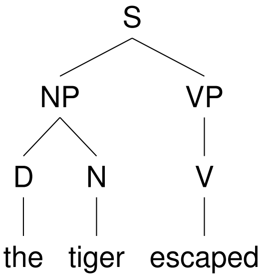
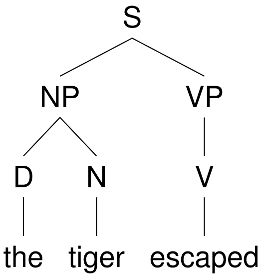
An Algebraic Approach to MCFGs

Alexander Clark and Ryo Yoshinaka

Department of Philosophy, King's College, London
Department of Informatics, Kyoto University

LACL 2014


S
NP
VP
D
N
V
the
tiger
escaped



Good questions

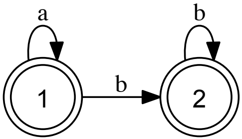
- What does the symbol NP mean?
- What does a production NP $\rightarrow$ D N mean?

Deterministic Finite Automata

Myhill-Nerode theorem

There is a one-to-one correspondence between the states of a minimal DFA and the residual languages.

Example $L = a^* b^*$



$$[[1]] = a^* b^*$$

$$[[2]] = b^*$$

Result for CFGs

[Clark, 2013]

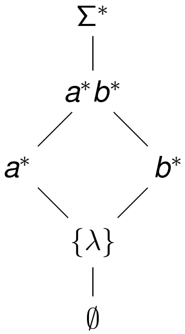
Mergeable nonterminals

Two nonterminals are mergeable if we can merge them without changing the generated language.

Theorem

If a CFG has no mergeable nonterminals then there is a one-to-one correspondence between the nonterminals and elements of the syntactic concept lattice.

$L = a^* b^*$

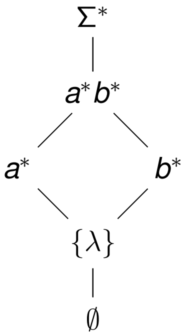


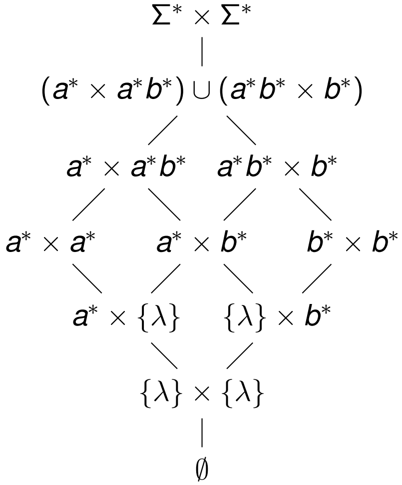
This paper

Theorem

If a MCFG of dimension 2 has no mergeable nonterminals then there is a one-to-one correspondence between the nonterminals and elements of the syntactic concept lattice of order 2.

$L = a^* b^*$


Σ^*
$a^* b^*$
a^*
b^*
$\{\lambda\}$
$\emptyset$


$\Sigma^* \times \Sigma^*$
$(a^* \times a^* b^*) \cup (a^* b^* \times b^*)$
$a^* \times a^* b^*$
$a^* b^* \times b^*$
$a^* \times a^*$
$a^* \times b^*$
$b^* \times b^*$
$a^* \times \{\lambda\}$
$\{\lambda\} \times b^*$
$\{\lambda\} \times \{\lambda\}$
$\emptyset$

1. An algebraic approach to CFGs
2. The extension to MCFGs of dimension 2

1. An algebraic approach to CFGs
2. The extension to MCFGs of dimension 2

Outline

1. Least fixed point semantics for (M)CFGs
2. Distributional learning
3. Simplicity and minimality of grammars

What Algebra?

Monoid: $\langle S, \circ, 1 \rangle$

Free monoid over Σ:

$$\langle \Sigma^*, \cdot, \lambda \rangle$$

What Algebra?

