

Sprachenhierarchie

Übersicht

- ◆ Begriffe
- ◆ Hierarchie verschiedener Formalismen (unvollständig)
 - ◆ Aufzählung — Endliche Sprachen
 - ◆ Endliche Automaten — Reguläre Sprachen
 - ◆ Kontextfreie Grammatiken — Kontextfreie Sprachen
- ◆ Wo liegen die natürlichen Sprachen?
 - ◆ Zürichdeutsch ist nicht kontextfrei
- ◆ Literatur

Zweck

- ◆ Interesse wecken; *kein Prüfungsstoff*

Begriffe

Sprache einer kontextfreien Grammatik

- ◆ Die Menge aller vom Startsymbol einer kontextfreien Grammatik G ableitbaren Ketten von Terminalsymbolen heisst *Sprache von G* , kurz $L(G)$.

Kontextfreie Sprache

- ◆ Eine kontextfreie Sprache ist eine Sprache, welche durch eine kontextfreie Grammatik beschrieben werden kann.

Reguläre Sprache

- ◆ Eine reguläre Sprache ist eine Sprache, die ein Endlicher Automat erkennen kann.

Kontextfreie \supset Reguläre Sprachen

Kontextfreie vs. reguläre Sprachen

- ◆ Jede reguläre Sprache ist auch kontextfrei
 - ◆ zu jedem Endlichen Automaten lässt sich eine gleichwertige kontextfreie Grammatik konstruieren
 - ▶ Verfahren (konstruktiver Beweis): [Lewis/Papadimitriou, 1981], Satz 3.2.1, p. 103
- ◆ Nicht jede kontextfreie Sprache ist regulär
 - ◆ die Sprache $a^n b^n$ kann nicht von einem Endlichen Automaten erkannt werden
- ◆ Somit sind die kontextfreien Sprachen eine echte Übermenge der regulären Sprachen

Sprachenhierarchie (unvollständig)

Kontextfreie Sprachen

Reguläre Sprachen

Endliche Sprachen

Wo liegen die natürlichen Sprachen?

Sind natürliche Sprachen endlich?

Endliche Sprachen: Aufzählung aller korrekten Sätze

Vorteil:

- ◆ Einfachster nur denkbarer Formalismus

Nachteil:

- ◆ Eine natürliche Sprache besitzt beliebig viele korrekte Sätze
⇒ natürliche Sprachen sind nicht endlich

Der Tisch ist gross

Der Tisch ist gross und schön

Der Tisch ist gross und schön und dunkel

Der Tisch ist gross und schön und dunkel und vierbeinig

Der Tisch ist gross und schön und dunkel und vierbeinig und ...

Sind natürliche Sprachen regulär?

Endliche Automaten als Grammatikformalismus?

Vorteil:

- ◆ sehr effizient zu verarbeiten
 - ◆ Die Verarbeitungszeit wächst linear mit der Länge der Eingabe

Nachteil:

- ◆ Sprachen vom Typ $u x^n v y^n w$ können nicht von Endlichen Automaten erkannt werden, wenn n beliebig gross sein darf
 - ◆ »Pumping-Lemma für Endliche Automaten«
- ◆ Wenn natürliche Sprachen solche sog. *Center-Embedding*-Konstruktionen besitzen, sind Endliche Automaten zu schwach

Sind natürliche Sprachen regulär?

Endliche Automaten als Grammatikformalismus?

- ◆ Das Englische besitzt Strukturen der Art $u x^n v y^n w$:

A white male hired another white male

A white male whom a white male hired hired another white male

*A white male whom a white male whom a white male hired hired
hired another white male*

*A white male whom a white male whom a white male whom a white male hired
hired hired hired another white male*

- ◆ u = "A white male"
- ◆ x = "whom a white male"
- ◆ v = ""
- ◆ y = "hired"
- ◆ w = "hired another white male"

Sind natürliche Sprachen kontextfrei?

Kontextfreie Grammatiken als Grammatikformalismus?

