

Learning Context Free Grammars with the Finite Context Property: A Correction of A. Clark's Algorithm

Hans Leiß

Centrum für Informations- und Sprachverarbeitung
Universität München
Oettingenstr.67, 80538 München
leiss@cis.uni-muenchen.de**

Abstract. A. Clark[2] has shown that the class of languages which have a context-free grammar whose nonterminals can be defined by a finite set of contexts can be identified in the limit, given an enumeration of the language and a test for membership. We show by example that Clark's algorithm may converge to a grammar that does not define the input language. We review the theoretical background, provide a non-obvious modification of the algorithm and prove its correctness.

1 Introduction

An important goal of structural linguistics was to analyse and describe a language in terms of distributions. Given an alphabet Σ , the *distribution* of a word $v \in \Sigma^*$ with respect to a language $L \subseteq \Sigma^*$ is the set

$$D(v) = \{(u, w) \in \Sigma^* \times \Sigma^* \mid uvw \in L\}$$

of all contexts where v appears in L . Having the same distribution with respect to L is a congruence relation \equiv_L on Σ^* , the *syntactic congruence*. It partitions Σ^* into disjoint distribution classes $[v] := \{v' \in \Sigma^* \mid v \equiv_L v'\}$. By the Myhill/Nerode theorem (c.f.[6]), \equiv_L has finitely many distribution classes if and only if L is a regular language.

When the monoid operations of $(\Sigma^*, \cdot, \epsilon)$ are lifted to word sets by $A \cdot B := \{a \cdot b \mid a \in A, b \in B\}$ and $1 = \{\epsilon\}$, one obtains a monoid $(\mathcal{P}(\Sigma^*), \cdot, 1)$, which is partially ordered by \subseteq . The operation $(u, w) \odot v := uvw$ of *filling* a context (u, w) with a word v is lifted to context sets C and word sets A by

$$C \odot A := \{(u, w) \odot v \mid (u, w) \in C, v \in A\}.$$

With respect to L , each set of contexts C has a largest set of fillers, $C^\triangleleft = \{v \in \Sigma^* \mid C \odot \{v\} \subseteq L\}$, and each set A of words has a largest set of contexts, $A^\triangleright = \{(u, w) \in \Sigma^* \times \Sigma^* \mid \{(u, w)\} \odot A \subseteq L\}$. Notice that $D(v) = \{v\}^\triangleright =: v^\triangleright$ and

$$\{v\}^{\triangleright\triangleleft} = \{u \mid v^\triangleright \subseteq u^\triangleright\} \supseteq \{u \mid v^\triangleright = u^\triangleright\} = [v].$$

** A shorter version will appear in Morill e.a (eds) Proc. 19th Conf. on Formal Grammar 2014. The final publication is available at link.springer.com

The maps $A \mapsto A^{\triangleright\triangleleft}$ and $C \mapsto C^{\triangleleft\triangleright}$ are closure operators. Via a Galois-connection between sets of words and sets of contexts, the partial orders of closed sets of words and closed sets of contexts are anti-isomorphic. Clark[2] defines a (*syntactic*) *concept of L* to be a pair $\langle A, C \rangle$ such that $A^{\triangleright} = C$ and $C^{\triangleleft} = A$. As here each component is closed and determines the other one, one can use the component which is the better representation for a given purpose.

Note that L is itself a concept, namely $L = \{(\epsilon, \epsilon)\}^{\triangleleft}$, and can be represented by a finite set of contexts. Suppose L has a context-free grammar G (in Chomsky normal form, CNF) whose nonterminals N are concepts of L , i.e. $N = N^{\triangleright\triangleleft}$ when N is identified with $\{w \in \Sigma^* \mid N \Rightarrow_G^* w\}$. A branching rule ($N \rightarrow AB$) of G then corresponds to $N \supseteq AB$, which is equivalent to $N \supseteq (AB)^{\triangleright\triangleleft}$. If N, A, B are represented by context sets C, D, E , this means $C^{\triangleleft} \supseteq (D^{\triangleleft} E^{\triangleleft})^{\triangleright\triangleleft}$.

A. Clark[2] developed an algorithm to identify in the limit a CNF-grammar for L from membership queries and an enumeration of L , provided L has a CNF-grammar whose nonterminals can be defined by finite sets of contexts. The basic idea is to extract from a finite subset $E \subseteq L$ a finite set F of contexts and a finite set K of subwords of L and relativize the operations \cdot^{\triangleright} and \cdot^{\triangleleft} of taking all contexts resp. fillers with respect to L to $\cdot^F : \mathcal{P}(\Sigma^*) \rightarrow \mathcal{P}(F)$ and $\cdot^K : \mathcal{P}(F) \rightarrow \mathcal{P}(K)$ by $A^F := A^{\triangleright} \cap F$ and $C^K := C^{\triangleleft} \cap K$. Then there are only finitely many relativized concepts of L , the pairs $\langle A, C \rangle \in \mathcal{P}(K) \times \mathcal{P}(F)$ where $A = C^K$ and $C = A^F$; using relativized closed sets $C = C^{KF}$ of contexts to represent them, Clark builds a CNF-grammar $G(K, L, F)$ by taking as branching rules those triples $(C \rightarrow DE)$ where $C^K \supseteq (D^K E^K)^{FK}$. It is claimed that as K and F increase, the grammars $G(K, L, F)$ converge to a grammar for L .

But there is a technical problem: Clark's criterion for C, D, E to form a grammar rule is right when working with infinite filler sets, i.e. $C^{\triangleleft} \supseteq (D^{\triangleleft} E^{\triangleleft})^{\triangleright\triangleleft}$, since $\cdot^{\triangleright\triangleleft}$ is a closure operator on $\mathcal{P}(\Sigma^*)$ and hence $D^{\triangleleft} E^{\triangleleft} \subseteq (D^{\triangleleft} E^{\triangleleft})^{\triangleright\triangleleft}$. But the criterion is not correct with finite filler sets, i.e. $C^K \supseteq (D^K E^K)^{FK}$ is not equivalent to $C^K \supseteq D^K E^K$: although \cdot^{FK} is a closure operator on $\mathcal{P}(K)$, it is generally not the case that $D^K E^K \subseteq (D^K E^K)^{FK}$, as the left hand side need not be a subset of K . We give an example where Clark's algorithm does not converge to a grammar of the intended language.

Clark's algorithm can be fixed by three modifications: (i) the criterion for when three concepts C, D, E constitute a grammar rule has to be changed from $C^K \supseteq (D^K E^K)^{FK}$ to $C \subseteq (D^K E^K)^F$. Since this works directly with context sets, it emphasizes the importance of the Galois correspondence between word sets and context sets. (ii) the criterion in the learning algorithm that makes the hypothesis grammar shrink is weakened; the effect is that the learner cannot converge to a grammar that defines a strict superset of the intended language. (iii) since for our modified definition, $L(G(K, L, F))$ is neither antitone in K nor monotone in F (as pointed out by R.Yoshinaka), we need a different line of reasoning to show the convergence of the grammar inference process.

We prove that the modified algorithm indeed identifies in the limit the class of context free languages with the finite context property. If we admit only concepts whose context sets are closures of bounded context sets, the algorithm

has polynomial update time. Ryo Yoshinaka [8] has a different modification of Clark's algorithm where nonterminals need not be closed sets. This makes the reasoning simpler, but results in a larger grammar.

2 Correspondence between word and context sets

Let Σ be a finite alphabet and $L \subseteq \Sigma^*$ a set of words. An L -context of $A \subseteq \Sigma^*$ is a word pair or *context* (u, w) such that $uAw \subseteq L$. The *largest set of L -contexts of $A \subseteq \Sigma^*$* is

$$A^\triangleright := \{(u, w) \mid \{(u, w)\} \odot A \subseteq L\} = \{(u, w) \mid uAw \subseteq L\}. \quad (1)$$

An L -filler of a set $C \subseteq \Sigma^* \times \Sigma^*$ of word pairs is a word $w \in \Sigma^*$ such that $uwv \in L$ for each $(u, v) \in C$. Let $C \odot A := \{uvw \mid (u, v) \in C, w \in A\}$. The *largest set of L -fillers of $C \subseteq \Sigma^* \times \Sigma^*$* is

$$C^\triangleleft := \{v \mid C \odot \{v\} \subseteq L\}. \quad (2)$$

If $C \neq \emptyset$, $C^\triangleleft \subseteq \text{Sub}(L)$, where $\text{Sub}(L) := \{v \in \Sigma^* \mid \exists u, w \in \Sigma^* uvw \in L\}$ is the set of all subwords of L . The following equivalence (3) is easy to check:

Proposition 1. *The functions $\cdot^\triangleright : (\mathcal{P}(\Sigma^*), \supseteq) \rightleftharpoons (\mathcal{P}(\Sigma^* \times \Sigma^*), \subseteq) : \cdot^\triangleleft$ form a Galois-connection, i.e., for all $A \in \mathcal{P}(\Sigma^*)$ and $C \in \mathcal{P}(\Sigma^* \times \Sigma^*)$ we have*

$$C^\triangleleft \supseteq A \iff C \subseteq A^\triangleright. \quad (3)$$

In particular, we have:

- (i) \cdot^\triangleright and \cdot^\triangleleft are antitone: $A^\triangleright \supseteq B^\triangleright$ for all word sets $A \subseteq B$ and $C^\triangleleft \supseteq D^\triangleleft$ for all context sets $C \subseteq D$.
- (ii) $A^{\triangleright\triangleleft} = A^\triangleright$ and $C^{\triangleleft\triangleright} = C^\triangleleft$ for all word sets A and context sets C .
- (iii) $\cdot^{\triangleright\triangleleft}$ is a closure operator on $(\mathcal{P}(\Sigma^*), \subseteq)$ and $\cdot^{\triangleleft\triangleright}$ is a closure operator on $(\mathcal{P}(\Sigma^* \times \Sigma^*), \subseteq)$.
- (iv) $\cdot^\triangleright : (\{C^\triangleleft \mid C \subseteq \Sigma^* \times \Sigma^*\}, \supseteq) \rightleftharpoons (\{A^\triangleright \mid A \subseteq \Sigma^*\}, \subseteq) : \cdot^\triangleleft$ form an order isomorphism between the closed word sets and the closed context sets.

Claims (i)-(iv) are standard consequences of a Galois-connection (see Appendix 1, Lemma 7). Since $A \subseteq L$ gives $(\epsilon, \epsilon) \in A^\triangleright$, hence $A^{\triangleright\triangleleft} \subseteq \{(\epsilon, \epsilon)\}^\triangleleft = L$, we have

$$A \subseteq L \iff A^{\triangleright\triangleleft} \subseteq L. \quad (4)$$

In particular, L itself is closed: $L^{\triangleright\triangleleft} = L$.

Proposition 2. *For all $A, B \subseteq \Sigma^*$, $(A^{\triangleright\triangleleft} B^{\triangleright\triangleleft})^{\triangleright\triangleleft} = (AB)^{\triangleright\triangleleft}$.*

Proof. Since $\cdot^{\triangleright\triangleleft}$ is a closure operator on $\mathcal{P}(\Sigma^*)$, it is sufficient to show that $A^{\triangleright\triangleleft} B^{\triangleright\triangleleft} \subseteq (AB)^{\triangleright\triangleleft}$. Let $a \in A^{\triangleright\triangleleft}$, $b \in B^{\triangleright\triangleleft}$, hence $A^\triangleright \subseteq a^\triangleright$ and $B^\triangleright \subseteq b^\triangleright$. Moreover, let $(u, w) \in (AB)^\triangleright$, so $uABw \subseteq L$. Since $\{u\} \times Bw \subseteq A^\triangleright \subseteq a^\triangleright$ we have $uBw \subseteq L$, so $(ua, w) \in B^\triangleright \subseteq b^\triangleright$, and $uabw \in L$. Thus, $ab \in (AB)^{\triangleright\triangleleft}$.

