Formal Models in NLP: Context-free Languages

Daniel Quernheim

April 24, 2012

Outline

Context-free languages

Properties of CFLs

Parsing of CFLs

Do we need more than CFLs?

Outline

Context-free languages

Properties of CFLs

Parsing of CFLs

Do we need more than CFLs?

Linguistic Motivation

Observation:

Human language utterances are recursively built from constituents.

Example:

a bug on the leaves of the plant that Sam loves

Some properties of context-free phrase structure

- contiguous phrases
- can only describe projective dependencies [$\rightsquigarrow$ DepGrammar]

Context-free Grammars

Context-free languages are generated by context-free grammars:

$$G = \langle N, \Sigma, P, S \rangle$$

- N a finite set of nonterminals
- Σ a finite alphabet of terminals
- $S \in N$ the start symbol
- P a finite set of productions of the form

$$A \to \alpha \quad \text{with } A \in N,\ \alpha \in (N \cup \Sigma)^*$$

For the sake of completeness: The equivalent automaton model is the pushdown automaton.

Example Grammar

$$
\begin{aligned}
S &\rightarrow NP\ VP \\
NP &\rightarrow Det\ N \\
NP &\rightarrow Det\ Adj\ N \\
NP &\rightarrow EN \\
VP &\rightarrow V\ NP
\end{aligned}
\qquad
\begin{aligned}
EN &\rightarrow Kim \\
V &\rightarrow feeds \\
Det &\rightarrow the \\
Adj &\rightarrow small \\
N &\rightarrow kangaroos \\
N &\rightarrow wombats
\end{aligned}
$$

Context-free Derivations

Derivation relation $\Rightarrow$ on $(N \cup \Sigma)^*$:

$$\alpha A \gamma \Rightarrow \alpha \beta \gamma \text{ for } A \rightarrow \beta \in P \text{ and } \alpha, \beta, \gamma \in (N \cup \Sigma)^*$$

Example (leftmost derivation)

S $\Rightarrow$ NP VP
$\Rightarrow$ EN VP
$\Rightarrow$ Kim VP
$\Rightarrow$ Kim V NP
$\Rightarrow$ Kim feeds NP
$\Rightarrow$ Kim feeds Det Adj N
$\Rightarrow$ Kim feeds the Adj N
$\Rightarrow$ Kim feeds the small N
$\Rightarrow$ Kim feeds the small wombats

Language of a CFG

- Reflexive and transitive closure:

$$\Rightarrow^*$$

- Example:

S $\Rightarrow^*$ Kim feeds the small wombats

- Language of G:

$$L(G) = \{w \in \Sigma^* : S \Rightarrow^* w\}$$

Outline

Context-free languages

Properties of CFLs

Parsing of CFLs

Do we need more than CFLs?

Closure Properties

Context-free languages are closed under

- Union
- Concatenation
- Star
- Substitution
- Homomorphisms

Context-free languages are not closed under

- Intersection
- Complement

Substitution and Homomorphism

Definition (Substitution)

A substitution is a mapping $f : \Sigma \to 2^{\Delta^*}$, i.e. from symbols to languages.

Generalisation (from strings to languages):

- $f(\epsilon) = \{\epsilon\}$
- $f(x\alpha) = f(x)f(\alpha)$

Generalisation (from languages to languages):

$$f(L) = \bigcup_{w \in L} f(w)$$

Definition (Homomorphism)

A homomorphism is a deterministic substitution.

Substitution and Homomorphism

Substitution example

- $f(0) = \mathbf{a}$
- $f(1) = \mathbf{b}^*$
- Then: $f(010) = \mathbf{ab}^*\mathbf{a}$
- If $L = \mathbf{0}^*(\mathbf{0} + \mathbf{1})\mathbf{1}^*$, then $f(L) = \mathbf{a}^*(\mathbf{a} + \mathbf{b}^*)(\mathbf{b}^*)^* = \mathbf{a}^*\mathbf{b}^*$

Homomorphism example

- $h(0) = aa$
- $h(1) = aba$
- Then $h(010) = aaabaaa$
- If $L = (01)^*$, then $h(L) = (aaaba)^*$

Intersection and Complement

Counterexample

- $L_1 = \{a^i b^i c^j : i \geq 1 \wedge j \geq 1\}$ and
- $L_2 = \{a^i b^j c^j : i \geq 1 \wedge j \geq 1\}$
- are both context-free, but their intersection $L_1 \cap L_2 = \{a^i b^i c^i : i \geq 1\}$ is not!