Vorteil:

- ◆ nur bei wenigen, recht exotischen Sprachen kann für exotische Phänomene gezeigt werden, dass sie nicht kontextfrei sind
- ◆ einfacher Parser in Prolog eingebaut

Nachteile:

- ◆ es gibt eben doch nicht-kontextfreie Phänomene
- ◆ beim besten bekannten Parsingalgorithmus steigt die Verarbeitungszeit kubisch mit der Eingabelänge
- ◆ mühsam für Sprachen mit freierer Wortstellung als Englisch

Zürichdeutsch ist nicht kontextfrei

Kontextfreie Sprachen sind unter Schnitt mit Regulären Sprachen abgeschlossen

- ◆ Wenn man die Menge der Sätze einer kontextfreien Sprache mit der Menge der Sätze schneidet, die in einer beliebigen regulären Sprache enthalten sind, ist das Ergebnis wiederum eine kontextfreie Sprache
- ◆ Umkehrschluss: Schneidet man irgendeine Sprache L mit einer regulären, und das Ergebnis ist *nicht* kontextfrei, dann kann auch L nicht kontextfrei sein

Zürichdeutsch ist nicht kontextfrei

Zürichdeutsch

- ◆ markiert Dativ- und Akkusativ-Objekte durch unterschiedliche Wortformen (wie Standard-Deutsch)
- ◆ manche Verben benötigen Dativ-, manche Akkusativobjekt (wie Standard-Deutsch)
- ◆ die Wortstellung ist freier als im Standard-Deutschen, auch sogenannte *Cross-Serial Dependencies* sind erlaubt (wie Niederländisch)

Zürichdeutsch ist nicht kontextfrei

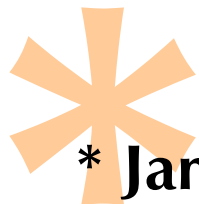
Die meisten Sprachen (z.B. Standard-Deutsch):

- ◆ Verschachtelte Konstruktionen sind möglich

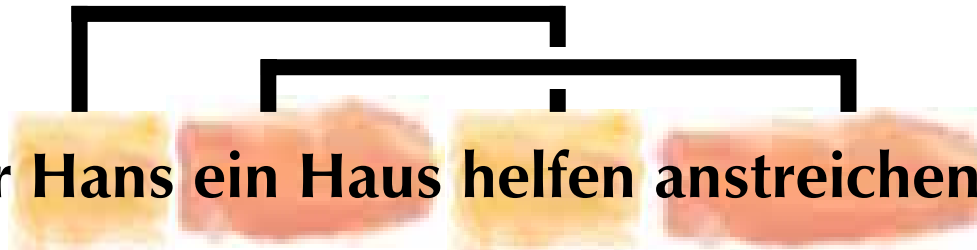
Jan sagt, daß wir Hans ein Haus anstreichen helfen.



- ◆ *Cross-Serial-Dependency*-Konstruktionen sind nicht zulässig



* Jan sagt, daß wir Hans ein Haus helfen anstreichen.



Zürichdeutsch ist nicht kontextfrei

Zürichdeutsch:

- ◆ Verschachtelte Konstruktionen sind möglich

De Jan säit, dass mer em Hans es Huus aastriiche hälfed.

- ◆ *Cross-Serial-Dependency*-Konstruktionen sind ebenfalls erlaubt

De Jan säit, dass mer em Hans es Huus hälfed aastriiche.

Zürichdeutsch ist nicht kontextfrei

Falsche Kasusmarkierung macht eine zürichdeutsche Cross-Serial-Dependency-Konstruktion ungrammatisch.

De Jan säit, dass mer em Hans es Huus hälfed aastriiche.

Dat. Akk. Dat.-V Akk.-V



* De Jan säit, dass mer em Hans em Huus hälfed aastriiche.