For the *residuation* $A/B := \{v \in \Sigma^* \mid \{v\}B \subseteq A\}$ of word sets A, B , we have:

Proposition 3. *For all $A, B \subseteq \Sigma^*$, $(A/B)^{\triangleright\triangleleft} \subseteq A^{\triangleright\triangleleft}/B$. If $A = A^{\triangleright\triangleleft}$, then $(A/B)^{\triangleright\triangleleft} = A/B$.*

Proof. Pick $v \in (A/B)^{\triangleright\triangleleft}$. We need to show $v \in A^{\triangleright\triangleleft}/B$, which is equivalent to $vB \subseteq A^{\triangleright\triangleleft}$, i.e. that $xvBy \subseteq L$ for all $(x, y) \in A^\triangleright$. So let $(x, y) \in A^\triangleright$. Then

$$x \cdot A/B \cdot B \cdot y \subseteq xAy \subseteq L,$$

hence $\{(x, by) \mid b \in B\} \subseteq (A/B)^\triangleright$. Therefore, since $v \in (A/B)^{\triangleright\triangleleft}$, we have $xvBy \subseteq L$. Thus, $vB \subseteq A^{\triangleright\triangleleft}$, which shows $v \in A^{\triangleright\triangleleft}/B$. So $(A/B)^{\triangleright\triangleleft} \subseteq A^{\triangleright\triangleleft}/B$. If A is closed, $A/B \subseteq (A/B)^{\triangleright\triangleleft} \subseteq A^{\triangleright\triangleleft}/B = A/B$.

2.1 The residuated lattice $\mathcal{B}(L)$ of all concepts of L

A syntactic *concept* of L is a pair $\langle A, C \rangle$ of a word set A and context set C such that $A^\triangleright = C$ and $C^\triangleleft = A$. Note that here $A = A^{\triangleright\triangleleft}$ and $C = C^{\triangleleft\triangleright}$ are closed sets, and by the order isomorphism of Proposition 1 (iv) and (ii), one can represent a concept $\langle A, C \rangle$ by its closed word set $A^{\triangleright\triangleleft}$ or by its closed context set $C^{\triangleleft\triangleright}$.

Theorem 1. (Clark[1]) *The set $B(L)$ of all concepts of L forms a monoidal, residuated and complete lattice,*

$$\mathcal{B}(L) := (B(L), \circ, \mathbb{1}, \setminus, /, \vee, \wedge, \top, \perp, \leq)$$

where the operations are, in terms of closed word sets, given by

$$\begin{array}{ll} A \circ B := (A \cdot B)^{\triangleright\triangleleft}, & A \vee B := (A \cup B)^{\triangleright\triangleleft}, \\ \mathbb{1} := \{\epsilon\}^{\triangleright\triangleleft}, & A \wedge B := A \cap B, \\ B \setminus A := \{v \in \Sigma^* \mid B \cdot \{v\} \subseteq A\}, & \top := \Sigma^*, \\ A/B := \{v \in \Sigma^* \mid \{v\} \cdot B \subseteq A\}, & \perp := \emptyset^{\triangleright\triangleleft}, \\ & A \leq B : \iff A \subseteq B. \end{array}$$

Proof. (Sketch) Monoid properties for \circ and $\mathbb{1}$ follow from Proposition 2:

$$\begin{aligned} (A \circ B) \circ C &= ((AB)^{\triangleright\triangleleft}C)^{\triangleright\triangleleft} = ((AB)^{\triangleright\triangleleft}C^{\triangleright\triangleleft})^{\triangleright\triangleleft} = (ABC)^{\triangleright\triangleleft}, \\ A \circ \mathbb{1} &= (A \cdot \{\epsilon\}^{\triangleright\triangleleft})^{\triangleright\triangleleft} = (A \cdot \{\epsilon\})^{\triangleright\triangleleft} = A. \end{aligned}$$

By proposition 3, we know that A/B is closed when A is. Since the residual laws hold in $\mathcal{P}(\Sigma^*)$ they hold in $B(L)$, because for concepts A, B, C ,

$$A \circ B \subseteq C \iff (A \cdot B)^{\triangleright\triangleleft} \subseteq C \iff A \cdot B \subseteq C. \quad (5)$$

(See also Jipsen e.a.[4], Lemma 7.1.)

The residuations make the syntactic concept lattices complete for the full Lambek calculus, see [7]. Clark emphasizes the lattice structure “as a good search space” for grammatical inference. To learn CFGs, it is sufficient that $(B(L), \vee, \perp, \circ, \mathbb{1})$ is a complete idempotent semiring, in which context-free grammars can be interpreted.

Proposition 4. *The syntactic concepts of L form a complete idempotent semiring*

$$\mathcal{B}(L) = (B(L), +^{\mathcal{B}(L)}, 0^{\mathcal{B}(L)}, \cdot^{\mathcal{B}(L)}, 1^{\mathcal{B}(L)}) := (B(L), \vee, \perp, \circ, \mathbb{1}),$$

and the mapping $h : \mathcal{P}(\Sigma^*) \rightarrow \mathcal{B}(L)$ given by $h(A) := \langle A^{\triangleright\triangleleft}, A^{\triangleright} \rangle$, is a continuous semiring-homomorphism.

A context-free grammar G with constants from Σ is a system of polynomial equations $X_i = p_i(X_1, \dots, X_n)$. Its least solution in $\mathcal{P}(\Sigma^*)$, the tuple of languages $L(G, A)$ for nonterminals A , is componentwise mapped by h to its least solution in $\mathcal{B}(L)$, the tuple of closed sets $L(G, A)^{\triangleright\triangleleft}$. For the main component, we have $h(L) = h(L(G, S)) = L(G, S)^{\triangleright\triangleleft} = L^{\triangleright\triangleleft} = L$.

Remark 1. Since the Kleene-closure A^* is the least solution of $AX + 1 \leq X$ in $\mathcal{P}(\Sigma^*)$ and h preserves least fixed-points, the semiring of syntactic concepts of L can be expanded to a Kleene algebra $(B(L), \vee, \perp, \circ, \mathbb{1}, \otimes)$, using $\langle A, C \rangle^{\otimes} := \langle (A^*)^{\triangleright\triangleleft}, ((C^{\triangleleft})^*)^{\triangleright} \rangle$.

2.2 The lattice $\mathcal{B}(L, F)$ of concepts of L relative to F

We restrict ourselves to context-free grammars in *weak Chomsky Normal Form* (CNF), where rules may have the forms $(C \rightarrow \epsilon)$, $(C \rightarrow a)$, $(C \rightarrow DE)$, where ϵ is the empty word, a is a terminal and C, D, E are nonterminals of the grammar. A motivating idea of Clark was that a finite set V of concepts in the monoidal lattice $B(L)$ gives rise to a grammar $G(L, V)$ that defines a sublanguage of L .

Proposition 5. (c.f. Lemma 1 in [2]) *Let $L \subseteq \Sigma^*$ and $V \subseteq B(L)$ be a finite set of concepts, here viewed as context sets $C \subseteq \Sigma^* \times \Sigma^*$ that are closed, i.e. $C = C^{\diamond}$. Let $G(L, V) = (\Sigma, V, P, S)$ be the grammar with*

$$\begin{aligned} S &:= \{(\epsilon, \epsilon)\}^{\diamond}, \\ P &:= \{(C \rightarrow w) \mid w \in \Sigma \cup \{\epsilon\}, C \in V, w \in C^{\triangleleft}\} \\ &\quad \cup \{(C \rightarrow DE) \mid C, D, E \in V, (D^{\triangleleft}E^{\triangleleft})^{\triangleright\triangleleft} \subseteq C^{\triangleleft}\}. \end{aligned}$$

Then $L(G(L, V)) \subseteq L$.

Proof. By induction, one shows that if $C \Rightarrow^* w$, then $w \in C^{\triangleleft}$, and $L = S^{\triangleleft}$.

In suitable cases, each concept $C = C^{\diamond}$ of V may be generated by a finite subset $C_f \subseteq C$ (c.f. the *diagnostic contexts* in 3.4 of Harris[3]), i.e. $C = C_f^{\diamond}$, and then V is determined by a collection of subsets of a finite set $F \subseteq \Sigma^* \times \Sigma^*$ of contexts. In particular, when L does have a grammar G whose nonterminals A define word sets $L(G, A)$ that are the filler sets C_A^{\triangleleft} of a finite set $C_A \subseteq F$ of contexts, we can hope to find such a grammar G from a finite fragment $E \subseteq L$ that provides each context (u, v) of F through some word $uvw \in E$.

The idea now is to construct from a finite set F of contexts a finite ‘‘approximation’’ $B(L, F)$ of $B(L)$ and define a variant $G(L, F)$ of $G(L, V)$ such that $L(G(L, F)) \subseteq L$. In suitable cases, $G(L, F)$ defines L .

Proposition 6. For sets $F \subseteq \Sigma^* \times \Sigma^*$ of contexts and $A \subseteq \Sigma^*$ of words, let $A^F := A^\triangleright \cap F$. The mappings $\cdot^F : (\mathcal{P}(\Sigma^*), \supseteq) \rightleftharpoons (\mathcal{P}(F), \subseteq) : \cdot^\triangleleft$ form a Galois-connection, i.e. for all $A \in \mathcal{P}(\Sigma^*)$ and $C \in \mathcal{P}(F)$ we have

$$C^\triangleleft \supseteq A \iff C \subseteq A^F.$$

In particular, properties (i) – (iv) of proposition 1 hold with \cdot^F instead of \cdot^\triangleright .

A concept of L relative to F is a pair $\langle A, C \rangle$ of $\mathcal{P}(\Sigma^*) \times \mathcal{P}(F)$ such that $A^F = C$ and $C^\triangleleft = A$. In this case, $A = C^\triangleleft = A^{F^\triangleleft}$ is closed with respect to \cdot^{F^\triangleleft} and $C = A^F = C^{\triangleleft F}$ is closed with respect to $\cdot^{\triangleleft F}$.

Proposition 7. Let $F \subseteq \Sigma^* \times \Sigma^*$. The set of all concepts of L relative to F forms a complete lattice $\mathcal{B}(L, F) = (B(L, F), \vee, \wedge, \top, \perp, \leq)$, where

$$\begin{aligned} \top &:= \langle \Sigma^*, \Sigma^{*F} \rangle & \perp &:= \langle F^\triangleleft, F \rangle \\ \langle A_1, C_1 \rangle \vee \langle A_2, C_2 \rangle &:= \langle (A_1 \cup A_2)^{F^\triangleleft}, C_1 \cap C_2 \rangle \\ \langle A_1, C_1 \rangle \wedge \langle A_2, C_2 \rangle &:= \langle A_1 \cap A_2, (C_1 \cup C_2)^{\triangleleft F} \rangle \\ \langle A_1, C_1 \rangle \leq \langle A_2, C_2 \rangle &: \iff A_1 \subseteq A_2 \wedge C_1 \supseteq C_2. \end{aligned}$$

(Notice that $\emptyset^F = F$ and $\emptyset^\triangleleft = \Sigma^*$.) Define an operation $\circ : B(L, F) \times B(L, F) \rightarrow B(L, F)$ by

$$\langle A_1, C_1 \rangle \circ \langle A_2, C_2 \rangle := \langle (A_1 A_2)^{F^\triangleleft}, (C_1^\triangleleft C_2^\triangleleft)^F \rangle.$$

Indeed, $(A_1 A_2)^{F^\triangleleft F} = (A_1 A_2)^F = (C_1^\triangleleft C_2^\triangleleft)^F$ and $(C_1^\triangleleft C_2^\triangleleft)^{F^\triangleleft} = (A_1 A_2)^{F^\triangleleft}$. Moreover, \circ is monotone with respect to \leq . However, proposition 2 does not extend from $\cdot^{\triangleright\triangleleft}$ to \cdot^{F^\triangleleft} , and \circ is not a monoid operation on $B(L, F)$; nor is $\mathbb{1} = \langle \{\epsilon\}^{F^\triangleleft}, \{\epsilon\}^F \rangle$ neutral with respect to \circ .

