

Automatentheorie und formale Sprachen

Wiebke Petersen

Seminarplan

- Lektüre: Theoretische Grundlagen der Informatik, Rolf Socher, 2008 (3. Auflage), Hanser-Verlag
- weitere hilfreiche Literatur:
 - B. Partee et al.: Mathematical Methods in Linguistics, Dordrecht: Kluwer Acad. Publ., 1990.
 - R. Klabunde: Formale Grundlagen der Linguistik. Tübingen: Gunter Narr Verlag, 1998.
 - U. Schöning (5. Aufl., 2008): Theoretische Informatik - kurzgefasst. Spektrum.

Inhalt:

- Chomskyhierarchie als Komplexitätshierarchie formaler Sprachen
- Grammatiken und Automaten
- formale Sprachen und ihre Abschlußeigenschaften
- Berechenbarkeit und Entscheidbarkeit
- Komplexitätsklassen

Formale Komplexität natürlicher Sprachen

- Deutsch, Englisch, Chinesisch, Finnisch, ...

Formale Komplexität natürlicher Sprachen

- Deutsch, Englisch, Chinesisch, Finnisch, ...
- Prolog, Pascal, ...

- Deutsch, Englisch, Chinesisch, Finnisch, ...
- Prolog, Pascal, ...
- Esperanto, Volapük, Interlingua, ...

Formale Komplexität natürlicher Sprachen

- Deutsch, Englisch, Chinesisch, Finnisch, ...
- Prolog, Pascal, ...
- Esperanto, Volapük, Interlingua, ...
- Aussagenlogik, Prädikatenlogik, ...
- ...

Formale Komplexität **natürlicher Sprachen**

- Deutsch, Englisch, Chinesisch, Finnisch, ...

Formale Komplexität natürlicher Sprachen

- Deutsch, Englisch, Chinesisch, Finnisch, ...
- vage, ambig,

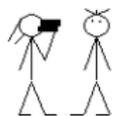
Formale Komplexität **natürlicher Sprachen**

- Deutsch, Englisch, Chinesisch, Finnisch, ...
- vage, ambig,
- Ambiguitäten
 - lexikalische Ambiguitäten (Ruf morgen an - Der Ruf der Möwen)

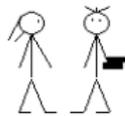
Formale Komplexität natürlicher Sprachen

- Deutsch, Englisch, Chinesisch, Finnisch, ...
- vage, ambig,
- Ambiguitäten
- lexikalische Ambiguitäten (Ruf morgen an - Der Ruf der Möwen)
- strukturelle Ambiguitäten:

- Die Frau sieht \lceil den Mann \rceil mit dem Fernrohr



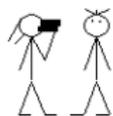
- Die Frau sieht \lceil den Mann mit dem Fernrohr \rceil



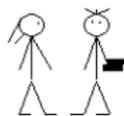
Formale Komplexität natürlicher Sprachen

- Deutsch, Englisch, Chinesisch, Finnisch, ...
- vage, ambig,
- Ambiguitäten
- lexikalische Ambiguitäten (Ruf morgen an - Der Ruf der Möwen)
- strukturelle Ambiguitäten:

- Die Frau sieht \lceil den Mann \rceil mit dem Fernrohr



- Die Frau sieht \lceil den Mann mit dem Fernrohr \rceil

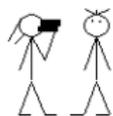


- einzige Experten: Menschen
- nur endlich viele Sätze abfragbar
- nicht immer konsistente Antworten

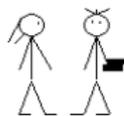
Formale Komplexität natürlicher Sprachen

- Deutsch, Englisch, Chinesisch, Finnisch, ...
- vage, ambig,
- Ambiguitäten
- lexikalische Ambiguitäten (Ruf morgen an - Der Ruf der Möwen)
- strukturelle Ambiguitäten:

- Die Frau sieht \lceil den Mann \rceil mit dem Fernrohr



- Die Frau sieht \lceil den Mann mit dem Fernrohr \rceil



- einzige Experten: Menschen
- nur endlich viele Sätze abfragbar
- nicht immer konsistente Antworten

- schwierig zu erlernen im Erstspracherwerb / Zweitspracherwerb

Formale **Komplexität natürlicher Sprachen**

- schwierig zu erlernen im Erstspracherwerb / Zweitspracherwerb
- komplexe Phonologie / Morphologie / Syntax / ...