Dat. Dat. Dat.-V Akk.-V

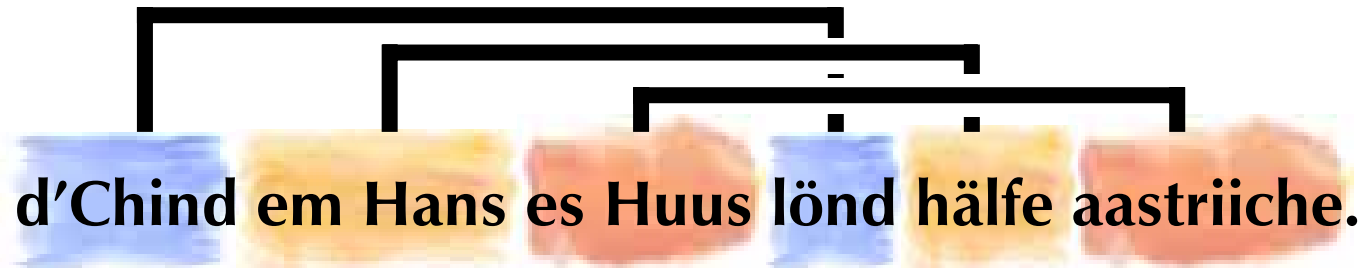
Zürichdeutsch ist nicht kontextfrei

Die zürichdeutschen Cross-Serial-Dependency-Konstruktionen können »aufgeblasen« werden.

De Jan säit, dass mer em Hans es Huus hälfed aastriiche.



De Jan säit, dass mer d'Chind em Hans es Huus lönd hälfe aastriiche.



Zürichdeutsch ist nicht kontextfrei

Wird Zürichdeutsch mit der regulären Sprache

De Jan säit, dass mer Akk.-Obj.* Dat.-Obj.*
es Huus händ wele Akk.-V* Dat.-V* aastriiche.

geschnitten, ergibt sich die Sprache

De Jan säit, dass mer Akk.-Obj.ⁿ Dat.-Obj.^m
es Huus händ wele Akk.-Vⁿ Dat.-V^m aastriiche.

Es ist beweisbar, ...

- ◆ dass diese Sprache nicht kontextfrei ist.
- ◆ Also kann die ursprüngliche Sprache, Zürichdeutsch, auch nicht kontextfrei sein.

Literaturhinweise

Welcher Formalismus ist adäquat für die Beschreibung natürlicher Sprachen?

- ◆ Walter J. Savitch/Emmon Bach/William Marsh [ed.]: The Formal Complexity of Natural Language. Studies in Linguistics and Philosophy, vol. 33. Dordrecht: Kluwer Academic Publishers, 1987. ISBN 1-55608-047-6 (Paperback); 1-55608-046-8 (gebunden).

Sammelband mit den wichtigsten Arbeiten. Die meisten Papiere sind ohne mathematischen Hintergrund verständlich, allerdings ist etwas Bekanntschaft mit der Automatentheorie nützlich. Eher für Fortgeschrittene.

Literaturhinweise



Nicht-Kontextfreiheit natürlicher Sprachen

◆ Zürichdeutsch

Stuart M. Shieber: Evidence Against the Context-Freeness of Natural Language. *Linguistics and Philosophy* **8** (1985), pp. 333–343.
Abgedruckt in [Savitch et al., 1987]

◆ Bambara

Christopher M. Culy: The Complexity of the Vocabulary of Bambara. *Linguistics and Philosophy* **8** (1985), pp. 345–351.
Abgedruckt in [Savitch et al., 1987]

Literaturhinweise

Nicht-Kontextfreiheit natürlicher Sprachen

◆ Niederländisch

Joan Bresnan/Ronald Kaplan/Stanley Peters/Annie Zaenen: Cross-Serial Dependencies in Dutch. *Linguistic Inquiry* Vol. 12, No. 4, Fall 1982. pp. 613–635. Abgedruckt in [Savitch et al., 1987]

Der Beweis basiert auf der Bedeutung der Sätze und einer bestimmten Konstituentenstruktur, ist also mathematisch weniger sauber als die Arbeiten zu Zürichdeutsch und Bambara.