Let $\mathcal{P}^{F^\triangleleft}(\Sigma^*)$ be the set of \cdot^{F^\triangleleft} -closed word sets and $\mathcal{P}^{\triangleleft F}(F)$ the set of $\cdot^{\triangleleft F}$ -closed sets of contexts from F . We sometimes use the component functions \circ^{F^\triangleleft} on $\mathcal{P}^{F^\triangleleft}(\Sigma^*)$ and $\circ^{\triangleleft F}$ on $\mathcal{P}^{\triangleleft F}(F)$ of \circ , defined by

$$A_1 \circ^{F^\triangleleft} A_2 := (A_1 A_2)^{F^\triangleleft} \quad \text{and} \quad C_1 \circ^{\triangleleft F} C_2 := (C_1^\triangleleft C_2^\triangleleft)^F.$$

If F is finite, a concept $\langle A, C \rangle$ of L relative to F has a finite representation by its closed context set C .

Lemma 1. If $F \subseteq \Sigma^* \times \Sigma^*$ is finite and $(\epsilon, \epsilon) \in F$, then $L(G(L, F)) \subseteq L$, where $G(L, F)$ is the CNF-grammar (V, Σ, P, S) with

$$\begin{aligned} V &:= \{C \mid \emptyset \neq C \subseteq F, C^{\triangleleft F} = C\}, \\ S &:= \{(\epsilon, \epsilon)\}^{\triangleleft F}, \\ P &:= \{(C \rightarrow w) \mid C \in V, w \in \Sigma \cup \{\epsilon\}, C \subseteq w^F\}, \\ &\quad \cup \{(C \rightarrow DE) \mid C, D, E \in V, C \subseteq (D^\triangleleft E^\triangleleft)^F\}. \end{aligned}$$

Proof. As for proposition 5, we show by induction on the length of derivations that if $C \Rightarrow^* w$, then $w \in C^\triangleleft$. The claim follows from $S^\triangleleft = \{(\epsilon, \epsilon)\}^{\triangleleft F^\triangleleft} =$

$\{(\epsilon, \epsilon)\}^\triangleleft = L$. For a derivation $C \Rightarrow DE \Rightarrow^* uE \Rightarrow^* uv$ with $u, v \in \Sigma^*$, we have, by induction,

$$uv \in D^\triangleleft E^\triangleleft \subseteq (D^\triangleleft E^\triangleleft)^{F^\triangleleft} \subseteq C^\triangleleft,$$

using that \cdot^{F^\triangleleft} is a closure operator on $\mathcal{P}(\Sigma^*)$, that $C \subseteq (D^\triangleleft E^\triangleleft)^F$, and that \cdot^\triangleleft is antitone.

If $G_1 = (V_1, \Sigma, P_1, S_1)$ and $G_2 = (V_2, \Sigma, P_2, S_2)$ are CNF-grammars over Σ , the mapping $\tilde{\cdot} : V_1 \rightarrow V_2$ induces a *grammar homomorphism* from G_1 to G_2 , if $\tilde{S}_1 = S_2$ and $(\tilde{C} \rightarrow a), (\tilde{C} \rightarrow \tilde{D}\tilde{E}) \in P_2$ for all rules $(C \rightarrow a), (C \rightarrow DE) \in P_1$. In this case, clearly $L(G_1) \subseteq L(G_2)$.

We remark that the mapping $C \mapsto C^\heartsuit$ induces a grammar isomorphism from $G(L, F)$ to $G(L, V)$, where $V = \{C^\heartsuit \mid C \in V_F\}$, and V_F is the set of nonterminals of $G(L, F)$. Moreover, if $(\epsilon, \epsilon) \in F_1 \subseteq F_2$, then $C \mapsto C^{\triangleleft F_2}$ induces a grammar homomorphism from $G(L, F_1)$ to $G(L, F_2)$. It follows that $L(G(L, F))$ is monotone in F , but we will not exploit this.

2.3 Grammars with the finite context property

A context-free grammar G has the *finite context property* (FCP), if the $\cdot^{\triangleright\triangleleft}$ -closure of the word set $L(G, A) := \{v \in \Sigma^* \mid A \Rightarrow_G^* v\}$ of every nonterminal A of G can be defined by a finite non-empty¹ set $C_A \subseteq \Sigma^* \times \Sigma^*$ of contexts, i.e.

$$L(G, A)^{\triangleright\triangleleft} = C_A^\triangleleft.$$

The grammar has the finite context property *with respect to the context set* F , if all the above C_A are subsets of F . Clearly, this is monotone in F , and we can replace the $C_A \subseteq F$ by their closures $C_A^{\triangleleft F}$, because $C_A^{\triangleleft F^\triangleleft} = C_A^\triangleleft$,

Lemma 2. *If L has a CNF-grammar G with the FCP with respect to the finite set F of contexts, then $G(L, F)$ contains a homomorphic image of G and $L = L(G(L, F))$.*

Proof. Suppose $G = (V, \Sigma, P, S)$, and for each $A \in V$, suppose $L(G, A)^{\triangleright\triangleleft} = C_A^\triangleleft$ for some $\emptyset \neq C_A \subseteq F$. We can assume that $C_A \in \mathcal{P}^{\triangleleft F}(F)$. Let $G(L, F) = (V(L, F), \Sigma, P(L, F), S(L, F))$. Under $A \mapsto C_A$, each rule of G is mapped to a rule of $G(L, F)$. If $(A \rightarrow BD) \in P$, then $L(G, B)L(G, D) \subseteq L(G, A)$, hence $C_B^\triangleleft C_D^\triangleleft \subseteq C_A^\triangleleft$ using proposition 2, hence $C_A = C_A^{\triangleleft F} \subseteq (C_B^\triangleleft C_D^\triangleleft)^F$, and therefore $(C_A \rightarrow C_B C_D) \in P(L, F)$. If $(A \rightarrow a) \in P$ for $a \in \Sigma \cup \{\epsilon\}$, then $a \in L(G, A) \subseteq C_A^\triangleleft$, so $C_A = C_A^{\triangleleft F} \subseteq a^F$, hence $(C_A \rightarrow a) \in P(L, F)$. Moreover, S is mapped to $S(L, F) = \{(\epsilon, \epsilon)\}^{\triangleleft F}$: since $L = L^{\triangleright\triangleleft} = L(G, S)^{\triangleright\triangleleft} = C_S^\triangleleft = \{(\epsilon, \epsilon)\}^\triangleleft$, we have $C_S = C_S^{\triangleleft F} = \{(\epsilon, \epsilon)\}^{\triangleleft F} = S(L, F)$. Thus $A \mapsto C_A$ induces a grammar homomorphism from G to $G(L, F)$. It follows that $L = L(G) \subseteq L(G(L, F))$. By lemma 1, $L(G(L, F)) \subseteq L$.

¹ Non-emptiness of C_A is not demanded in [2], [8], but is needed for $E \subseteq L$ in lemma 3. The stronger condition $L(G, A) = C_A^\triangleleft$ is used in [2], [5].

The set of contexts derived from $v \in L$ resp. $E \subseteq \Sigma^*$ is

$$\begin{aligned} \text{Con}(v) &:= \{(u, w) \in \Sigma^* \times \Sigma^* \mid \exists \tilde{v} \in \Sigma^* v = u\tilde{v}w\}, \\ \text{Con}(E) &:= \bigcup \{\text{Con}(v) \mid v \in E\}. \end{aligned}$$

Lemma 3. *Suppose $G = (V, \Sigma, P, S)$ is a CNF-grammar without unnecessary nonterminals, and $L = L(G) \neq \emptyset$. If G has the FCP, there are finite sets $E \subseteq L$ and $F \subseteq \text{Con}(E)$ such that G has the FCP with respect to F .*

Proof. For each nonterminal A of G , there is a finite set of contexts $C_A \neq \emptyset$ such that $L(G, A)^{\triangleright\triangleleft} = C_A^{\triangleleft}$. Let $F = \bigcup \{C_A \mid A \in V\}$ be the union of all C_A . Since each $A \in V$ is necessary, $C_A^{\triangleleft} \neq \emptyset$, so there is $v_A \in \Sigma^*$ such that $v_A \in C_A^{\triangleleft}$. Then $C_A \subseteq \text{Con}(C_A \odot v_A)$, and $C_A \odot v_A \subseteq L$ is finite. It follows that $F \subseteq \text{Con}(E)$ for $E = \bigcup \{C_A \odot v_A \mid A \in V\} \subseteq L$.

It follows that if L has a grammar with the FCP, then in order to find one, we can search finite subsets E of L and consider $G(L, F)$ with $F = \text{Con}(E)$. We know that $L(G(L, F))$ is a subset of L , and equals L when F is large. However, to construct $G(L, F)$ we must avoid computing infinite filler sets C^{\triangleleft} in order to find the closed sets $C = C^{\triangleleft F}$. We need a truly finite representation of $G(L, F)$.

3 Lattices $\mathcal{B}(K, L, F)$ of relativized concepts of L

Let $K \subseteq \Sigma^*$ and $F \subseteq \Sigma^* \times \Sigma^*$ be word- and context sets. Put

$$A^F := A^{\triangleright} \cap F, \quad C^K := C^{\triangleleft} \cap K, \quad A^{FK} := (A^F)^K, \quad C^{KF} := (C^K)^F.$$

Then A^F and C^K are monotone in F and K , but antitone in A and C .

Proposition 8. *The mappings $\cdot^F : (\mathcal{P}(K), \supseteq) \rightleftarrows (\mathcal{P}(F), \subseteq) : \cdot^K$ form a Galois-connection, i.e. for all $A \in \mathcal{P}(K)$ and $C \in \mathcal{P}(F)$ we have*

$$C^K \supseteq A \iff C \subseteq A^F.$$

In particular,

- (i) \cdot^F and \cdot^K are antitone: $A^F \supseteq B^F$ for all $A \subseteq B \subseteq K$ and $C^K \supseteq D^K$ for all $C \subseteq D \subseteq F$,
- (ii) $A^{FKF} = A^F$ and $C^{KFK} = C^K$ for all $A \subseteq K$ and $C \subseteq F$.
- (iii) \cdot^{FK} resp. \cdot^{KF} is a closure operator on $(\mathcal{P}(K), \subseteq)$ resp. $(\mathcal{P}(F), \subseteq)$.
- (iv) $\cdot^F : (\{C^K \mid C \subseteq F\}, \supseteq) \rightleftarrows (\{A^F \mid A \subseteq K\}, \subseteq) : \cdot^K$ form an order isomorphism.