Formale **Komplexität natürlicher Sprachen**

- schwierig zu erlernen im Erstspracherwerb / Zweitspracherwerb
- komplexe Phonologie / Morphologie / Syntax / ...
- schwierig maschinell zu parsen

Formale Komplexität natürlicher Sprachen

- Komplexität der Berechnung / Verarbeitungskomplexität (computational complexity)

Formale Komplexität natürlicher Sprachen

- Komplexität der Berechnung / Verarbeitungskomplexität (computational complexity)
- Komplexität der Struktur

Formale Komplexität natürlicher Sprachen

- Komplexität der Berechnung / Verarbeitungskomplexität (computational complexity)
- Komplexität der Struktur
 - Natürliche Sprachen werden als abstrakte Symbolsysteme betrachtet, bestehend aus elementaren Zeichen und Kombinationsvorschriften.
 - Fragen nach der Grammatikalität natürlichsprachlicher Sätze entsprechen Fragen nach der syntaktischen Korrektheit von Programmen oder der Wohlgeformtheit logischer Ausdrücke.

Warum ist die formale Komplexität natürlicher Sprachen von Interesse?

- gibt Information über die Struktur von natürlichen Sprachen (NL)

Warum ist die formale Komplexität natürlicher Sprachen von Interesse?

- gibt Information über die Struktur von natürlichen Sprachen (NL)
- erlaubt Rückschlüsse auf Adäquatheit eines Grammatikformalismus für NL

Warum ist die formale Komplexität natürlicher Sprachen von Interesse?

- gibt Information über die Struktur von natürlichen Sprachen (NL)
- erlaubt Rückschlüsse auf Adäquatheit eines Grammatikformalismus für NL
- unter CL Aspekten sind möglichst effizient verarbeitbare Analysen gefragt

Warum ist die formale Komplexität natürlicher Sprachen von Interesse?

- gibt Information über die Struktur von natürlichen Sprachen (NL)
- erlaubt Rückschlüsse auf Adäquatheit eines Grammatikformalismus für NL
- unter CL Aspekten sind möglichst effizient verarbeitbare Analysen gefragt
- erlaubt Rückschlüsse auf menschliche Sprachverarbeitung

Was eine Grammatik erklären muß

- 1 Anne sieht Peter
- 2 Anne Peter sieht
- 3 sieht Peter Anne
- 4 sieht Anne Peter
- 5 Peter sieht Anne
- 6 Peter Anne sieht

weitere Sprachdaten

- 1 Anne sieht Peter am Rathaus mit dem Fernrohr
- 2 Anne sieht Peter mit dem Fernrohr am Rathaus
- 3 am Rathaus sieht Anne Peter mit dem Fernrohr
- 4 mit dem Fernrohr sieht Anne Peter am Rathaus
- 5 mit dem Fernrohr am Rathaus sieht Anne Peter
- 6 ...

Auch wenn das Deutsche eine relativ freie Wortstellung hat, so wird doch deutlich, daß die Zahl der grammatisch korrekten Sätze verschwindend klein ist im Vergleich zu den ungrammatischen Wortketten.

Wie komplex sind denn nun Sätze des Deutschen?

- 1 Anne sieht Peter
- 2 Anne sieht Peter am Rathaus mit dem Fernrohr
- 3 Anne sieht Peter, den sie vorgestern kennengelernt hat, am Rathaus mit dem Fernrohr

Wie komplex sind denn nun Sätze des Deutschen?

- 1 Anne sieht Peter
- 2 Anne sieht Peter am Rathaus mit dem Fernrohr
- 3 Anne sieht Peter, den sie vorgestern kennengelernt hat, am Rathaus mit dem Fernrohr
- 4 Anne sieht Peter und Hans und Sabine und Joachim und Elfriede und Johanna und Maria und Jochen und Thomas und Andrea

Wie komplex sind denn nun Sätze des Deutschen?

- 1 Anne sieht Peter
- 2 Anne sieht Peter am Rathaus mit dem Fernrohr
- 3 Anne sieht Peter, den sie vorgestern kennengelernt hat, am Rathaus mit dem Fernrohr
- 4 Anne sieht Peter und Hans und Sabine und Joachim und Elfriede und Johanna und Maria und Jochen und Thomas und Andrea

Satzlänge spielt zwar bei der Verarbeitungskomplexität eine Rolle, ist aber kein strukturelles Komplexitätsmerkmal!

Klassische Grammatiken versus Theorie formaler Sprachen

Klassische Grammatiken

- sollen sprachliche Daten erklären
- sind einzelsprachspezifisch

Klassische Grammatiken versus Theorie formaler Sprachen

Klassische Grammatiken

- sollen sprachliche Daten erklären
- sind einzelsprachspezifisch

Theorie formaler Sprachen

- ist eine Theorie über den Aufbau und die Struktur von Symbolkettenmengen
- ist grundsätzlicher und allgemeiner ausgelegt als jede Grammatiktheorie
- erlaubt Aussagen über die Mechanismen der Erzeugung und Erkennung von Symbolkettenmengen

- Formale Sprachen sind Mengen von Wörtern (entspricht in natürlichen Sprachen den Sätzen), die ihrerseits aus Zeichen (in natürlichen Sprachen Wörter) aufgebaut sind. Was in der Menge ist, ist ein “grammatisch korrektes Wort”, alles andere nicht.

