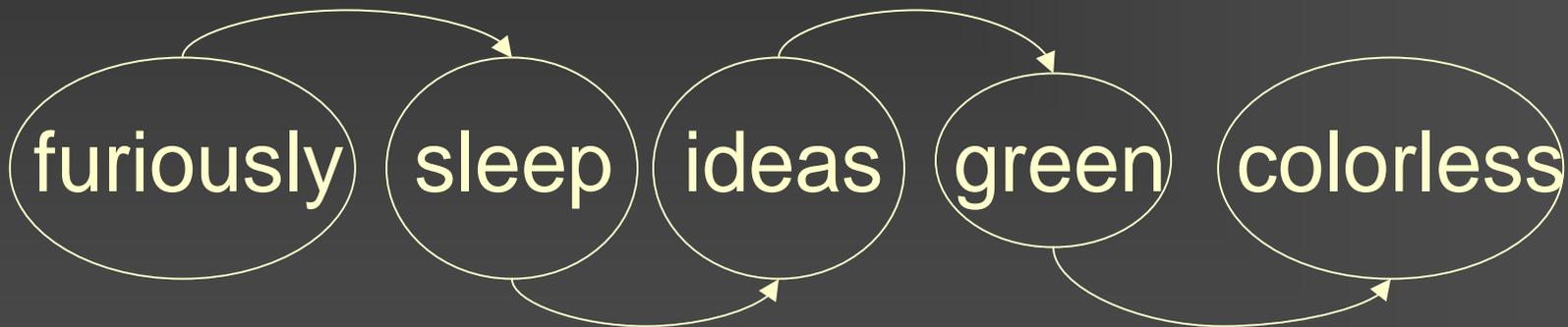
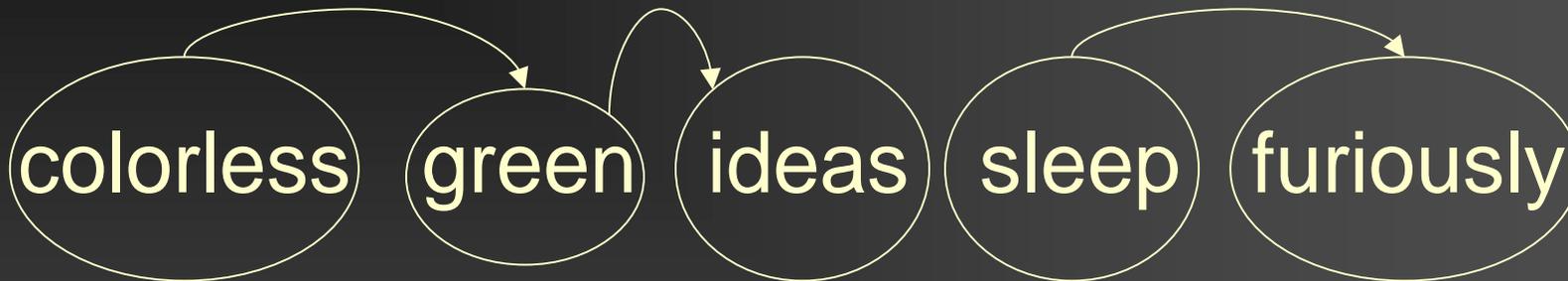
- Weder (1') noch (2') sind (m.W.) je bei „6 aus 49“ gezogen worden:

(1') 49, 48, 47, 46, 45, 44

(2') 49, 49, 49, 49, 49, 49

$$P(1') = 1/13\,983\,816$$

$$P(2') = 0$$



# Lernbarkeit

---

- Wahrscheinlichkeitsabschätzung muss auf induzierten Parametern basieren
- Größenordnung schon für simples Markov-Modell unrealistisch

„We cannot seriously propose that a child learns the values of  $10^9$  parameters in a childhood lasting only  $10^8$  seconds.“