We call a pair $\langle A, C \rangle$ such that $A^F = C$ and $C^K = A$ a *relativized concept of L* . Let $\mathcal{B}(K, L, F)$ be the set of all relativized concepts of L . The components of a relativized concept $\langle A, C \rangle$ are closed with respect to \cdot^{FK} and \cdot^{KF} , respectively,

as $A = A^{FK}$ and $C = C^{KF}$, and via (ii) and (iv) one can represent a relativized concept of L by its closed word set or its closed context set. Writing

$$\mathcal{P}^{FK}(K) := \{A \subseteq K \mid A = A^{FK}\} \quad \text{and} \quad \mathcal{P}^{KF}(F) := \{C \subseteq F \mid C = C^{KF}\}$$

for the set of closed members of $\mathcal{P}(K)$ and $\mathcal{P}(F)$, respectively, (iv) gives an order isomorphism

$$\cdot^F : (\mathcal{P}^{FK}(K), \supseteq) \rightleftharpoons (\mathcal{P}^{KF}(F), \subseteq) : \cdot^K .$$

Proposition 9. *Let $K \subseteq \Sigma^*$ and $F \subseteq \Sigma^* \times \Sigma^*$. The set of all relativized concepts of L forms a complete lattice $\mathcal{B}(K, L, F) = (B(K, L, F), \vee, \wedge, \top, \perp, \leq)$ with the following operations:*

$$\begin{aligned} \top &:= \langle K, K^F \rangle & \perp &:= \langle \emptyset^{FK}, F \rangle \\ \langle A_1, C_1 \rangle \vee \langle A_2, C_2 \rangle &:= \langle (A_1 \cup A_2)^{FK}, C_1 \cap C_2 \rangle \\ \langle A_1, C_1 \rangle \wedge \langle A_2, C_2 \rangle &:= \langle A_1 \cap A_2, (C_1 \cup C_2)^{KF} \rangle \\ \langle A_1, C_1 \rangle \leq \langle A_2, C_2 \rangle &: \iff A_1 \subseteq A_2 \wedge C_1 \supseteq C_2. \end{aligned}$$

When $K = \Sigma^*$, we have $C^K = C^\triangleleft$ for all $C \subseteq F$, and $\mathcal{B}(\Sigma^*, L, F) = \mathcal{B}(L, F)$. When F is finite, we want to use $\mathcal{B}(L, F)$ as a finite ‘‘approximation’’ of the generally infinite $\mathcal{B}(L)$, and when K is also finite, $\mathcal{B}(K, L, F)$ is a finite, effective approximation of $\mathcal{B}(L, F)$. To relativize to $B(K, L, F)$ the monoid operation \circ of $\mathcal{B}(L)$ with its component functions $\circ^{\triangleright\triangleleft}$ and $\circ^{\triangleleft\triangleright}$,

$$\langle A_1, C_1 \rangle \circ \langle A_2, C_2 \rangle = \langle A_1 \circ^{\triangleright\triangleleft} A_2, C_1 \circ^{\triangleleft\triangleright} C_2 \rangle := \langle (A_1 A_2)^{\triangleright\triangleleft}, (C_1^\triangleleft C_2^\triangleleft)^{\triangleright} \rangle,$$

Clark [1] (Def. 7) defines a *partial* operation \circ on $\mathcal{B}(K, L, F)$ by

$$\langle A_1, C_1 \rangle \circ \langle A_2, C_2 \rangle := \langle A_1 \circ^{FK} A_2, C_1 \circ^{KF} C_2 \rangle := \langle (A_1 A_2)^{FK}, (C_1^K C_2^K)^F \rangle,$$

which need not be a monoid operation. (Proposition 2 for $\cdot^{\triangleright\triangleleft}$ does not extend to all \cdot^{FK} .) It is only a partial operation on $\mathcal{B}(K, L, F)$, because although $(A_1 A_2)^{FK}$ is closed in $\mathcal{P}(K)$ and the \cdot^K -image of $(C_1^K C_2^K)^F$, the latter need not belong to $\mathcal{P}^{KF}(F)$: $C_1^K C_2^K$ need not be a subset of K , whence its \cdot^F -image need not be \cdot^{KF} -closed in $\mathcal{P}(F)$.

Moreover, the embedding $\langle A, C \rangle \mapsto \langle C^\triangleleft, C \rangle$ from $B(K, L, F)$ to $B(L, F)$ need not preserve \circ . However, if K is large enough, everything is fine:

Lemma 4. *(cf. Clark[1], Lemma 6): For any $L \subseteq \Sigma^*$ and finite $F \subseteq \Sigma^* \times \Sigma^*$, there is a finite $K \subseteq \Sigma^*$ such that*

$$(B(K, L, F), \vee, \wedge, \perp, \top, \leq, \circ) \simeq (B(L, F), \vee, \wedge, \top, \perp, \leq, \circ).$$

We detail the four line proof sketch from [1] for later reference.

Proof. First note that if $C \subseteq F$ is closed in $\mathcal{P}^{KF}(F)$, it is closed in $\mathcal{P}^{\triangleleft F}(F)$ as well: if $C = C^{KF}$, then since $C^\triangleleft \supseteq C^K$ gives $C^{\triangleleft F} \subseteq C^{KF} = C$, so $C = C^{\triangleleft F}$

as $\cdot^{\triangleleft F}$ is a closure operator on $\mathcal{P}(F)$. Therefore $\langle A, C \rangle \mapsto \langle A^{F\triangleleft}, C \rangle = \langle C^{\triangleleft}, C \rangle$ embeds $B(K, L, F)$ into $B(L, F)$.

To make this embedding be onto $B(L, F)$, we have to choose K sufficiently large. Recall that C^{KF} is antitone in K , and $C^{\triangleleft F} \subseteq C^{KF}$ for any $C \subseteq F$ and any K . If $C^{KF} \not\subseteq C^{\triangleleft F}$, there is $(u, v) \in C^{KF} \setminus C^{\triangleleft F}$, and hence some $w \in C^{\triangleleft} \setminus C^K$ such that $uwv \not\subseteq L$, hence $(u, v) \notin C^{(K \cup \{w\})^F}$. By adding at most $|C^{KF} \setminus C^{\triangleleft F}|$ many elements from C^{\triangleleft} to K we obtain $K' \supseteq K$ such that $C^{\triangleleft F} = C^{K'F} \subseteq C^{KF}$. Since $\mathcal{P}(F)$ is finite, there is thus a finite set $K \subseteq \Sigma^*$ such that

$$C^{KF} = C^{\triangleleft F}, \quad \text{for all } C \subseteq F. \quad (6)$$

It follows that $\mapsto: (B(K, L, F), \vee, \wedge, \perp, \top, \leq) \simeq (B(L, F), \vee, \wedge, \perp, \top, \leq)$, and that $A^{FKF} = A^{F\triangleleft F} = A^F$ for any $A \subseteq \Sigma^*$ (not just for $A \subseteq K$). In particular, \circ is total on $B(K, L, F)$, as for each $\langle A_1, C_1 \rangle, \langle A_2, C_2 \rangle \in B(K, L, F)$ we get

$$\langle A_1, C_1 \rangle \circ \langle A_2, C_2 \rangle = \langle (A_1 A_2)^{FK}, (C_1^K C_2^K)^F \rangle = \langle (A_1 A_2)^{FK}, (C_1^K C_2^K)^{FKF} \rangle.$$

Moreover, by similar means we can achieve that

$$(C_1^K C_2^K)^F = (C_1^{\triangleleft} C_2^{\triangleleft})^F, \quad \text{for all } C_1, C_2 \subseteq F, \quad (7)$$

so that \mapsto preserves \circ since for each $\langle A_1, C_1 \rangle, \langle A_2, C_2 \rangle \in B(K, L, F)$, we get

$$\begin{aligned} \langle A_1, C_1 \rangle \circ^{\mathcal{B}(K, L, F)} \langle A_2, C_2 \rangle &= \langle (A_1 A_2)^{FK}, (C_1^K C_2^K)^F \rangle \\ \mapsto \langle (C_1^{\triangleleft} C_2^{\triangleleft})^{F\triangleleft}, (C_1^{\triangleleft} C_2^{\triangleleft})^F \rangle &= \langle C_1^{\triangleleft}, C_1 \rangle \circ^{\mathcal{B}(L, F)} \langle C_2^{\triangleleft}, C_2 \rangle. \end{aligned}$$

To see (7), notice that $(C_1^K C_2^K)^F \supseteq (C_1^{\triangleleft} C_2^{\triangleleft})^F$ for any K , and if we have \neq here, there are $u \in C_1^{\triangleleft} \setminus C_1^K$ and $v \in C_2^{\triangleleft} \setminus C_2^K$. Adding these to K lets $(C_1^K C_2^K)^F$ shrink strictly. By adding at most $|(C_1^K C_2^K)^F \setminus (C_1^{\triangleleft} C_2^{\triangleleft})^F|$ many elements from $C_1^{\triangleleft} \cup C_2^{\triangleleft}$ to K we obtain $K' \supseteq K$ such that $(C_1^{K'} C_2^{K'})^F = (C_1^{\triangleleft} C_2^{\triangleleft})^F$. Since $\mathcal{P}(F)$ is finite, we can do this for all $C_1, C_2 \subseteq F$, and achieve (7).

The monotonicity properties ensure that (6) and (7) and hence $\mathcal{B}(K, L, F) \simeq \mathcal{B}(L, F)$ are preserved under extensions of K .

Corollary 1. *For any finite F where \emptyset is not $\cdot^{\triangleleft F}$ -closed, there is a finite $K \subseteq \text{Sub}(L)$ such that*

$$(B(K, L, F), \vee, \wedge, \perp, \top, \leq, \circ) \simeq (B(L, F), \vee, \wedge, \top, \perp, \leq, \circ).$$

Proof. For $\emptyset \neq C \subseteq F$ we have $C^{\triangleleft} \subseteq \text{Sub}(L)$. To achieve $C^{\triangleleft F} = C^{KF}$ and $(C_1^K C_2^K)^F = (C_1^{\triangleleft} C_2^{\triangleleft})^F$, only elements from C^{\triangleleft} and C_i^{\triangleleft} are added to K in the above proof, so we can do with some $K \subseteq \text{Sub}(L)$, provided $\emptyset \neq \emptyset^{\triangleleft F}$.

If \emptyset is $\cdot^{\triangleleft F}$ -closed, as it often happens, there may be no finite $K \subseteq \text{Sub}(L)$ with $\emptyset^{KF} = \emptyset^{\triangleleft F} = \emptyset$. For example, when $K \subseteq \text{Sub}(L) = L \neq \Sigma^*$ and $F = \{(\epsilon, \epsilon)\}$, then $\emptyset^{\triangleleft F} = (\Sigma^*)^F = \emptyset$, but $\emptyset^{KF} = K^F = F$. Also, when $K \subseteq \text{Sub}(L)$, it may be impossible that \circ^{KF} is total on $\mathcal{P}^{KF}(F)$, i.e. there may be $C_1, C_2 \in \mathcal{P}^{KF}(F)$ such that $C_1 \circ^{KF} C_2 = (C_1^K C_2^K)^F = \emptyset \neq \emptyset^{KF}$.

3.1 Clark's learning algorithm

For finite sets $K \subseteq \Sigma^*$ and $F \subseteq \Sigma^* \times \Sigma^*$ with $(\epsilon, \epsilon) \in F$, Clark relativized $G(L, V)$ to a grammar $G(K, L, F) = (V, \Sigma, P, S)$ where²

$$\begin{aligned} V &:= \{C \mid C \subseteq F, C = C^{KF}\}, \\ S &:= \{(\epsilon, \epsilon)\}^{KF}, \\ P &:= \{(C \rightarrow w) \mid C \in V, w \in \Sigma \cup \{\epsilon\}, w \in C^K\} \\ &\quad \cup \{(C \rightarrow DE) \mid C, D, E \in V, (D^K E^K)^{FK} \subseteq C^K\}. \end{aligned}$$

He then shows (Lemma 2,3,4 in [2]) that $L(G(K, L, F))$ depends monotonically on F and antitonically on $K \supseteq \Sigma \cup \{\epsilon\}$. Theorem 1 in [2] claims that the following algorithm identifies L in the limit, i.e. that for any oracles T and χ_L , $\langle G_n \mid n \in \mathbb{N} \rangle$ gets constant at some G_n such that $L(G_n) = L$.

Let Σ be a finite alphabet and $L \subseteq \Sigma^*$, a language with a CNF-grammar with the finite context property, be given by oracles $\chi_L : \Sigma^* \rightarrow \mathbb{B}$ and $T : \mathbb{N} \rightarrow \Sigma^*$ for recognition and enumeration of L .