Proof.

- $L_1 = \mathbf{a^i b^i} \cdot \mathbf{c^j}$. Since $\mathbf{a^i b^i}$ is context-free, and $\mathbf{c^j}$ is regular, L_1 is context-free. The same holds for L_2.
- Use the pumping lemma to show that $L_1 \cap L_2$ is not context-free.

□

This implies: Context-free languages are not closed under complement because $L_1 \cap L_2 = \overline{\overline{L_1} \cup \overline{L_2}}$ (contradiction).

Decision Problems

Decidable properties of context-free languages

Word problem	$w \in L\,?$
Emptiness	$L = \emptyset\,?$
Finiteness	$\lvert L \rvert < \infty\,?$

Undecidable properties of context-free languages

Equivalence	$L_1 = L_2\,?$
Universality	$L_1 = \Sigma^*\,?$
Subset	$L_1 \subseteq L_2\,?$
Disjointness	$L_1 \cap L_2 = \emptyset\,?$

Normal Forms

What are they for?

- Normal forms can simplify proofs
- Normal forms can simplify parsing

- Chomsky normal form
  every rule is of the form $A \to BC$ or $A \to b$
- extended CNF:
  $S \to \epsilon$ is allowed and no other rule derives S
- Greibach normal form
  each rule is of the form $A \to bC_1 \dots C_n$

Chomsky Normal Form

$S \rightarrow NP\ VP$

$NP \rightarrow Det\ N$

$NP \rightarrow Det\ Adj\ N$

$NP \rightarrow Det\ AdjN$

$AdjN \rightarrow Adj\ N$

$NP \rightarrow EN$

$NP \rightarrow Kim$

$VP \rightarrow V\ NP$

$EN \rightarrow Kim$

$V \rightarrow feeds$

$Det \rightarrow the$

$Adj \rightarrow small$

$N \rightarrow kangaroos$

$N \rightarrow wombats$

Pumping Lemma

If L is context-free, then there is $n \in \mathbb{N}$ such that:

For every $z \in L$ with $|z| \geq n$ there are $u, v, w, x, y \in \Sigma^*$ such that:

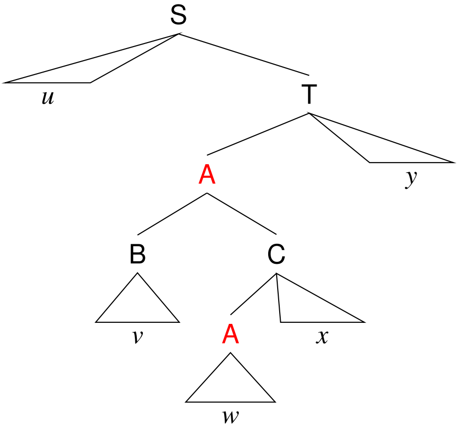
1. $z = uvwxy$
2. $|vx| \geq 1$
3. $|vwx| \leq n$
4. $\forall i \geq 0 : uv^iwx^iy \in L$

Proof sketch

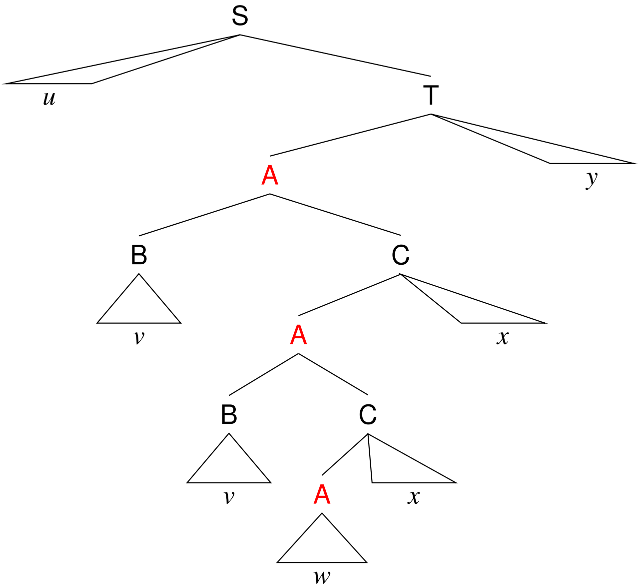
If L is finite, then we can pick n such that $n > |z|$ for every $z \in L$.
If L is infinite, assume that there is a grammar in CNF for L and ...