Monoid: $\langle S, \circ, 1 \rangle$

Free monoid over Σ:

$$\langle \Sigma^*, \cdot, \lambda \rangle$$

Complete Idempotent Semiring: $\langle S, \circ, 1, \vee, \bot \rangle$

Free CIS over Σ:

$$\langle \mathcal{P}(\Sigma^*), \cdot, \{\lambda\}, \cup, \emptyset \rangle$$

Context-Free Grammars

Context-Free Grammar

$G = \langle \Sigma, V, S, P \rangle$
$\mathcal{L}(G, A) = \{w \in \Sigma^* \mid A \stackrel{*}{\Rightarrow}_G w\}$

Example

$\Sigma = \{a, b\}, V = \{S\}$
$P = \{S \rightarrow ab, S \rightarrow aSb, S \rightarrow \epsilon\}$

$$\mathcal{L}(G, S) = \{a^n b^n \mid n \geq 0\}$$

Least fixed point semantics

Ginsburg and Rice, 1962

Interpret this as a set of equations in $\mathcal{P}(\Sigma^*)$

$$S = (\{a\} \circ \{b\}) \vee (\{a\} \circ S \circ \{b\}) \vee \{\lambda\}$$

Least fixed point semantics

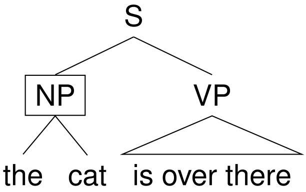
Ginsburg and Rice, 1962

Interpret this as a set of equations in $\mathcal{P}(\Sigma^*)$

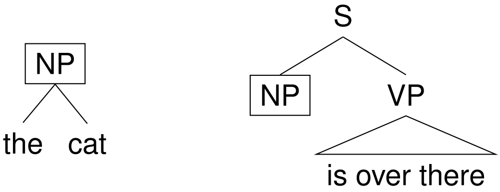
$$S = (\{a\} \circ \{b\}) \vee (\{a\} \circ S \circ \{b\}) \vee \{\lambda\}$$

The least solution of this equation is $S = \{a^n b^n \mid n \geq 0\}$.

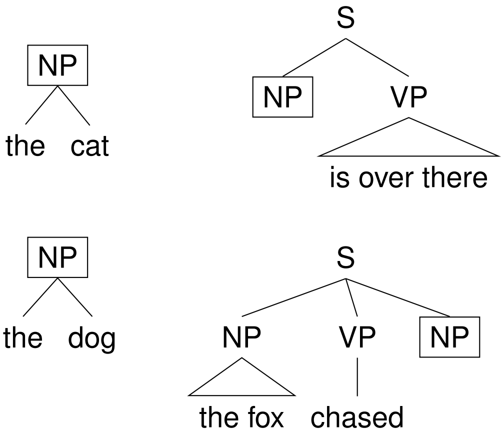
Distributional Learning



Distributional Learning



Distributional Learning



Contexts

A context is a string with a hole:

$$l\Box r$$

Derivation contexts

The derivation contexts of CFGs are just string contexts:

$$S \stackrel{*}{\Rightarrow}_G lNr$$

Definition

Filling the hole

$l\Box r \odot u = lur$

A factorisation of a language L

C is a set of contexts; S is a set of strings

$$C \odot S \subseteq L$$

Context free grammars

Contexts and yields

$\mathcal{L}(G, N) = \{w \in \Sigma^* \mid N \overset{*}{\Rightarrow} w\}$
$\mathcal{C}(G, N) = \{l\Box r \mid S \overset{*}{\Rightarrow} lNr\}$.

Nonterminals in a context-free grammar

$$\mathcal{C}(G, N) \odot \mathcal{L}(G, N) \subseteq L$$

Maximal decompositions into contexts and substrings

Definition

A *syntactic concept* is a pair $\langle S, C \rangle$ where

- where S is a set of strings, and C is a set of contexts
- $C \odot S \subseteq L$
- C and S are both maximal

Polar maps

If S is a set of strings:

$$S^{\rhd} = \{l\Box r \mid \forall u \in S, lur \in L\} \tag{1}$$

If C is a set of contexts:

$$C^{\lhd} = \{u \in \Sigma^* \mid \forall l\Box r \in C, lur \in L\} \tag{2}$$

Polar maps

If S is a set of strings:

$$S^{\triangleright} = \{l\Box r \mid \forall u \in S, lur \in L\} \quad (1)$$

If C is a set of contexts:

$$C^{\triangleleft} = \{u \in \Sigma^* \mid \forall l\Box r \in C, lur \in L\} \quad (2)$$

$\cdot^{\triangleright}$ and $\cdot^{\triangleleft}$ form a Galois connection between sets of strings and sets of contexts.