Produce a sequence $\langle G_n \mid n \in \mathbb{N} \rangle$ of CNF-grammars, where $w_n = T(n)$:

$$\begin{aligned} E_0 &:= \emptyset, & E_{n+1} &:= E_n \cup \{w_n\}, \\ K_0 &:= \Sigma \cup \{\epsilon\}, & K_{n+1} &:= \begin{cases} \text{Sub}(E_{n+1}), & \text{if } E_{n+1} \not\subseteq L(G_n), \text{ or} \\ G(\text{Sub}(E_{n+1}), L, F_n) \neq G_n & \\ K_n, & \text{else} \end{cases} \\ F_0 &:= \{(\epsilon, \epsilon)\}, & F_{n+1} &:= \begin{cases} \text{Con}(E_{n+1}), & \text{if } E_{n+1} \not\subseteq L(G_n), \\ F_n, & \text{else} \end{cases} \\ G_0 &:= G(K_0, L, F_0). & G_{n+1} &:= G(K_{n+1}, L, F_{n+1}). \end{aligned}$$

Table 1. Clark's grammar inference algorithm

Why should $\langle G_n \mid n \in \mathbb{N} \rangle$ converge to a grammar for L ? Call a finite set F of contexts *adequate for L* if $L \subseteq L(G(K, L, F))$ for every finite $\Sigma \cup \{\epsilon\} \subseteq K \subseteq \Sigma^*$. By lemma 3, there is some F such that $L \subseteq L(G(L, F))$. Then any $F_n \supseteq F$ is adequate for L , since $G(L, F) = G(\Sigma^*, L, F)$, and $L(G(K, L, F))$ is monotone in F and antitone in $K \subseteq \Sigma^*$. So one would first like to show:

$\langle F_n \mid n \in \mathbb{N} \rangle$ gets constant in some F_n that is adequate for L ,

so that $L \subseteq L(G(K_m, L, F_n))$ for all m , and then show:

$$\langle K_n \mid n \in \mathbb{N} \rangle \text{ gets constant in some } K_m \text{ such that } L = L(G(K_m, L, F_n)). \quad (8)$$

² We omit a size bound f on a generating subset for the C^{KF} in the definition of V , which only serves to bound $|G(K, L, F)|$ by a polynomial in $|K|$ and $|F|$.

To achieve (8), Lemma 5 in [2] claims:

$$\text{For any } L \text{ and } F, \text{ there is } K \subseteq \Sigma^* \text{ such that } L(G(K, L, F)) \subseteq L. \quad (9)$$

The proof sketched in [2] only works for the infinite set $K = \Sigma^*$, but the claim is needed with finite $K \subseteq \text{Sub}(L)$ in the grammar inference algorithm. However, this strengthening of (9) is wrong:

Example 1. There is a language $L \subseteq \Sigma^*$ and a finite set of contexts F with

$$L(G(K, L, F)) \not\subseteq L, \quad \text{for all finite } K \text{ such that } \Sigma \cup \{\epsilon\} \subseteq K \subseteq \text{Sub}(L).$$

Let $\Sigma = \{a\}$, $F = \{\epsilon, \epsilon\}$, $L = \{\epsilon, a\}$. The only finite set K with $\text{Sub}(L) \supseteq K \supseteq \Sigma \cup \{\epsilon\}$ is $K = L$.

- (i) The set V of closed elements of $\mathcal{P}(F)$ contains F , since $F \subseteq F^{KK}$ is maximal in $\mathcal{P}(F)$. Since $F^K = \{k \in K \mid \epsilon k \epsilon \in L\} = K \cap L = L$ and $a \in L$, $(F \rightarrow a)$ is a non-branching rule of $G(K, L, F)$.
- (ii) To see that $(F \rightarrow FF)$ is a branching rule of $G(K, L, F)$, notice that since $LL \not\subseteq L$, we have $(F^K F^K)^F = (LL)^F = \emptyset$, so $(F^K F^K)^{FK} = \emptyset^K = K = L = F^K$.

Since F is the start symbol of $G(K, L, F)$, it follows that $aa \in L(G(K, L, F)) \setminus L$. This grammar, extended by the rule $F \rightarrow \epsilon$, is G_0 , so $L \subset L(G_0)$ and $K_1 = K_0$, $F_1 = F_0$, and $G_1 = G(K_1, L, F_1) = G_0$. By induction, $\langle G_n \mid n \in \mathbb{N} \rangle$ gets constant in G_0 , a grammar that does *not* define L .

3.2 Correcting Clark's algorithm

Example 1 shows that Clark's condition $(D^K E^K)^{FK} \subseteq C^K$ in branching rules of $G(K, L, F)$ is too permissive. For concepts C, D, E , being a branching rule $(C \rightarrow DE)$ of $G(L, V)$ amounts to any of the equivalent conditions $D^\triangleleft E^\triangleleft \subseteq C^\triangleleft$, $(D^\triangleleft E^\triangleleft)^{\triangleright\triangleleft} \subseteq C^\triangleleft$ or $C \subseteq (D^\triangleleft E^\triangleleft)^\triangleright$. These are no longer equivalent in the relativized situation; for concepts C, D, E relativized to K, F we only have

$$D^K E^K \subseteq C^K \implies C \subseteq (D^K E^K)^F \implies (D^K E^K)^{FK} \subseteq C^K.$$

Though \cdot^{FK} is a closure operation on $\mathcal{P}(K)$, we may have $D^K E^K \not\subseteq (D^K E^K)^{FK}$ when $D^K E^K \not\subseteq K$. We modify the definition of $G(K, L, F)$ by replacing the condition $(D^K E^K)^{FK} \subseteq C^K$ by the stronger condition³ $C \subseteq (D^K E^K)^F$. Moreover, we exclude the empty context set from the nonterminals.

Definition 1. Let $K, L \subseteq \Sigma^*$ be arbitrary sets of words, and $F \subseteq \Sigma^* \times \Sigma^*$ a finite set of contexts such that $(\epsilon, \epsilon) \in F$. Then $G(K, L, F)$ is the binary grammar

³ It excludes $(F \rightarrow FF)$ in example 1, where $F \not\subseteq (F^K F^K)^F = \emptyset$. The weaker $C \subseteq (D^K E^K)^{FK}$ is equivalent to Clark's $(D^K E^K)^{FK} \subseteq C^K$.

(V, Σ, P, S) where

$$\begin{aligned} V &:= \{C \mid \emptyset \neq C \subseteq F, C^{KF} = C\} \\ S &:= \{(\epsilon, \epsilon)\}^{KF}, \\ P &:= \{(C \rightarrow w) \mid C \in V, w \in \Sigma \cup \{\epsilon\}, C \subseteq w^F\} \\ &\quad \cup \{(C \rightarrow DE) \mid C, D, E \in V, C \subseteq (D^K E^K)^F\}. \end{aligned}$$

Note that for $C = C^{KF}$, the condition $C \subseteq w^F$ is equivalent to $w \in C^K$. When K and F are finite, we can determine V , S and P from a decision algorithm for membership in L . To determine V , we need to know $(F \odot K) \cap L$, and to determine P , we need to know $(F \odot KK) \cap L$.

Our conditions $C \subseteq w^F$ for non-branching rules $(C \rightarrow w)$ and $C \subseteq (D^K E^K)^F$ for branching rules $(C \rightarrow DE)$ of $G(K, L, F)$ are monotone in F and antitone in K . Yet, as pointed out by R. Yoshinaka⁴, $L(G(K, L, F))$ is neither monotone in F nor antitone in K . The reason is that for $F_1 \subseteq F_2$, say, the set V_1 of nonterminals of $G(K, L, F_1)$ is not a subset of the set V_2 of nonterminals of $G(K, L, F_2)$, and the embedding $\tilde{\cdot} : V_1 \rightarrow V_2$ given by $\tilde{C} := C^{KF_2}$ gives $C \subseteq \tilde{C}$ and hence can lead from $C \subseteq (D^K E^K)^{F_1}$ to $\tilde{C} \not\subseteq (\tilde{D}^K \tilde{E}^K)^{F_2}$ and does not induce a grammar homomorphism.

Before presenting a correction of Clark's algorithm, let us recapitulate his idea. A finite amount of positive information $E \subseteq L$ about L gives a finite set $F = \text{Con}(E)$ of contexts and, through $\mathcal{B}(L, F)$, a grammar $G(L, F)$ of a sublanguage of L . If F is big enough, $G(L, F)$ defines L . Each F and each finite $K \subseteq \Sigma^*$ provide, through the finite $\mathcal{B}(K, L, F)$, a grammar $G(K, L, F)$ whose language is monotone in F and antitone in K . If F is big enough, $L \subseteq L(G(K, L, F))$, and if K is big enough, one has $\mathcal{B}(L, F) \simeq \mathcal{B}(K, L, F)$, in which case $G(K, L, F)$ is $G(L, F)$ and defines L . Thus, when $E \not\subseteq L(G(K, L, \text{Con}(E)))$, one needs to increase E , and otherwise one should keep $F = \text{Con}(E)$ fixed and increase K to make it big enough.

Since the input to the inference process consists of positive information about L only, we cannot use lemma 4 directly to get $\mathcal{B}(K, L, F) \simeq \mathcal{B}(L, F)$, but need a refinement with $K \subseteq \text{Sub}(L)$ instead of $K \subseteq \Sigma^*$. (We get no clue on which $K \not\subseteq \text{Sub}(L)$ would satisfy $\emptyset = \emptyset^{\triangleleft F} = \emptyset^{KF}$, so we exclude \emptyset resp. \top from the nonterminals.) Moreover, it will in general be impossible to define L with $G(K, L, F)$ where $K = \text{Sub}(E)$ and $F = \text{Con}(E)$ for some finite $E \subseteq L$; we may need a finite K with $\text{Sub}(E) \subset K \subseteq \text{Sub}(L)$. So we have to switch between increasing F on the one hand and increasing K while keeping F fixed on the other. Finally, of course we cannot explicitly test whether F is big enough so that $L = L(G(L, F))$, or whether K gives $\mathcal{B}(K, L, F) \simeq \mathcal{B}(L, F)$. We need computable substitutes for such tests.

We say that $\langle \mathcal{P}^{KF}(F), \circ^{KF} \rangle$ and $\langle \mathcal{P}^{\triangleleft F}(F), \circ^{\triangleleft F} \rangle$ *almost agree*, in symbols: $\langle \mathcal{P}^{KF}(F), \circ^{KF} \rangle \equiv \langle \mathcal{P}^{\triangleleft F}(F), \circ^{\triangleleft F} \rangle$, if for all non-empty $C, C_1, C_2 \subseteq F$,

$$C^{KF} = C^{\triangleleft F} \quad \text{and} \quad C_1 \circ^{KF} C_2 = C_1 \circ^{\triangleleft F} C_2.$$

⁴ Personal communication, February 2013

Like (6) and (7) in the proof of lemma 4, this property is monotone in K ; but now we only consider non-empty subsets of F .

Lemma 5. *Let $F \subseteq \Sigma^* \times \Sigma^*$ and $K \subseteq \text{Sub}(L)$ be finite, and $(\epsilon, \epsilon) \in F$.*

- (i) *There is a finite $K \subseteq \tilde{K} \subseteq \text{Sub}(L)$ with $\langle \mathcal{P}^{\tilde{K}F}(F), \circ^{\tilde{K}F} \rangle \equiv \langle \mathcal{P}^{\triangleleft F}(F), \circ^{\triangleleft F} \rangle$.*
- (ii) *If $\langle \mathcal{P}^{KF}(F), \circ^{KF} \rangle \equiv \langle \mathcal{P}^{\triangleleft F}(F), \circ^{\triangleleft F} \rangle$, then $G(K, L, F) = G(L, F)$.*

Proof. (i) We need to satisfy the restrictions of (6) and (7) to non-empty sets $C, C_1, C_2 \subseteq F$. As the proof of lemma 4 shows, we have to extend K by elements of $C^{\triangleleft}, C_1^{\triangleleft}, C_2^{\triangleleft}$, and these are subsets of $\text{Sub}(L)$ when C, C_1, C_2 are non-empty.

(ii) Since $\langle \mathcal{P}^{KF}(F), \circ^{KF} \rangle$ and $\langle \mathcal{P}^{\triangleleft F}(F), \circ^{\triangleleft F} \rangle$ almost agree, $G(K, L, F)$ and $G(L, F)$ have the same nonterminals, start symbols, and non-branching rules and branching rules.