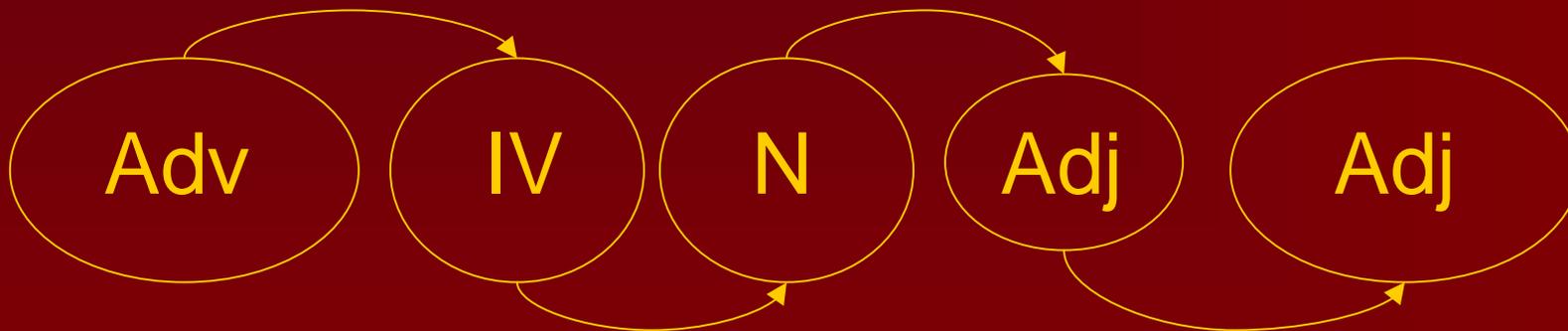
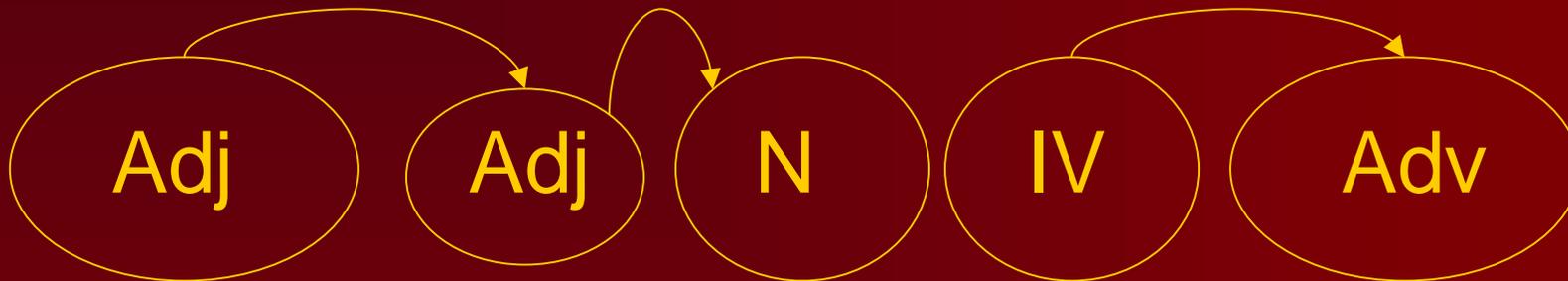
*Chomsky 1963*

---

# Wörter und Kategorien

---

- Versteckte Parameter reduzieren Größenordnung des Lernproblems
  - Einfachste Implementierung:
    - Wörter gehören zu syntaktischen Kategorien
    - Gelernt werden Übergangswahrscheinlichkeiten von Kategorien statt von Wörtern
-