Pumping Lemma (Proof sketch)

…pick $n = 2^{|N|}$. Derivations of strings of length $\geq n$ have height $\geq |N|$. Some nonterminal must occur more than once on a spine (by the pigeonhole principle).



Pumping Lemma (Proof sketch)



Using the Pumping Lemma

You can use the pumping lemma to show that some languages are not context-free:

- $\mathbf{a^n b^n c^n}$ is not context-free
- $\{ww : w \in \Sigma^*\}$ is not context-free
- $\mathbf{a^m b^n c^m d^n}$ is not context-free

Pumping lemmas also exist for some other language classes (e.g. regular languages)

Outline

Parsing of Context-free Languages

Popular algorithms

- CYK (Cocke, Younger, Kasami): $O(n^3)$
  (only for grammars in Chomsky normal form)
- LR(k) (Knuth): $O(n)$
  (only for LR(k) grammars)
- GLR (Tomita): $O(n^3)$
- Chart parsing (Earley): $O(n^3)$, for unambiguous grammars: $O(n^2)$
- and many others…

Deductive Parsing

How can we systematically compare parsing algorithms?

Write them as inference rules! [Shieber et al., 1995]

Deduction system

- Items are the statements that we prove
- Axioms are the statements that do not have to be proven
- Goals are the statements we want to prove
- Inference rules are used in order to prove statements

CYK Parsing

Item form $[A, i, j]$

Goal $[S, 0, n]$

Axiom $\dfrac{A \rightarrow w_{i+1}}{[A, i, i+1]}$

Rules $\dfrac{[B, i, j] \quad [C, j, k] \quad A \rightarrow BC}{[A, i, k]}$

CYK Parsing

(using the CNF version of the example grammar)

$_0$ Kim $_1$ feeds $_2$ the $_3$ wombats $_4$

1	$[\text{NP}, 0, 1]$	AXIOM
2	$[\text{V}, 1, 2]$	AXIOM
3	$[\text{Det}, 2, 3]$	AXIOM
4	$[\text{N}, 3, 4]$	AXIOM
5	$[\text{NP}, 2, 4]$	from 3 and 4
6	$[\text{VP}, 1, 4]$	from 2 and 5
7	$[\text{S}, 0, 4]$	from 1 and 6

Recursive Descent Parsing

Item form $[\bullet\beta, j]$

Goal $[\bullet, n]$

Axiom $[\bullet S, 0]$

Rules

Scanning $$\frac{[\bullet w_{j+1}\beta, j]}{[\bullet\beta, j+1]}$$

Prediction $$\frac{[\bullet B\beta, j] \quad B \to \gamma}{[\bullet\gamma\beta, j]}$$

Recursive Descent Parsing

$_0$ Kim $_1$ feeds $_2$ the $_3$ wombats $_4$

1	$[\bullet \text{S}, 0]$	AXIOM
2	$[\bullet \text{NP VP}, 0]$	PREDICT from 1
3	$[\bullet \text{EN VP}, 0]$	PREDICT from 2
4	$[\bullet \text{Kim VP}, 0]$	PREDICT from 3
5	$[\bullet \text{VP}, 1]$	SCAN from 4
6	$[\bullet \text{V NP}, 1]$	PREDICT from 5
7	$[\bullet \text{feeds NP}, 1]$	PREDICT from 6
8	$[\bullet \text{NP}, 2]$	SCAN from 7
9	$[\bullet \text{Det N}, 2]$	PREDICT from 8
10	$[\bullet \text{the N}, 2]$	PREDICT from 9
11	$[\bullet \text{N}, 3]$	SCAN from 10
12	$[\bullet \text{wombats}, 3]$	PREDICT from 11
13	$[\bullet , 4]$	SCAN from 12

Shift-Reduce Parsing

Item form $[\alpha \bullet, j]$

Goal $[S \bullet, n]$

Axiom $[\bullet, 0]$

Rules

Shift $$\frac{[\alpha \bullet, j]}{[\alpha w_{j+1} \bullet, j+1]}$$

Reduce $$\frac{B \to \gamma \quad [\alpha\gamma \bullet, j]}{[\alpha B \bullet, j]}$$