Although $\langle \mathcal{P}^{\triangleleft F}(F), \supseteq, \circ^{\triangleleft F} \rangle$ is a finite structure, we generally cannot compute it, given an oracle for membership in L , because possibly infinite word sets C^{\triangleleft} are involved. So we cannot test whether $\langle \mathcal{P}^{KF}(F), \circ^{KF} \rangle$ almost agrees with $\langle \mathcal{P}^{\triangleleft F}(F), \circ^{\triangleleft F} \rangle$. Nor can we test whether $G(K, L, F)$ equals $G(L, F)$. But we *can* test the following weaker property, which however is *not* monotone in K .

We say \circ^{KF} is almost total on $\mathcal{P}^{KF}(F)$, if for all non-empty $C_1, C_2 \in \mathcal{P}^{KF}(F)$, $C_1 \circ^{KF} C_2$ belongs to $\mathcal{P}^{KF}(F) \cup \{\emptyset\}$. Using an oracle for membership in L one can check whether \circ^{KF} is almost total.

Proposition 10. *The following conditions are equivalent:*

- (i) *For all non-empty $C \subseteq F$, $C^{KF} = C^{\triangleleft F}$.*
- (ii) *$\mathcal{P}^{KF}(F) \setminus \{\emptyset\} = \mathcal{P}^{\triangleleft F}(F) \setminus \{\emptyset\}$.*

Proof. (i) \Rightarrow (ii): We have $\mathcal{P}^{KF}(F) \subseteq \mathcal{P}^{\triangleleft F}(F)$, since for all $C \subseteq F$, we have $C^K \subseteq C^{\triangleleft}$, hence $C \subseteq C^{\triangleleft F} \subseteq C^{KF}$, and when $C \in \mathcal{P}^{KF}(F)$, we also have $C^{KF} \subseteq C$. Now suppose $C \in \mathcal{P}^{\triangleleft F}(F) \setminus \{\emptyset\}$. Then $\emptyset \neq C = C^{\triangleleft F}$, so by (i) $C^{KF} = C^{\triangleleft F} = C$, whence $C \in \mathcal{P}^{KF}(F) \setminus \{\emptyset\}$.

(ii) \Rightarrow (i): Assume $\emptyset \neq C \subseteq F$. Then by (ii), $\emptyset \neq C^{\triangleleft F} \in \mathcal{P}^{\triangleleft F}(F) \subseteq \mathcal{P}^{KF}(F)$, so $C^{KF} \subseteq (C^{\triangleleft F})^{KF} = C^{\triangleleft F} \subseteq C^{KF}$.

Proposition 11. *Let $F \subseteq \Sigma^* \times \Sigma^*$ and $K \subseteq \Sigma^*$. If $\langle \mathcal{P}^{KF}(F), \circ^{KF} \rangle$ and $\langle \mathcal{P}^{\triangleleft F}(F), \circ^{\triangleleft F} \rangle$ almost agree, then \circ^{KF} is almost total on $\mathcal{P}^{KF}(F)$.*

Proof. Suppose $C_1, C_2 \in \mathcal{P}^{KF}(F)$ are nonempty, and $(C_1 \circ^{KF} C_2) \neq \emptyset$. Then

$$C_1 \circ^{KF} C_2 = C_1 \circ^{\triangleleft F} C_2 = (C_1^{\triangleleft} C_2^{\triangleleft})^F = (C_1^{\triangleleft} C_2^{\triangleleft})^{F \triangleleft F} = (C_1^{\triangleleft} C_2^{\triangleleft})^{FKF} = (C_1 \circ^{KF} C_2)^{KF},$$

and so $(C_1 \circ^{KF} C_2) \in \mathcal{P}^{KF}(F)$.

Lemma 6. *Suppose $F \subseteq \Sigma^* \times \Sigma^*$ is finite with $(\epsilon, \epsilon) \in F$ and $L \subseteq \Sigma^*$ has a CNF-grammar G with the FCP with respect to F . If $\Sigma \cup \{\epsilon\} \subseteq K \subseteq \Sigma^*$ is finite and \circ^{KF} is almost total on $\mathcal{P}^{KF}(F)$, then $L \subseteq L(G(K, L, F))$.*

Proof. By assumption, for each nonterminal A of G there is a finite set $\emptyset \neq C_A \subseteq F$ with $L(G, A)^{\triangleright\triangleleft} = C_A^{\triangleleft}$. We show that $A \mapsto C_A^{KF}$ induces a grammar homomorphism from G to $G(K, L, F)$. For each nonterminal A of G , $\emptyset \neq C_A^{KF}$ is \cdot^{KF} -closed, hence a nonterminal of $G(K, L, F)$. Let $(A \rightarrow DE)$ be a rule of G , so $C_A^{\triangleleft} \supseteq C_D^{\triangleleft} C_E^{\triangleleft}$, using proposition 2. Then

$$\emptyset \neq C_A \subseteq C_A^{\triangleleft F} \subseteq (C_D^{\triangleleft} C_E^{\triangleleft})^F \subseteq (C_D^K C_E^K)^F = (C_D^{KFK} C_E^{KFK})^F = C_D^{KF} \circ^{KF} C_E^{KF}.$$

By monotonicity of \cdot^{KF} and since \circ^{KF} is almost total on $\mathcal{P}^{KF}(F)$, we get

$$C_A^{KF} \subseteq (C_D^{KF} \circ^{KF} C_E^{KF})^{KF} \subseteq C_D^{KF} \circ^{KF} C_E^{KF},$$

and so $(C_A^{KF} \rightarrow C_D^{KF} C_E^{KF})$ is a rule of $G(K, L, F)$. Likewise, let $(A \rightarrow a)$ be a rule of G . Then $a \in L(G, A) \cap K \subseteq C_A^{\triangleleft} \cap K = C_A^K$, so $C_A^{KF} \subseteq a^F$, and $(C_A^{KF} \rightarrow a)$ is a rule of $G(K, L, F)$. Hence, if $A \Rightarrow_G^* w$, then $C_A^{KF} \Rightarrow_{G(K, L, F)}^* w$. We may assume that $C_S = \{(\epsilon, \epsilon)\}$, as $C_S^{\triangleleft} = L = \{(\epsilon, \epsilon)\}^{\triangleleft} = \{(\epsilon, \epsilon)\}^{\triangleleft F \triangleleft}$ since $(\epsilon, \epsilon) \in F$. Then $C_S^{KF} = \{(\epsilon, \epsilon)\}^{KF}$ is the start symbol of $G(K, L, F)$, and we have $L = L(G) \subseteq L(G(K, L, F))$.

The idea for the corrected grammar inference is as follows. Start with $F = \{(\epsilon, \epsilon)\}$ and consume increasing finite subsets E of L until $K = \text{Sub}(E)$ makes \circ^{KF} almost total. We find such a K by lemma 5 and proposition 11. If then $E \not\subseteq L(G(K, L, F))$, we know by lemma 6 that L does not have a grammar with the FCP with respect to F (i.e. “ F is not adequate for L ”). So we update F to $\text{Con}(E)$ and repeat this process, until we have $K \subseteq \text{Sub}(L)$ where \circ^{KF} is almost total and $E \subseteq L(G(K, L, F))$. Then we keep F fixed and increase E and K until \circ^{KF} is almost total, and check if $E \subseteq L(G(K, L, F))$. If we no more run into the case $E \not\subseteq L(G(K, L, F))$ where F is increased, we exhaust the finite subsets K of $\text{Sub}(L)$ and hence reach $G(K, L, F) = G(L, F)$ by lemma 5; since this is monotone in K , the grammar $G(L, F)$ is the limit grammar. Then on the one hand, $E \subseteq L(G(K, L, F)) = G(L, F)$ for all finite $E \subseteq L$, hence $L \subseteq L(G(L, F))$, and on the other hand $L(G(L, F)) \subseteq L$.

Since L has a grammar with the FCP, after finitely many updates of $F = \text{Con}(E)$ it has a grammar with the FCP with respect to F . Then by lemma 6, $L \subseteq L(G(K, L, F))$ for all K where \circ^{KF} is almost total, and so we do not run into the case $E \not\subseteq L(G(K, L, F))$ any more.

Theorem 2. *If $\emptyset \neq L \subseteq \Sigma^*$ has a CNF-grammar with the finite context property, then the algorithm of table 2 identifies L in the limit, i.e. for the sequence $\langle G_n \mid n \in \mathbb{N} \rangle$ of grammars produced for any membership oracle and enumeration of L , there is m such that $L(G_m) = L$ and $G_n = G_m$ for all $n \geq m$.*

Proof. Let G be a grammar for L with the FCP. We may assume that G has no unnecessary symbols. By lemma 3, G has the FCP with respect to some $F \subseteq \text{Con}(E)$ for some finite $E \subseteq L$.

Claim 1 The sequence $\langle F_n \mid n \in \mathbb{N} \rangle$ gets constant at a finite $F_{\bar{n}} \subseteq \text{Con}(L)$.

Proof of Claim 1: By induction, $F_n \subseteq F_{n+1} \subseteq \text{Con}(E_{n+1})$ for all n . Assume that $\langle F_n \mid n \in \mathbb{N} \rangle$ does not get constant. Then neither do $\langle E_n \mid n \in \mathbb{N} \rangle$ nor $\langle K_n \mid$

Let Σ be a finite alphabet and $\emptyset \neq L \subseteq \Sigma^*$, a language with a CNF-grammar with the finite context property, be given by oracles $\chi_L : \Sigma^* \rightarrow \mathbb{B}$ and $T : \mathbb{N} \rightarrow \Sigma^*$ for recognition and enumeration of L .

Produce a sequence $\langle G_n \mid n \in \mathbb{N} \rangle$ of CNF-grammars, where $w_n = T(n)$:

$$\begin{array}{ll}
E_0 := \emptyset, & E_{n+1} := E_n \cup \{w_n\}, \\
K_0 := \Sigma \cup \{\epsilon\}, & K_{n+1} := \text{Sub}(E_{n+1}) \cup K_0, \\
F_0 := \{(\epsilon, \epsilon)\}, & G_{n+1} := \begin{cases} G(K_{n+1}, L, F_n) & \text{if } \circ^{K_{n+1}F_n} \text{ is} \\ & \text{almost total,} \\ G_n, & \text{else,} \end{cases} \\
G_0 := G(K_0, L, F_0). & F_{n+1} := \begin{cases} \text{Con}(E_{n+1}) & \text{if } E_{n+1} \not\subseteq L(G_{n+1}) \text{ and} \\ & \circ^{K_{n+1}F_n} \text{ is almost total,} \\ F_n, & \text{else.} \end{cases}
\end{array}$$

Table 2. Grammar inference algorithm

$n \in \mathbb{N}$), and there are infinitely many n such that $F_n \subset F_{n+1} = \text{Con}(E_{n+1})$. Let n be one of those. Then $\circ^{K_{n+1}F_n}$ is almost total and hence $G_{n+1} = G(K_{n+1}, L, F_n)$. We may assume that $F \subseteq F_{n+1}$, so G has the FCP with respect to F_{n+1} .

Since $\langle F_n \mid n \in \mathbb{N} \rangle$ does not get constant, there is a least $m \geq n + 1$ such that $\circ^{K_{m+1}F_m}$ is almost total, and then

$$G_{m+1} = G(K_{m+1}, L, F_m) = G(K_{m+1}, L, F_{n+1}).$$

By lemma 6, $L \subseteq L(G(K_{m+1}, L, F_{n+1})) = L(G_{m+1})$, so $E_{m+1} \subseteq L(G_{m+1})$ and $F_{m+1} = F_m = F_{n+1}$. Let \tilde{m} be the least $k \geq m + 1$ such that $\circ^{K_{k+1}F_k}$ is almost total. Then $G_{\tilde{m}} = G_{m+1}$, $F_{\tilde{m}} = F_{m+1} = F_{n+1}$ and

$$G_{\tilde{m}+1} = G(K_{\tilde{m}+1}, L, F_{\tilde{m}}) = G(K_{\tilde{m}+1}, L, F_{n+1}).$$

Since $\circ^{K_{\tilde{m}+1}F_{\tilde{m}}}$ is almost total, we also have $E_{\tilde{m}+1} \subseteq L \subseteq L(G_{\tilde{m}+1})$, again by lemma 6, so $F_{\tilde{m}+1} = F_{\tilde{m}} = F_{n+1}$. By induction, $F_{\tilde{m}+1} = F_{n+1}$ for all $\tilde{m} \geq n + 1$ where $\circ^{K_{\tilde{m}+1}F_{\tilde{m}}}$ is almost total, hence for all $\tilde{m} \geq n + 1$. This contradicts the assumption that $\langle F_n \mid n \in \mathbb{N} \rangle$ does not get constant.