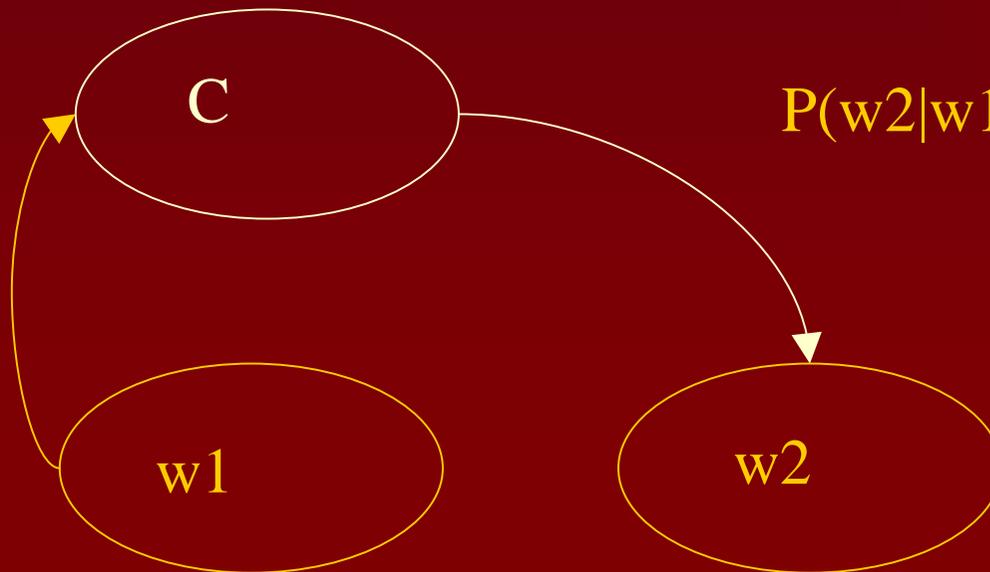
„To say this would be to adopt [...] the viewpoint of ordinary grammatical analysis, in which one considers form classes instead of states [...]. But this is no minor revision; it amounts to a total rejection of the statistical theory of grammaticalness [...].

„In fact, there is little reason to believe that there is any interesting relation between the statistics of language use and the grammatical pattern.“

*Chomsky 1957a*

# Saul & Pereira 1997

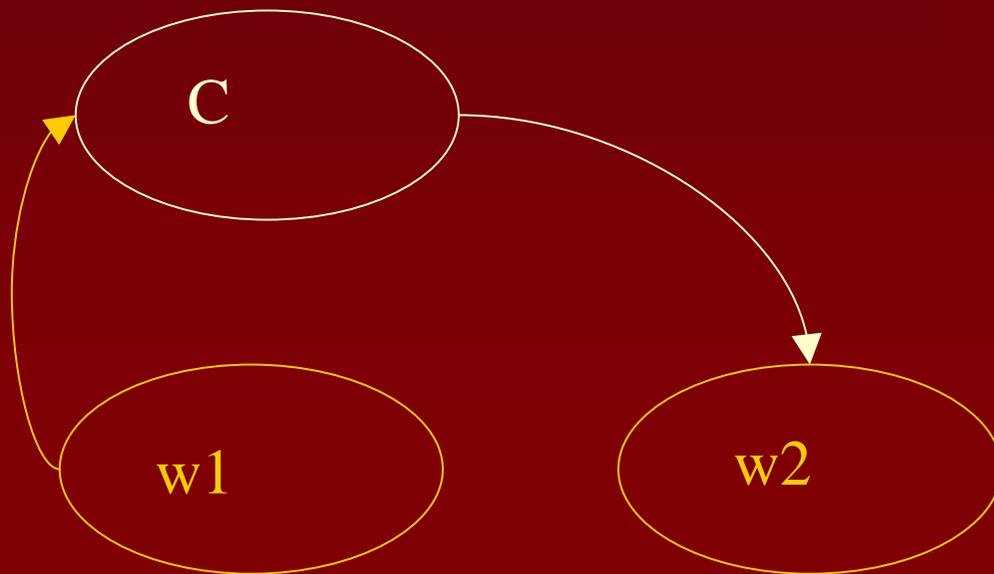
- Aggregate Markov Models:
  - Zuordnung Wort/Kategorie ebenfalls stochastischer Parameter