Closed sets of strings

$\cdot^{\triangleright\triangleleft}$ is a closure operator on the sets of strings.

definition

A set of strings is closed if $X = X^{\triangleright\triangleleft}$.

L is always closed.

Closed sets of strings

$\cdot^{\rhd\lhd}$ is a closure operator on the sets of strings.

definition

A set of strings is closed if $X = X^{\rhd\lhd}$.
L is always closed.

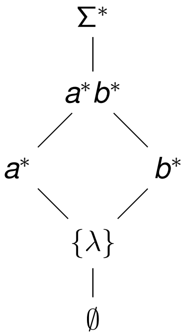
The syntactic concept lattice

The set of all closed sets of strings form a complete idempotent semiring: $\mathfrak{B}(L)$.

- A residuated lattice which gives a sound and complete semantics for the Lambek calculus.

Language is regular iff lattice is finite

$L = a^* b^*$



Some different CIS

Free CIS

$\mathcal{P}(\Sigma^*)$

Syntactic Concept Lattice

$\mathfrak{B}(L)$

Some different CIS

Free CIS
$\mathcal{P}(\Sigma^*)$

Syntactic Concept Lattice
$\mathfrak{B}(L)$

Trivial one element CIS
$\mathbf{1}_{CIS}$
(equal to $\mathfrak{B}(\Sigma^*)$)

Homomorphism

Suppose h is a CIS-homomorphism from $\mathcal{P}(\Sigma^*)$ to some other CIS:

$$S = (\{a\} \circ \{b\}) \vee (\{a\} \circ S \circ \{b\}) \vee \{\lambda\}$$

$$h(S) = (h(\{a\}) \circ h(\{b\})) \vee (h(\{a\}) \circ h(S) \circ h(\{b\})) \vee h(\{\lambda\})$$

Residual

Define the residual of f, f^*.

- $f : A \to B$ is a CIS-homomorphism
- $f^* : B \to A$ is $f(y) = \bigvee\{x \mid f(x) \leq y\}$

- Suppose B is the trivial one element CIS with unique element 1: $f^*(1) = \Sigma^*$.

Recognising a language

Definition

We say that a CIS B recognizes L if there is a surjective morphism h from $\mathcal{P}(\Sigma^*) \to B$ such that $h^*(h(L)) = L$.

Examples

- $\mathcal{P}(\Sigma^*)$ recognizes L through the identity.
- $\mathfrak{B}(L)$ recognizes L through the function $X \to X^{\triangleright\triangleleft}$
- Unless $L = \Sigma^*$, $\mathbf{1}_{CIS}$ does not recognize L since $h^*(h(L) = \Sigma^*$

Recognising a language

Definition

We say that a CIS B recognizes L if there is a surjective morphism h from $\mathcal{P}(\Sigma^*) \to B$ such that $h^*(h(L)) = L$.

Examples

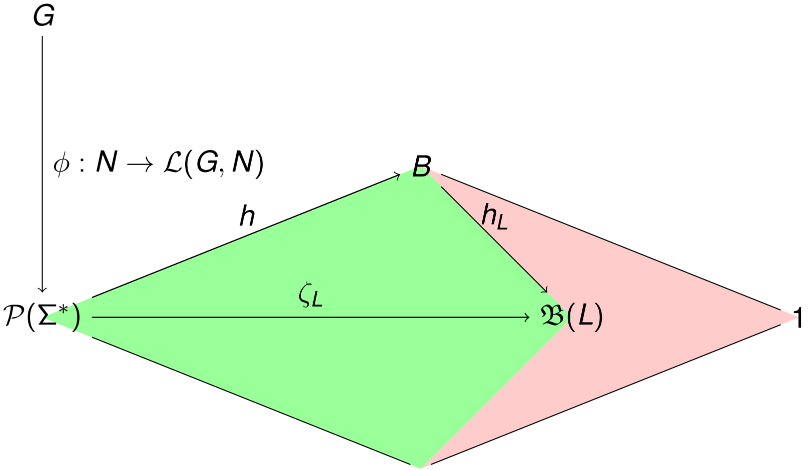
- $\mathcal{P}(\Sigma^*)$ recognizes L through the identity.
- $\mathfrak{B}(L)$ recognizes L through the function $X \to X^{\triangleright\triangleleft}$
- Unless $L = \Sigma^*$, $\mathbf{1}_{CIS}$ does not recognize L since $h^*(h(L) = \Sigma^*$

B has enough structure to represent the CFG derivations of *L*.