Shift-Reduce Parsing

$_0$ Kim $_1$ feeds $_2$ the $_3$ wombats $_4$

1	$[\bullet, 0]$	AXIOM
2	[Kim $\bullet$, 1]	SHIFT from 1
3	[EN $\bullet$, 1]	REDUCE from 2
4	[NP $\bullet$, 1]	REDUCE from 3
5	[NP feeds $\bullet$, 2]	SHIFT from 4
6	[NP V $\bullet$, 2]	REDUCE from 5
7	[NP V the $\bullet$, 3]	SHIFT from 6
8	[NP V Det $\bullet$, 3]	REDUCE from 7
9	[NP V Det wombats $\bullet$, 4]	SHIFT from 8
10	[NP V Det N $\bullet$, 4]	REDUCE from 9
11	[NP V NP $\bullet$, 4]	REDUCE from 10
12	[NP VP $\bullet$, 4]	REDUCE from 11
13	[S $\bullet$, 4]	REDUCE from 12

Earley Parsing

Item form $[i, A \to \alpha \bullet \beta, j]$

Goal $[0, S' \to S \bullet, n]$

Axiom $[0, S' \to \bullet S, 0]$

Rules

Scanning $$\frac{[i, A \to \alpha \bullet w_{j+1}\beta, j]}{[i, A \to \alpha w_{j+1} \bullet \beta, j+1]}$$

Prediction $$\frac{B \to \gamma}{[j, B \to \bullet\gamma, j]} \quad [i, A \to \alpha \bullet B\beta, j]$$

Completion $$\frac{[i, A \to \alpha \bullet B\beta, k] \quad [k, B \to \bullet\gamma, j]}{[i, A \to \alpha B \bullet \beta, j}$$

Earley Parsing

$_0$ Kim $_1$ feeds $_2$ the $_3$ wombats $_4$

1	$[0, \mathrm{S}' \to \bullet \mathrm{S}, 0]$	AXIOM
2	$[0, \mathrm{S} \to \bullet \mathrm{NP\ VP}, 0]$	PREDICT from 1
3	$[0, \mathrm{NP} \to \bullet \mathrm{EN}, 0]$	PREDICT from 2
4	$[0, \mathrm{EN} \to \bullet \mathrm{Kim}, 0]$	PREDICT from 3
5	$[0, \mathrm{EN} \to \mathrm{Kim} \bullet, 1]$	SCAN from 4
6	$[0, \mathrm{NP} \to \mathrm{EN} \bullet, 1]$	COMPLETE from 3 and 5
7	$[0, \mathrm{S} \to \mathrm{NP} \bullet \mathrm{VP}, 1]$	COMPLETE from 2 and 6
8	$[1, \mathrm{VP} \to \bullet \mathrm{V\ NP}, 1]$	PREDICT from 7
9	$[1, \mathrm{V} \to \bullet \mathrm{feeds}, 1]$	PREDICT from 8
10	$[1, \mathrm{V} \to \mathrm{feeds} \bullet, 2]$	SCAN from 9
11	$[1, \mathrm{VP} \to \mathrm{V} \bullet \mathrm{NP}, 2]$	COMPLETE from 8 and 10
12	$[2, \mathrm{NP} \to \bullet \mathrm{Det\ N}, 2]$	PREDICT from 11
13	$[2, \mathrm{Det} \to \bullet \mathrm{the}, 2]$	PREDICT from 12
14	$[2, \mathrm{Det} \to \mathrm{the} \bullet, 3]$	SCAN from 13
15	$[2, \mathrm{NP} \to \mathrm{Det} \bullet \mathrm{N}, 3]$	COMPLETE from 12 and 14
16	$[3, \mathrm{N} \to \bullet \mathrm{wombats}, 3]$	PREDICT from 15
17	$[3, \mathrm{N} \to \mathrm{wombats} \bullet, 4]$	SCAN from 16
18	$[2, \mathrm{NP} \to \mathrm{Det\ N} \bullet, 4]$	COMPLETE from 15 and 17
19	$[1, \mathrm{VP} \to \mathrm{V\ NP} \bullet, 4]$	COMPLETE from 11 and 18
20	$[0, \mathrm{S} \to \mathrm{NP\ VP} \bullet, 4]$	COMPLETE from 7 and 19
21	$[0, \mathrm{S}' \to \mathrm{S} \bullet, 4]$	COMPLETE from 1 and 20