Claim 2: If $\langle F_n \mid n \in \mathbb{N} \rangle$ converges to $F_{\tilde{n}}$, then $\langle G_n \mid n \in \mathbb{N} \rangle$ converges to $G(L, F_{\tilde{n}})$ and $L = L(G(L, F_{\tilde{n}}))$.

Proof of Claim 2: Suppose $\langle F_n \mid n \in \mathbb{N} \rangle$ converges to $F_{\tilde{n}}$. By lemma 5 there is some finite $K \subseteq \text{Sub}(L)$ such that

$$\langle \mathcal{P}^{KF_{\tilde{n}}}(F_{\tilde{n}}), \circ^{KF_{\tilde{n}}} \rangle \equiv \langle \mathcal{P}^{\triangleleft F_{\tilde{n}}}(F_{\tilde{n}}), \circ^{\triangleleft F_{\tilde{n}}} \rangle. \quad (10)$$

Since (10) is monotone in K and $\langle K_n \mid n \in \mathbb{N} \rangle$ is non-decreasing and majorizes all finite subsets of $\text{Sub}(L)$, there is $m_0 \geq \tilde{n}$ such that for all $m \geq m_0$,

$$\langle \mathcal{P}^{K_{m+1}F_{\tilde{n}}}(F_{\tilde{n}}), \circ^{K_{m+1}F_{\tilde{n}}} \rangle \equiv \langle \mathcal{P}^{\triangleleft F_{\tilde{n}}}(F_{\tilde{n}}), \circ^{\triangleleft F_{\tilde{n}}} \rangle.$$

By lemma 11, $\circ^{K_{m+1}F_{\bar{n}}} = \circ^{K_{m+1}F_m}$ is almost total for $m \geq m_0$. Therefore $G_{m+1} = G(K_{m+1}, L, F_{\bar{n}})$ for all $m \geq m_0$, and $G(K_{m+1}, L, F_{\bar{n}}) = G(L, F_{\bar{n}})$ by lemma 5. So $\langle G_n \mid n \in \mathbb{N} \rangle$ gets constant in $G(L, F_{\bar{n}})$.

We have $L(G(L, F_{\bar{n}})) \subseteq L$ by lemma 1. Suppose $L \not\subseteq L(G(L, F_{\bar{n}}))$, and pick $w \in L \setminus L(G(L, F_{\bar{n}}))$. Since $\langle E_n \mid n \in \mathbb{N} \rangle$ majorizes every finite subset of L , for some $m \geq m_0$ we have $E \cup \{w\} \subseteq E_{m+1} \not\subseteq L(G(L, F_{\bar{n}})) = L(G(K_{m+1}, L, F_m))$. Then $F_{\bar{n}} = F_{m+1} = \text{Con}(E_{m+1}) \supseteq \text{Con}(E) \supseteq F$. So the given G has the FCP with respect to $F_{\bar{n}}$. Then $L \subseteq L(G(K_{m+1}, L, F_m)) = L(G(L, F_{\bar{n}}))$ by lemma 6, in contradiction to the assumption.

The proof does not rely on whether $L(G(K, L, F))$ is monotone in F or antitone in K . In Appendix 2 we show how a grammar for the Dyck language of well-bracketed strings is correctly inferred.

To obtain polynomial update complexity, we need to limit the number of concepts. A grammar has the *f-finite context property* (*f*-FCP), if for each of its nonterminals A there is a set C_A of contexts such that $A^{\text{p}^d} = C_A^d$ and $|C_A| \leq f$, where $f \geq 1$. Restrict the algorithm of table 2 to languages having a grammar with the *f*-FCP, and to build the hypotheses G_n , use grammars $G_f(K, L, F)$ defined like $G(K, L, F)$, but whose nonterminals are non-empty elements from $\mathcal{P}_f^{KF}(F) := \{C^{KF} \mid C \subseteq F, |C| \leq f\}$. Then the algorithm has *polynomial update time*, i.e. the number of steps to generate the hypothesis grammar G_n is bounded by a polynomial in $|E_n|$ and $\max\{|w| \mid w \in E_n\}$.

First observe that the number of steps to compute $G_f(K, L, F)$ from K and F is bounded by a polynomial in $|K|$ and $|F|$. Clearly, $\mathcal{P}_f^{KF}(F)$ is of size $O(|F|^f)$ and its elements can be represented in a trie of bitvectors of length $|F|$. For each $C \subseteq F$ with $|C| \leq f$ we need $O(f|K||F|)$ membership queries to determine C^{KF} and $|F|$ steps to insert it into the trie. Likewise, for $C_1, C_2 \in \mathcal{P}_f^{KF}(F)$ we can compute $C_1 \circ^{KF} C_2$ and check if it belongs to $\mathcal{P}_f^{KF}(F)$ in $O(|K|^2|F|)$ steps. Therefore, we can compute $(\mathcal{P}_f^{KF}(F), \circ^{KF})$ and check if \circ^{KF} is almost total on $\mathcal{P}_f^{KF}(F)$ in $O(|K|^2|F|^{2f+1})$ steps. If \circ^{KF} is almost total, we can read off $G_f(K, L, F)$ from $(\mathcal{P}_f^{KF}(F), \circ^{KF})$ in $O(|V|^3|F|) = O(|F|^{3f+1})$ steps.

Next, let $e = |E_{n+1}|$ and m be the maximal length of words in E_{n+1} . Then $\text{Sub}(E_{n+1})$ and $\text{Con}(E_{n+1})$ are determined in $O(em^2)$ steps, so K_{n+1} and $F_n \subseteq \text{Con}(E_n)$ are of size $O(em^2)$, and $G_{n+1} = G_f(K_{n+1}, L, F_n)$ is polynomial in e and m . Finally, to determine F_{n+1} , we must check whether $E_{n+1} \not\subseteq L(G_{n+1})$, which can be done in time polynomial in e and m using a CYK-recognizer ([6]).

4 Conclusion

We have pointed out that Clark's grammar inference algorithm may converge to a grammar for a superset of the intended language. We modified Clark's grammars $G(K, L, F)$, replaced major parts of the reasoning for the inference process, and provided proofs of the correctness of the algorithm. We have thus shown that one can learn a grammar for L , if L does have a CNF grammar

with the finite context property, and can do so in the framework of relativized syntactic concepts of L . Some experts seemed to think this was impossible⁵, as [2] and [8] rely heavily on the (anti)monotonicity of $L(G(K, L, F))$ in $(K)F$.

Yoshinaka's [8] "dual" approach uses grammars $H(K, L, F)$ that differ from our $G(K, L, F)$ by admitting arbitrary $C \subseteq F$ as nonterminals in order to make $L(H(K, L, F))$ monotone in F and antitone in K . This simplifies the reasoning, but leads to a limit grammar with many "copies" of the same rule.

If we consider syntactic concepts of L as the only linguistically relevant notions to be used in describing L , we would like to do syntactic analysis in terms of "concept arithmetic", i.e. using $\mathcal{B}(L)$ rather than $\mathcal{P}(\Sigma^*)$. It remains to be developed what this amounts to, in particular when L is not context-free.

Appendix 1

For completeness, we here provide a proof of the basic properties of Galois-connections.

Lemma 7. *Let $f : (A, \leq^A) \rightleftarrows (B, \leq^B) : g$ be a Galois connection, i.e. (A, \leq^A) and (B, \leq^B) are partial orders, and for all $a \in A$ and $b \in B$*

$$gb \leq^A a \iff b \leq^B fa. \quad (11)$$

Then the following hold:

- (i) $g \circ f$ is shrinking and $f \circ g$ is extending, i.e. $g(fa) \leq^A a$ and $b \leq^B f(gb)$ for all $a \in A, b \in B$.
- (ii) f and g are monotone., i.e. $fa \leq^B fa'$ for $a \leq^A a'$, and $gb \leq^A gb'$ for $b \leq^B b'$.
- (iii) $f \circ g \circ f = f$ and $g \circ f \circ g = g$.
- (iv) $g \circ f$ is a closure operator on (A, \geq^A) and $f \circ g$ is a closure operator on (B, \geq^B) .
- (v) $(\{gb \mid b \in B\}, \leq^A)$ and $(\{fa \mid a \in A\}, \leq^B)$ are isomorphic partial orders.

Proof. (i) From $fa \leq^B fa$ and (11) we get $g(fa) \leq^A a$, and from $gb \leq^A gb$ and (11) we get $b \leq^B f(gb)$.

(ii) For $a \leq^A a'$ we get $g(fa) \leq^A a'$ by (i), and with (11) then $fa \leq^B fa'$. Hence f is monotone. Likewise for g : From $b \leq^B b'$ and $b' \leq^B f(gb')$ of (i) we get $b \leq^B f(gb')$, so $gb \leq^A gb'$ by (11).

(iii) $f \circ g \circ f = f$: By (i) we have $g(fa) \leq^A a$ and by (ii) f is monotone, so $f(g(fa)) \leq^B fa$. For $b = fa$ in (i) we get $fa \leq^B f(g(fa))$, so in combination we have $f(g(fa)) = fa$.

$g \circ f \circ g = g$: By (i) we have $b \leq^B f(gb)$, and by (ii) g is monotone, so $gb \leq^A g(f(gb))$. For $a = gb$ in (i) we get $g(f(gb)) \leq^A gb$. Taken together, this gives $g(f(gb)) = gb$.

⁵ Ryo Yoshinaka: "I am afraid relativized lattices do not have quite a right property to base on for learning.", personal communication, February 2013

(iv) By (i), $b \leq f(gb)$, so $f \circ g$ is extending. Since by (ii) g and f are monotone, so is $f \circ g$. By (iii), $f \circ g \circ f \circ g = f \circ g$, so $f \circ g$ is idempotent. It follows that $f \circ g$ is a closure operator on (B, \leq^B) .

Since $g \circ f$ is shrinking with respect to \leq^A , it is extending with respect to \geq^A . As f and g are monotone with respect to \leq^A and \leq^B , $g \circ f$ is monotone with respect to \leq^A , hence also with respect to \geq^A . By (iii), $g \circ f \circ g \circ f = g \circ f$, so $g \circ f$ is idempotent. Thus, it is a closure operator on (A, \geq^A) .

(v) Let $g(B) = \{gb \mid b \in B\}$ be the set of “small” elements of A and $f(A) = \{f(a) \mid a \in A\}$ be the set of “large” elements of B . Then

$$f : (g(B), \leq^A) \rightleftharpoons (f(A), \leq^B) : g$$

is a pair of order preserving embeddings. Moreover, $g : f(A) \rightarrow g(B)$ is the inverse of $f : g(B) \rightarrow f(A)$, since by (iii) for each $b \in B$ we have $f(g(b)) \in f(A)$, and $g(f(g(b))) = g(b)$. Likewise is $f : g(B) \rightarrow f(A)$ the inverse of $g : f(A) \rightarrow g(B)$.