Universal Property

There is a unique-up-to-isomorphism ’smallest’ CIS that recognizes L: this is $\mathfrak{B}(L)$.

The universal CIS-morphism



CIS-homomorphisms define CFG-morphisms

Given a CIS B and a homomorphism $h : \mathcal{P}(\Sigma^*) \to B$, we can define a new grammar $\phi_h(G)$ by merging nonterminals M, N if

$$h(\mathcal{L}(G, M)) = h(\mathcal{L}(G, N))$$

CIS-homomorphisms define CFG-morphisms

Given a CIS B and a homomorphism $h : \mathcal{P}(\Sigma^*) \to B$, we can define a new grammar $\phi_h(G)$ by merging nonterminals M, N if

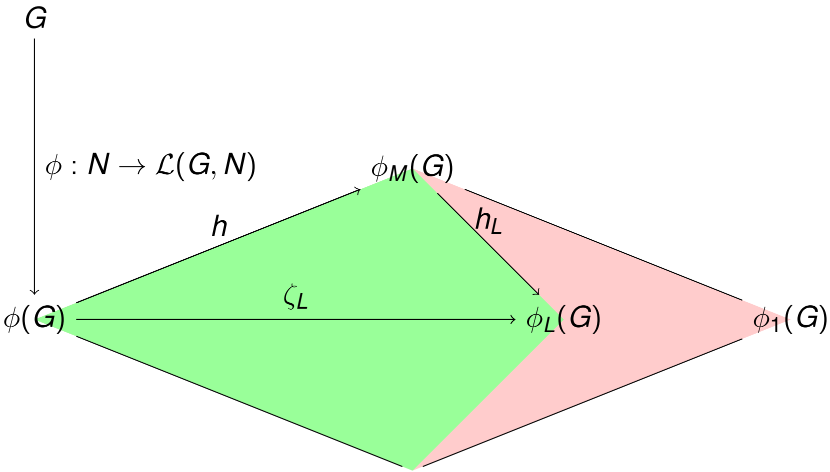
$$h(\mathcal{L}(G, M)) = h(\mathcal{L}(G, N))$$

- If we use $\mathcal{P}(\Sigma^*)$ then we merge only nonterminals that generate the same language, and the resulting grammar will generate the same language.
- If we use $\mathbf{1}_{CIS}$ we merge all of the nonterminals into a single nonterminal and presumably the grammar will overgenerate.

Theorem

Let G be a CFG over Σ and h a homomorphism $h : \mathcal{P}(\Sigma^*) \to B$. Then if B recognizes L through h, $\phi_h(G)$ defines the same language as G.

The universal CFG-morphism



[Clark, 2013]

Mergeable nonterminals

If

$$\mathcal{L}(G, M)^{\triangleright\triangleleft} = \mathcal{L}(G, N)^{\triangleright\triangleleft}$$

then we can merge *M* and *N* without increasing the language defined by *G*,

[Clark, 2013]

Mergeable nonterminals

If

$$\mathcal{L}(G, M)^{\triangleright\triangleleft} = \mathcal{L}(G, N)^{\triangleright\triangleleft}$$

then we can merge *M* and *N* without increasing the language defined by *G*,

Minimal grammars correspond to maximal factorisations

A grammar without mergable nonterminals will have nonterminals that correspond to syntactic concepts.

Multiple context free grammars of dimension 2

A mild extension of context free grammars; the class of well-nested 2-MCFGs is equivalent to TAG, LIG, CCG ...

- Some nonterminals have dimension 1 and derive strings
- Some nonterminals have dimension 2 and derive pairs of strings

Notation for productions

$$N \to PQ$$

$$N(uv) :- P(u)Q(v)$$

- What algebra?

