

Grammatikformalismen und ihre Parskomplexität

2. Sitzung: Mild kontextsensitive Grammatiken

Christoph Teichmann

Institut für Linguistik, Universität Potsdam
Grammatikformalismen und ihre Parskomplexität,
Thomas Hanneforth

29. April 2009

Christoph Teichmann

2: Mild kontextsensitive Grammatiken

29.04.2009 1 / 15

Dependenz

Warum ist Dependenz wichtig?

- ▶ Gibt Subkategorisierung wieder
- ▶ Stellt eine Abhängigkeit von Elementen über lange Distanzen her
- ▶ Zentral für die Analyse von Fragesätzen u.ä. Konstrukten, die häufig durch Bewegung analysiert werden
- ▶ Einfache Darstellung von Skopusbeziehungen

Christoph Teichmann

2: Mild kontextsensitive Grammatiken

29.04.2009 2 / 15

Dependenz bei kontextfreien Grammatiken

Problem:

Die Standardinterpretation für die von kontextfreien Grammatiken erzeugten Bäume generiert nur projektive Dependenz, aber natürliche Sprachen scheinen komplexere Dependenz zu haben.

- ▶ Von *dieser Frau* behauptet Tom, Hans hätte gesagt, *sie hätte* ihn gerettet.
- ▶ Von *dir* behauptet Tom, Hans hätte gesagt, *du hättest* ihn gerettet.

Christoph Teichmann

2: Mild kontextsensitive Grammatiken

29.04.2009 3 / 15

Polynomielle Zeit

Grammatikformalismen, die verwendet werden, um menschliche Sprache zu modellieren, sollten Parsingalgorithmen in PTIME haben.

- ▶ Gut für Anwendungen
- ▶ sinnvolle Begrenzung für die vom Menschen erwartete Rechenleistung
- ▶ Phänomene, die nicht kontext-frei sind, sind immer noch in PTIME

Christoph Teichmann

2: Mild kontextsensitive Grammatiken

29.04.2009 4 / 15

Kontextfreie Grammatiken

Man sollte die Möglichkeiten kontextfreier Sprachen für alle Versuche, einen Grammatikformalismus für natürliche Sprache zu finden, bewahren.

- ▶ es gibt viele Phänomene, die durch kontextfreie Grammatiken gut beschrieben werden
- ▶ Bäume sind eine gute Basis für weitere Verarbeitung und Beschreibung von Sprachen

Christoph Teichmann

2: Mild kontextsensitive Grammatiken

29.04.2009 5 / 15

Aufbau aus Abschnitte

Sätze natürlicher Sprache sollten aus wiederkehrenden Abschnitten bestehen. Daher wird erwartet, dass die Sätze sehr gleichmäßig wachsen.

Christoph Teichmann

2: Mild kontextsensitive Grammatiken

29.04.2009 6 / 15

Mild kontextsensitive Grammatiken

Mit diesen Einsichten formulierte Joshi folgende Eigenschaften mild kontextsensitiver Grammatiken:

1. Die kontextfreien Sprachen sind eine Teilmenge der mild kontextsensitiven Sprachen
2. Sie erlauben projektive Dependenz und begrenzte kreuzende Dependenz
3. Ihr Erkennungsproblem liegt in PTIME
4. Alle von ihnen generierten Sprachen wachsen konstant

Christoph Teichmann

2: Mild kontextsensitive Grammatiken

29.04.2009 7 / 15

Teilmenge

1. sollte kaum Fragen aufwerfen.

Christoph Teichmann

2: Mild kontextsensitive Grammatiken

29.04.2009 8 / 15

Dependenz

2. ist ein wenig unscharf. Es sollte eine Obergrenze für die kreuzenden Dependenz pro Wort für eine Grammatik gegeben sein.

Christoph Teichmann

2: Mild kontextsensitive Grammatiken

29.04.2009 9 / 15

Dependenzen

Obwohl 2. keine Bedingung ist, die unabhängig von Grammatiken wirklich formuliert werden kann wird gerne als Beispiel angegeben, dass für jedes n :

$$\{w^n : w \in \Sigma^*\}$$

eine mild kontext-sensitive Sprache ist.

Dass aber:

$$\{w^m : w \in \Sigma^* \wedge m \in \mathbb{N}\}$$

keine mild kontext-sensitive Sprache ist.

Polynomielle Zeit

versteht sich von selbst

Konstantes Wachstum

Wenn man alle Wörter einer mild kontext-sensitiven Sprache nach Länge sortiert, dann gibt es eine Konstante k , so dass kein Wort mehr als k Zeichen kürzer ist als sein Nachfolger.

Generell

- ▶ Es gibt viele Formalismen, die Mengen von Sprachen generieren, die die vier Eigenschaften haben.
- ▶ Die Unschärfe der Eigenschaften macht unklar, wann genau eine Menge von Sprachen nicht mehr mild kontext-sensitiv ist.

Beispiele

- ▶ Tree Adjoining Grammars (TAG)
- ▶ Head Grammars (HG)
- ▶ Linear indexed Grammars (LIG)
- ▶ Combinatory categorial Grammars (CCG)
- ▶ Linear context-free rewriting Systems (LCFRS)

Beziehungen

Bei schwacher generativer Kapazität gilt (Joshi et al. 1991):

$$\text{CFG} \subset \text{TAG} = \text{HG} = \text{LIG} = \text{CCG} \subset \text{LCFRS}$$

das können wir bei Bedarf auch alles noch diskutieren.