Weighted Parsing

Semiring parsing [Goodman, 1999]

- Items are weighted (e.g. probabilities)
- In each inference, the weights of the main conditions are multiplied (the side conditions only have to be nonzero)
- Different ways to derive an item are added

Semirings

Addition and multiplication are abstract terms for the operations in a *semiring*. Some semirings:

Boolean	$(\{1, 0\}, \vee, \wedge, 0, 1)$
probabilistic	$(\mathbb{R}_0^+, +, \cdot, 0, 1)$
Viterbi	$(\{\mathbb{R}_0^1, \max, \cdot, 0, 1)$ [$\rightsquigarrow$ ViterbiParsing]

Outline

Context-free languages

Properties of CFLs

Parsing of CFLs

Do we need more than CFLs?

Non-context-freeness in Natural Languages

Swiss German [Shieber, 1985]

(1)	Jan säit	das mer	d'chind	em Hans	es huus	lönd
	Jan says that we		the children.ACC	Hans.DAT	the house.ACC	let

hälfe aastriiche
help paint

Jan says that we let the children help Hans paint the house.

- Subcategorisation: Dative and accusative
- Order reflects argument agreement (colour-coded)
- → Cross-serial dependencies
- can be mapped by homomorphism and intersection with regular languages to $\mathbf{w} \cdot \mathbf{a}^{\mathbf{m}}\mathbf{b}^{\mathbf{n}} \cdot \mathbf{x} \cdot \mathbf{c}^{\mathbf{m}}\mathbf{d}^{\mathbf{n}} \cdot \mathbf{y}$ which is not context-free

Non-context-freeness in Natural Languages

German [Groenink, 1997]

(2) . . . , dass Frank Julia Fred schwimmen helfen ließ und
. . . , that Frank Julia Fred swim help let and
ertrinken lassen sah.
drown let saw.

. . . , that Frank let Julia help Fred swim and saw her let him drown.

(3) . . . , dass Hans Frank Julia Fred schwimmen helfen hören
ließ und ertrinken lassen sehen ließ.

Beyond Context-freeness

Mildly context-sensitive languages

A class of languages is mildly context-sensitive (fuzzy definition!) if:

1. it includes the class of context-free languages,
2. it admits a limited form of cross-serial dependencies,
3. it is parsable in polynomial time, and
4. if it has the *constant growth property*, i.e. if you sort words by length, the length difference between adjacent words is bounded by a constant.

- This is a list of desirable properties (from a practical/linguistic point of view), rather than a sharp characterisation!

Some Mildly Context-sensitive Formalisms

Equivalent to Embedded Pushdown Automata[1]

- Combinatory Categorial Grammar [$\rightsquigarrow$ CCG]
- Tree-Adjoining Grammar [$\rightsquigarrow$ TAG]
- Linear Indexed Grammar
- Head Grammar

Equivalent to Thread Automata (more powerful)

- Simple Range Concatenation Grammar [$\rightsquigarrow$ RCG]
- Multi-Component Tree-Adjoining Grammar [$\rightsquigarrow$ TAG]
- Minimalist Grammar
- Linear Context-Free Rewriting Systems
- Multiple Context-Free Grammars

---

[1] [Vijay-Shanker and Weir, 1994]

References

Goodman, J. (1999).
Semiring parsing.
Computational Linguistics, 25:573–605.

Groenink, A. V. (1997).
Surface without Structure: Word order and tractability issues in natural language analysis.
PhD thesis, Universiteit Utrecht.

Hopcroft, J. E. and Ullman, J. D. (1979).
Introduction to Automata Theory, Languages and Computation.
Addison-Wesley.

Shieber, S. M. (1985).
Evidence against the context-freeness of natural language.
Linguistics and Philosophy, 8:333–343.

Shieber, S. M., Schabes, Y., and Pereira, F. C. N. (1995).
Principles and implementation of deductive parsing.
J. Log. Program., 24(1&2):3–36.

Vijay-Shanker, K. and Weir, D. J. (1994).
The equivalence of four extensions of context-free grammars.
Mathematical Systems Theory, 27(6):511–546.