- What algebra?

A two sorted algebra.

- Nonterminals of dim 1 correspond to elements of sort 1 "sets of strings"
- Nonterminals of dim 2 correspond to elements of sort 2 "sets of pairs of strings"

A tupling operation $\otimes$ that takes two elements of sort 1 and produces an element of sort 2.

Paired CIS

$$\langle A_1, A_2, \circ, \otimes, \vee^1, \vee^2, \epsilon, \bot \rangle$$

Paired CIS

$$\langle A_1, A_2, \circ, \otimes, \vee^1, \vee^2, \epsilon, \perp \rangle$$

Free Paired CIS over Σ

$$\langle \mathcal{P}(\Sigma^*), \mathcal{P}(\Sigma^* \times \Sigma^*), \cdot, \times, \cup, \cup, \{\lambda\}, \emptyset \rangle$$

Two nontrivial axioms

Axiom E

For all $x \in A_2$,

$$x = \bigvee^{2} \{u \otimes v \mid u, v \in A_1, u \otimes v \leq_2 x\}.$$

Two axioms

Axiom V

For any set $J \subseteq A_1 \times A_1$, and any $p, q, a, b, c, d \in A_1$ if

$$p \otimes q \leq_2 \bigvee \{u \otimes v \mid (u, v) \in J\}$$

then

$$p \circ q \leq_1 \bigvee \{u \circ v \mid (u, v) \in J\}$$

and

$$(a \circ p \circ b) \otimes (c \circ q \circ d) \leq_2 \bigvee \{(a \circ u \circ b) \otimes (c \circ v \circ d) \mid (u, v) \in J\} .$$

Representing rules

Production

$$N(z_{1,1}z_{2,1}, z_{2,2}z_{1,2}) :- P(z_{1,1}, z_{1,2}), Q(z_{2,1}, z_{2,2})$$

Definition of Algebraic operation

$$x \ominus y = \bigvee\{(z_{1,1} \circ z_{2,1}) \otimes (z_{2,2} \circ z_{1,2}) \mid (z_{1,1} \otimes z_{1,2}) \leq x, (z_{2,1} \otimes z_{2,2}) \leq y\}$$

Lemma

For any homomorphism *h*:

$$h(x \ominus y) = h(x) \ominus h(Y)$$

Representing rules

Production

$$N(z_{1,1}z_{2,1}, z_{2,2}z_{1,2}) :- P(z_{1,1}, z_{1,2}), Q(z_{2,1}, z_{2,2})$$

Definition of Algebraic operation

$$x \ominus y = \bigvee \{(z_{1,1} \circ z_{2,1}) \otimes (z_{2,2} \circ z_{1,2}) \mid (z_{1,1} \otimes z_{1,2}) \leq x, (z_{2,1} \otimes z_{2,2}) \leq y\}$$

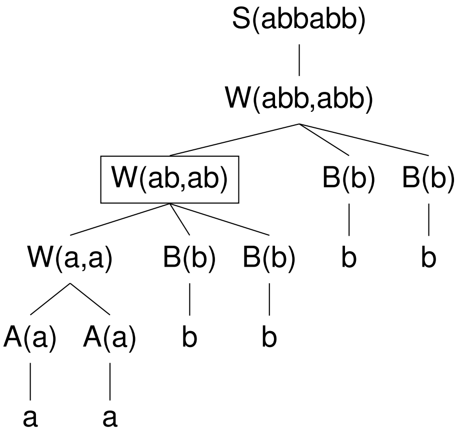
Lemma

For any homomorphism *h*:

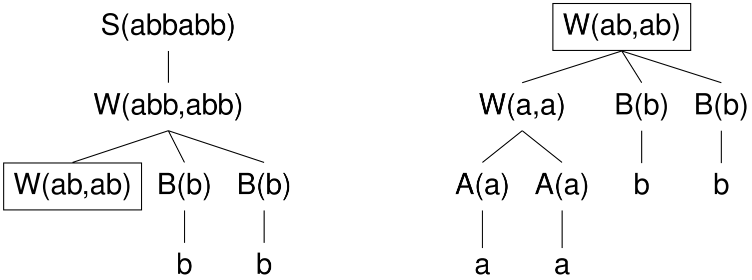
$$h(x \ominus y) = h(x) \ominus h(Y)$$

and similarly for all of the infinitely many possible functions ...