$$P(w_2|w_1) = \sum_C P(C|w_1) * P(w_2|C)$$

# Saul & Pereira 1997

$$P(w_2|w_1) = \sum_C P(C|w_1) * P(w_2|C)$$



# Zwei Arten der Parameter-Schätzung

---

- Überwachtes Lernen
  - Unüberwachtes Lernen
-

# Überwachtes Lernen

---

- Einfach
  - Kategorien sind bekannt
  - Wahrscheinlichkeit = relative Häufigkeit
  - Setzt linguistisches Vorwissen voraus
-

# Überwachtes Lernen

---

- überwacht gelerntes Modell kann zur Annotation neuer Daten verwendet werden
  - Für jedes Wortpaar: wähle Kategorie aus, die Übergangswahrscheinlichkeit maximiert
-

# Unüberwachtes Lernen

---

- Schwierig
  - Nur Wörter sind bekannt, keine Kategorien
  - Zuordnung *Wort*  $\mapsto$  *Kategorie* muss gelernt werden
  - Lösbar durch schrittweise Approximation
-

# Unüberwachtes Lernen

---

1. Wähle beliebiges Initialmodell
  2. Annotiere Beobachtungsdaten mit Hilfe des provisorischen Modells
  3. Führe überwachtes Lernen durch
  4. Wenn gelerntes Modell besser ist als vorheriges Modell, *goto 2.*
-

# Unüberwachtes Lernen

---

- Im Allg. verbessert jeder Iterationsschritt die Qualität des Modells
- Keinesfalls Verschlechterung
- Maß für Qualität des Modells:

## *Perplexität*

- Grad der Vorhersagbarkeit des folgenden Wortes auf der Basis des vorangehenden Textes
-

# Unüberwachtes Lernen

---

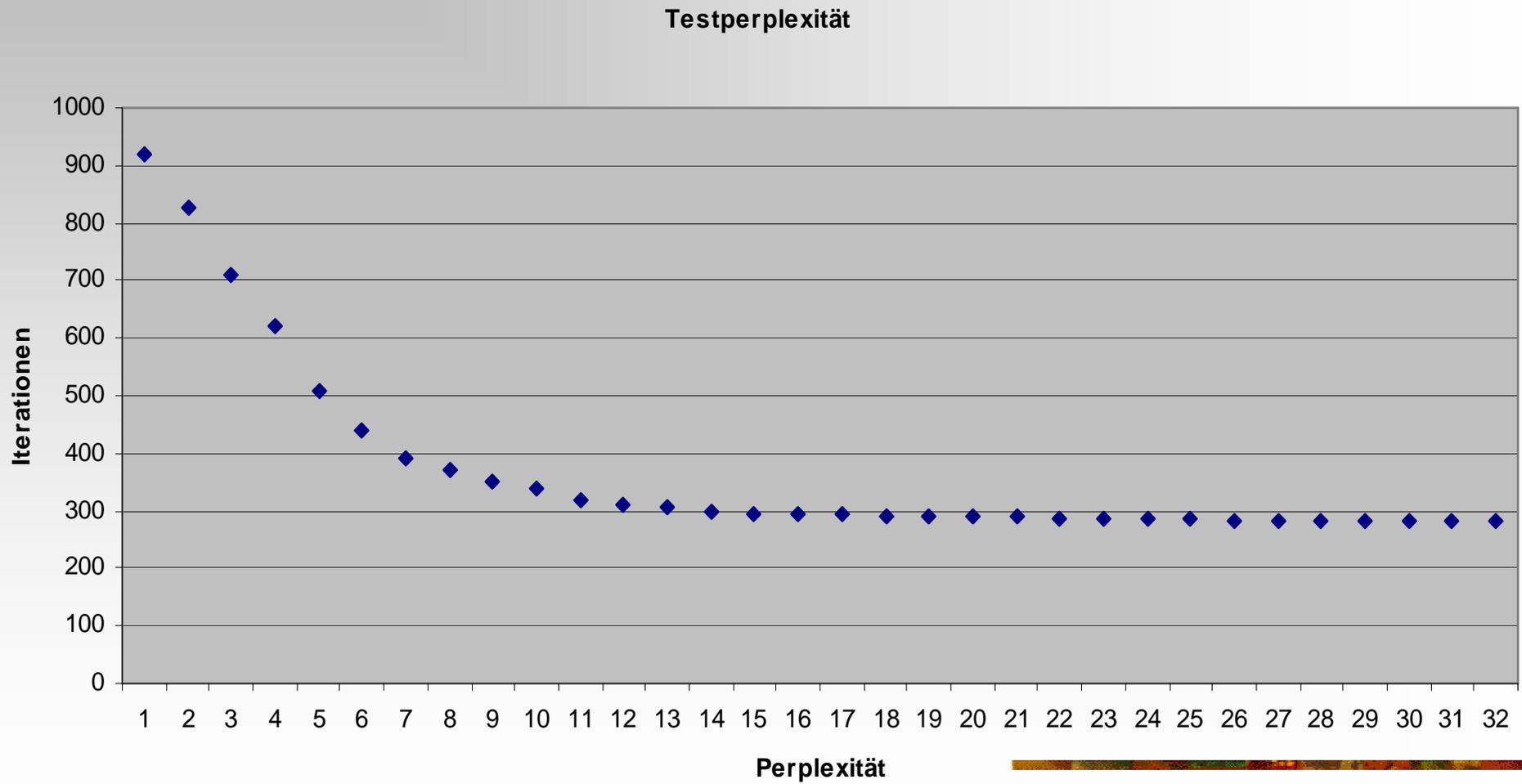
- *A-priori-Setzungen:*
    - *Zahl der Kategorien*
    - *Zufällige Initialwerte für zu lernende Parameter*
  - **Reduziert Zahl der zu lernenden Parameter um mehrere Größenordnungen**
-

