

Automatentheorie und Formale Sprachen

Einleitung

Dozentin: Wiebke Petersen

22.4.2009

Seminarplan

- Grundlagen: formalen Sprachen, Grammatiken, Maschinen (siehe Partee, ter Meulen und Wall 1990 oder Klabunde 1998 (Vorsicht Fehler!))
 - reguläre Sprachen und ihre Abschlußeigenschaften, reguläre Ausdrücke und endliche Automaten, Pumpinglemma für reguläre Sprachen, Transduktoren
 - kontextfreie Sprachen und ihre Abschlußeigenschaften, Phrasenstrukturregeln, (Kellerautomaten), Pumpinglemma für kontextfreie Sprachen, Chomsky-Normalform
 - Chomsky-Hierarchie (kontextsensitive Sprachen, rekursiv aufzählbare Sprachen)
 - formale Komplexität natürlicher Sprachen

Seminarplan

- Grundlagen: formalen Sprachen, Grammatiken, Maschinen (siehe Partee, ter Meulen und Wall 1990 oder Klabunde 1998 (Vorsicht Fehler!))
 - reguläre Sprachen und ihre Abschlußeigenschaften, reguläre Ausdrücke und endliche Automaten, Pumpinglemma für reguläre Sprachen, Transduktoren
 - kontextfreie Sprachen und ihre Abschlußeigenschaften, Phrasenstrukturregeln, (Kellerautomaten), Pumpinglemma für kontextfreie Sprachen, Chomsky-Normalform
 - Chomsky-Hierarchie (kontextsensitive Sprachen, rekursiv aufzählbare Sprachen)
 - formale Komplexität natürlicher Sprachen
- Warum ist das Schweizerdeutsche eine besonders komplexe Sprache?

Formale Komplexität natürlicher Sprachen

Formale Komplexität natürlicher Sprachen

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

Formale Komplexität natürlicher Sprachen

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

Formale Komplexität natürlicher Sprachen

- 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,

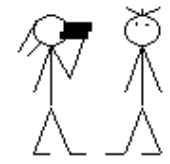
Formale Komplexität **natürlicher Sprachen**

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

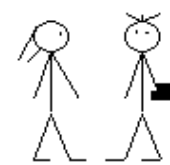
Formale Komplexität **natürlicher Sprachen**

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

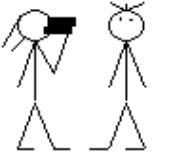
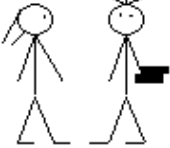
○ Die Frau sieht \lceil den Mann \rceil mit dem Fernrohr



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

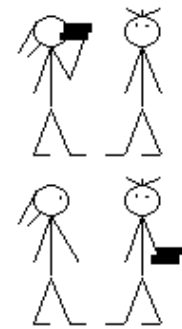


Formale Komplexität **natürlicher Sprachen**

- 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
- Kunstsprachen werden starr definiert – natürliche Sprachen entwickeln sich (fließende Sprachverschiebungen)



- die Entwicklung natürlicher Sprachen hängt ab von klimatischen, politischen, gesellschaftlichen, . . . Bedingungen.
- für Kunstsprachen gilt das Prinzip der semantischen Kompositivität; bei natürlichen Sprachen ist dies nicht so offensichtlich (wissenschaftlicher Streitpunkt).

Formale **Komplexität natürlicher Sprachen**

- 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

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.

Fragen?