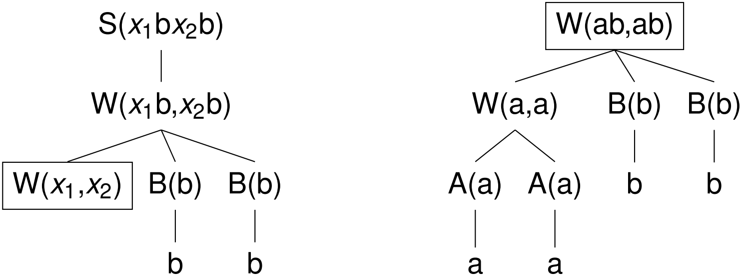
MCFG derivation contexts



MCFG derivation contexts



MCFG derivation contexts



$$\Box b \Box b \odot \langle ab, ab \rangle = abbabb$$

Derivation contexts

The derivation contexts of MCFGs are just string contexts:

$$N(x_1, x_2) \vdash_G S(l x_1 m x_2 r)$$

Contexts

A context is a string with one or two holes:

$$l \Box r$$

$$l \Box m \Box r$$

Main result

Lattice of order 2

- Sort 1 is just $\mathfrak{B}(L)$
- Sort 2

$$\{X \in \mathcal{P}(\Sigma^* \times \Sigma^*) \mid X = X^{\triangleright\triangleleft}\}$$

Unique smallest PCIS that recognizes the language.

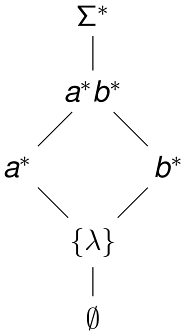
Theorem

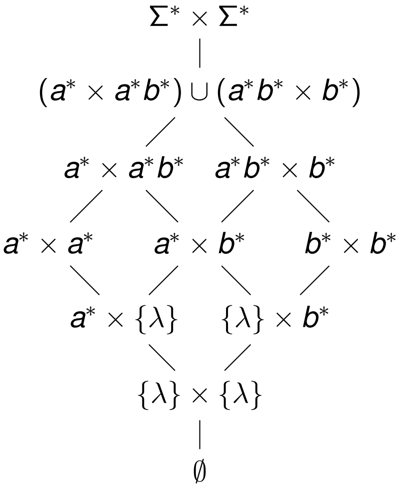
A 2-MCFG without mergeable nonterminals has nonterminals (of dimension d) that correspond to elements of the lattice (of order d):

$$N \to \mathcal{L}(G, N)^{\triangleright\triangleleft}$$

Language is regular iff lattice is finite

$L = a^* b^*$





An answer to our naive questions

What do the nonterminals mean?

They refer to elements of the Galois lattice between derivation contexts and yields.

An answer to our naive questions

What do the nonterminals mean?

They refer to elements of the Galois lattice between derivation contexts and yields.

What do the productions mean?

They state inequalities in this structure.

$$N(z_{1,1}z_{2,1}, z_{2,2}z_{1,2}) :- P(z_{1,1}, z_{1,2}), Q(z_{2,1}, z_{2,2})$$

means that

$$[[N]] \geq [[P]] \ominus [[Q]]$$

Conclusions

- Minimal MCFGs correspond to elements of these lattices. (Minimal in a very weak sense)
- Under the reasonable assumption that we don't need mergeable nonterminals, MCFGs have distributionally definable nonterminals.
- Nonterminals in MCFGs represent elements of these lattices.

Open questions

- The extension to all MCFGs of dimension $k > 2$ seems straightforward. Is it?
- Is it reasonable to assume that the grammar has no mergeable nonterminals?
- Are these models for (a fragment of) Morrill's displacement calculus? Completeness?
- Can we use this to get strong learning results for MCFGs?

Bibliography I

Clark, A. (2013).
The syntactic concept lattice: Another algebraic theory of the context-free languages?
Journal of Logic and Computation.