# Empirischer Befund

---

- Annahme von 16-32 Kategorien
  - Training anhand von Zeitungstexten
-

# Empirischer Befund



# Empirischer Befund

---

$$\frac{P(\text{Colorless green ideas sleep furiously})}{P(\text{Furiously sleep ideas green colorless})} \approx 2 \cdot 10^5$$

---

# Rekursivität

- Auch Saul & Pereiras Modell ist *finite state*
- Natürliche Sprache erlaubt Zentraleinbettung

(3) If S1, then S2.



# Stochastische Grammatiken

---

- Problem ist unabhängig von Frage „stochastisch versus kategorisch“
  - Kontextfreie Regeln, elementare Bäume, OT-Constraints ... können ebenfalls als versteckte Parameter behandelt werden
  - Implikationen für „logisches Problem des Spracherwerbs“
-

# Poverty of stimulus

---

- Gold 1967:
    - Kontextfreie Grammatiken können nicht allein auf der Basis positiver Evidenz gelernt werden.
-

## ■ Horning 1969:

- Stochastische Kontextfreie Grammatiken können allein auf der Basis positiver Evidenz gelernt werden
- Stochastische Grammatikmodelle kommen mit weniger restriktiver UG aus

# Zusammenfassung

---

## Chomskys Argumente gegen stochastische Grammatikmodelle

- Stochastische Modelle arbeiten nur mit beobachtbaren Parametern
    - Von Entwicklung überholt
    - Versteckte Variable können induktiv gelernt werden
-

- 
- Stochastische Modelle enthalten zu viele zu lernende Parameter, um psychologisch relevant zu sein
    - Abhängig von konkreter Grammatikarchitektur
    - Praktisch nach wie vor ernstes Problem, aber kein grundsätzlich gültiger Einwand
-

---

- Stochastische Modelle sind zwangsläufig  
*finite state*

- Sachlich falsch