Formale Sprachen

- Formale Sprachen sind Mengen von Wörtern (entspricht in natürlichen Sprachen den Sätzen), die ihrerseits aus Zeichen (in natürlichen Sprachen Wörter) aufgebaut sind. Was in der Menge ist, ist ein “grammatisch korrektes Wort”, alles andere nicht.
- Für “sinnvolle”/“strukturierte” formale Sprachen lassen sich endliche Mengen von Regeln/Grammatiken angeben, die diese beschreiben.

Formale Sprachen

- Formale Sprachen sind Mengen von Wörtern (entspricht in natürlichen Sprachen den Sätzen), die ihrerseits aus Zeichen (in natürlichen Sprachen Wörter) aufgebaut sind. Was in der Menge ist, ist ein “grammatisch korrektes Wort”, alles andere nicht.
- Für “sinnvolle”/“strukturierte” formale Sprachen lassen sich endliche Mengen von Regeln/Grammatiken angeben, die diese beschreiben.
- Wir gehen in der Regel davon aus, daß alle natürlichen Sprachen durch endlich viele Regeln beschreibbar sind, da wir sie ansonsten nicht sprechen/verstehen könnten.

Chomsky-Hierarchie

- Wenn man die Form der Regeln einschränkt erhält man Teilmengen der Menge aller durch eine Grammatik erzeugten Sprachen.

Chomsky-Hierarchie

- Wenn man die Form der Regeln einschränkt erhält man Teilmengen der Menge aller durch eine Grammatik erzeugten Sprachen.
- Die Chomsky-Hierarchie ist eine Hierarchie über die Regelbedingungen (den verschiedenen Sprachklassen entsprechen Einschränkungen über die rechten und linken Regelseiten).

Chomsky-Hierarchie

- Wenn man die Form der Regeln einschränkt erhält man Teilmengen der Menge aller durch eine Grammatik erzeugten Sprachen.
- Die Chomsky-Hierarchie ist eine Hierarchie über die Regelbedingungen (den verschiedenen Sprachklassen entsprechen Einschränkungen über die rechten und linken Regelseiten).
- Die Chomsky Hierarchie reflektiert eine spezielle Form der Komplexität, andere Kriterien sind denkbar und führen zu anderen Hierarchien.

Chomsky-Hierarchie

- Wenn man die Form der Regeln einschränkt erhält man Teilmengen der Menge aller durch eine Grammatik erzeugten Sprachen.
- Die Chomsky-Hierarchie ist eine Hierarchie über die Regelbedingungen (den verschiedenen Sprachklassen entsprechen Einschränkungen über die rechten und linken Regelseiten).
- Die Chomsky Hierarchie reflektiert eine spezielle Form der Komplexität, andere Kriterien sind denkbar und führen zu anderen Hierarchien.
- Die Sprachklassen der Chomsky Hierarchie sind in der Informatik intensiv untersucht worden (Berechnungskomplexität, effektive Parser).

Chomsky-Hierarchie

- Wenn man die Form der Regeln einschränkt erhält man Teilmengen der Menge aller durch eine Grammatik erzeugten Sprachen.
- Die Chomsky-Hierarchie ist eine Hierarchie über die Regelbedingungen (den verschiedenen Sprachklassen entsprechen Einschränkungen über die rechten und linken Regelseiten).
- Die Chomsky Hierarchie reflektiert eine spezielle Form der Komplexität, andere Kriterien sind denkbar und führen zu anderen Hierarchien.
- Die Sprachklassen der Chomsky Hierarchie sind in der Informatik intensiv untersucht worden (Berechnungskomplexität, effektive Parser).
- Für Linguisten ist die Chomsky Hierarchie besonders interessant, da sie die Form der Regeln zentral stellt, und somit Aussagen über Grammatikformalisten zuläßt.



Noam Chomsky

(* 7.12.1928, Philadelphia)

Noam Chomsky, *Three Models for the Description of Language*,
IRE Transactions on Information Theory (1956).

Chomsky-Hierarchie (grober Überblick)

reguläre Sprachen (regular languages)	Typ 3, REG	$A \rightarrow bA$ $A \rightarrow a$	$a^* b^*$
kontextfreie Sprachen (context-free languages)	Typ 2, CF	$A \rightarrow \beta$	$a^n b^n, w^R w$
kontextsensitive Sprachen context-sensitive languages	Typ 1, CS	$\alpha A \nu \rightarrow \alpha \beta \nu$	$a^n b^n c^n, ww,$ $a^n b^m c^n d^m$
allgemeine Regelsprachen recursively enumerable languages	Typ 0, RE	$\alpha \rightarrow \beta$	

$a \in T, A \in N, \alpha, \beta, \dots \in (N \cup T)^*, S$ Startsymbol