Appendix 2

Example 2. Let L be the Dyck-language of well-bracketed strings over the alphabet $\Sigma = \{a, b\}$ of opening bracket a and closing bracket b , given in the enumeration $L = \{w_1, w_2, w_3, w_4, \dots\} = \{ab, aabb, abab, \epsilon, \dots\}$. To determine $G(K, L, F) = (V, \Sigma, P, S)$, we compute $V = \mathcal{P}^{KF}(F) \setminus \{\emptyset\}$ and $X \circ^{KF} Y = (X^K Y^K)^F$ for $X, Y \in V$. Then we need to check if \circ^{KF} is almost total and compute the next grammar and context set.

$n = 0$: $E_0 = \emptyset$, $K_0 = \{\epsilon, a, b\}$, $F_0 = \{(\epsilon, \epsilon)\}$. To determine $G_0 = G(K_0, L, F_0) = (V_0, \Sigma, P_0, S_0)$, we compute $V_0 = \mathcal{P}^{K_0 F_0}(F_0) \setminus \{\emptyset\}$ and $X \circ^{K_0 F_0} Y$ for $X, Y \in V_0$.

$$\begin{array}{c} \frac{C \subseteq F_0 \mid C^{K_0} \quad C^{K_0 F_0} \parallel V_0}{\emptyset \mid K_0 \quad \emptyset \parallel -} \\ (\epsilon, \epsilon) \mid \epsilon \quad (\epsilon, \epsilon) \parallel \perp \end{array} \quad \begin{array}{c} \top \\ \uparrow \\ \perp \end{array} \quad \begin{array}{c} \frac{X^{K_0} Y^{K_0} \mid (\epsilon, \epsilon)}{(\epsilon, \epsilon) \mid \epsilon} \quad \frac{X \circ^{K_0 F_0} Y \mid (\epsilon, \epsilon)}{(\epsilon, \epsilon) \mid (\epsilon, \epsilon)} \end{array}$$

Hence, $S_0 = \perp$ and $P_0 = \{\perp \rightarrow \perp \perp \mid \epsilon\}$, as $\perp \subseteq (\perp \circ^{K_0 F_0} \perp)$, $\perp \subseteq \perp = \epsilon^{F_0}$, and $\perp \not\subseteq \emptyset = a^{F_0} = b^{F_0}$.

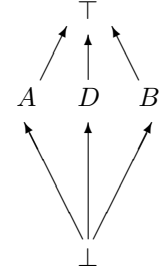
$n = 1$: We have $E_1 = \{ab\}$, $K_1 = \{\epsilon, a, b, ab\}$, and

$$\frac{C \subseteq F_0 \mid C^{K_1} \quad C^{K_1 F_0} \parallel V_1}{\emptyset \mid K_1 \quad \emptyset \parallel -} \quad \frac{X^{K_1} Y^{K_1} \mid (\epsilon, \epsilon)}{(\epsilon, \epsilon) \mid \epsilon + ab + (ab)^2} \quad \frac{X \circ^{K_1 F_0} Y \mid (\epsilon, \epsilon)}{(\epsilon, \epsilon) \mid (\epsilon, \epsilon)}$$

Since $(\epsilon, \epsilon) = (\epsilon, \epsilon)^{K_1 F_0}$, $\circ^{K_1 F_0}$ is almost total, hence $G_1 = G(K_1, L, F_0)$ has $S_1 = \perp$ and $P_1 = \{\perp \rightarrow \epsilon \mid \perp \perp\}$. So $E_1 \not\subseteq L(G_1) = \{\epsilon\}$ and $F_1 = \text{Con}(E_1) = \{(\epsilon, \epsilon), (\epsilon, b), (a, \epsilon), (a, b), (\epsilon, ab), (ab, \epsilon)\}$

$n = 2$: We have $E_2 = \{ab, aabb\}$ and $K_2 = \{\epsilon, a, b, ab, aa, bb, aab, abb, aabb\}$.

$C \subseteq F_1$	C^{K_2}	$C^{K_2 F_1}$	V_2
\emptyset	K_2	\emptyset	—
(ϵ, ϵ)	$\epsilon + ab + a^2b^2$	$(\epsilon, \epsilon) + (a, b) + (\epsilon, ab) + (ab, \epsilon)$	—
(ϵ, b)	$a + a^2b$	(ϵ, b)	A
(a, ϵ)	$b + ab^2$	(a, ϵ)	B
(a, b)	$\epsilon + ab + a^2b^2$	$(\epsilon, \epsilon) + (a, b) + (\epsilon, ab) + (ab, \epsilon)$	—
(ϵ, ab)	$\epsilon + ab + a^2b^2$	$(\epsilon, \epsilon) + (a, b) + (\epsilon, ab) + (ab, \epsilon)$	—
(ab, ϵ)	$\epsilon + ab + a^2b^2$	$(\epsilon, \epsilon) + (a, b) + (\epsilon, ab) + (ab, \epsilon)$	—
$(\epsilon, \epsilon) + (a, b)$	$\epsilon + ab + a^2b^2$	$(\epsilon, \epsilon) + (a, b) + (\epsilon, ab) + (ab, \epsilon)$	—
...			—
$(\epsilon, \epsilon) + (a, b)$	$\epsilon + ab + a^2b^2$	$(\epsilon, \epsilon) + (a, b) + (\epsilon, ab) + (ab, \epsilon)$	D
$+(\epsilon, ab) + (ab, \epsilon)$			—
...	\emptyset	F_1	—
F_1	\emptyset	F_1	\perp



For $X^{K_2} Y^{K_2}$ and $X \circ^{K_2 F_1} Y = (X^{K_2} Y^{K_2})^{F_1}$ we obtain

$X^{K_2} Y^{K_2}$	A	B	D	\perp
A	$a^2 + a^3b$ $+a^2ba + a^2ba^2b$	$ab + a^2b^2$ $+a^2bab^2$	$a^2b + a^3b^2$ $+a^2bab + a^2ba^2b^2$	\emptyset
B	$ba + ba^2b$ $+ab^2a + ab^2a^2b$	$b^2 + bab^2$ $+ab^3 + ab^2ab^2$	$bab + ba^2b^2$ $+ab^2ab + ab^2a^2b^2$	\emptyset
D	$aba + aba^2b$ $+a^2b^2a + a^2b^2a^2b$	$ab^2 + abab^2$ $+a^2b^3 + a^2b^2ab^2$	$(ab)^2 + aba^2b^2$ $+a^2b^2ab + (a^2b^2)^2$	\emptyset
\perp	\emptyset	\emptyset	\emptyset	\emptyset

and

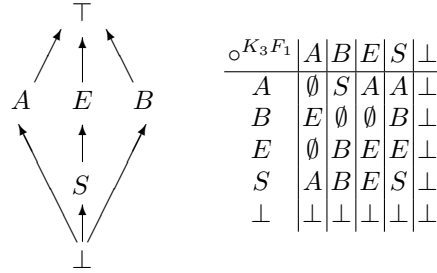
$\circ^{K_2 F_1}$	A	B	D	\perp
A	\emptyset	D	A	\perp
B	(a, b)	\emptyset	B	\perp
D	A	B	D	\perp
\perp	\perp	\perp	\perp	\perp

Since $(B \circ^{K_2 F_1} A) = (a, b) \neq (a, b)^{K_2 F_1}$, $\circ^{K_2 F_1}$ is not almost total. Hence $G_2 = G_1$ and $F_2 = F_1$.

$n = 3$: We have $E_3 = \{ab, a^2b^2, abab\}$, $K_3 = \{\epsilon, a, b, ab, aa, bb, ba, a^2b, ab^2, aba, bab, a^2b^2, (ab)^2\}$.

$C \subseteq F_1$	C^{K_3}	$C^{K_3F_1}$	V_3
\emptyset	K_2	\emptyset	—
(ϵ, ϵ)	$\epsilon + ab + a^2b^2 + (ab)^2$	$(\epsilon, \epsilon) + (a, b) + (\epsilon, ab) + (ab, \epsilon)$	—
(ϵ, b)	$a + a^2b + aba$	(ϵ, b)	A
(a, ϵ)	$b + ab^2 + bab$	(a, ϵ)	B
(a, b)	$\epsilon + ab + a^2b^2 + ba + (ab)^2$	(a, b)	E
(ϵ, ab)	$\epsilon + ab + a^2b^2 + (ab)^2$	$(\epsilon, \epsilon) + (a, b) + (\epsilon, ab) + (ab, \epsilon)$	—
(ab, ϵ)	$\epsilon + ab + a^2b^2 + (ab)^2$	$(\epsilon, \epsilon) + (a, b) + (\epsilon, ab) + (ab, \epsilon)$	—
$(\epsilon, \epsilon) + (a, b)$	$\epsilon + ab + a^2b^2 + (ab)^2$	$(\epsilon, \epsilon) + (a, b) + (\epsilon, ab) + (ab, \epsilon)$	—
...	—
$(\epsilon, \epsilon) + (a, b) + (\epsilon, ab) + (ab, \epsilon)$	$\epsilon + ab + a^2b^2 + (ab)^2$	$(\epsilon, \epsilon) + (a, b) + (\epsilon, ab) + (ab, \epsilon)$	S
...	\emptyset	F_1	—
F_1	\emptyset	F_1	\perp

This leads to the following lattice $\mathcal{B}(L, F_1)$, where the operation $\circ^{K_3F_1}$ is almost total on $\mathcal{P}^{K_3F_1}(F_1)$:



We get $\epsilon^{F_1} = (\epsilon, \epsilon) + (a, b) + (ab, \epsilon) + (\epsilon, ab) = S$, $a^{F_1} = (\epsilon, b) = A$, $b^{F_1} = (a, \epsilon) = B$, $S_3 = S$, and

$$P_3 = \left\{ \begin{array}{l} A \rightarrow a \mid AE \mid AS \mid SA, \\ B \rightarrow b \mid BS \mid EB \mid SB, \\ E \rightarrow \epsilon \mid AB \mid BA \mid EE \mid ES, \mid SE \mid SS, \\ S \rightarrow \epsilon \mid AB \mid SS \end{array} \right\}$$

where we omitted rules for \perp , like $\perp \rightarrow A\perp$, since⁶ \perp is not reachable from S , whence \perp is an unnecessary symbol. This is a grammar for L .

Since the inferred grammars contain *all* rules that are “correct” with respect to $\mathcal{B}(L, F)$, the grammars often may be simplified by omitting some rules.

Acknowledgement

I wish to thank A. Clark for email discussions and W. Buszkowski and R. Yoshinaka for pointing out mistakes in a draft version and for a hint to Jipsen e.a.[4].

⁶ Slight mistakes with \perp corrected on Nov. 11th, 2014

References

1. Alexander Clark. A learnable representation for syntax using residuated lattices. In *Proceedings of the 14th Conference on Formal Grammar*, Bordeaux, France, 2009. LNCS 5591, pages 183–198. Springer Verlag, Berlin, 2009.
2. Alexander Clark. Learning context free grammars with the syntactic concept lattice. In *10th International Colloquium on Grammatical Inference, ICGI 2010, Valencia*. LNAI 6339, pages 38–51. Springer Verlag, Berlin, 2010.
3. Zellig S. Harris. From morpheme to utterance. *Language*, 22(3):161–83, 1946.
4. P. Jipsen and C. Tsinakis. A survey of residuated lattices. In J. Martinez, editor, *Ordered Algebraic Structures*, pages 19–56. Kluwer, 2002.
5. Hans Leiß. Learning CFGs with the finite context property. A note on A. Clark’s algorithm. Universität München, CIS, Manuscript, July 2012.
6. John E. Hopcroft and J.D. Ullman. Introduction to Automata Theory, Languages, and Computation. Addison-Wesley, 1979.
7. Christian Wurm. Completeness of full Lambek calculus for syntactic concept lattices. In *Formal Grammar. 17th and 18th International Conferences, FG 2012 and FG 2014, Proceedings*, LNCS 8036, pages 126–141. Springer-Verlag, Berlin, 2013.
8. Ryo Yoshinaka. Towards dual approaches for learning context-free grammars based on syntactic concept lattices. In G. Maury and A. Leporati, editors, *Developments in Language Theory. 15th International Conference, DLT 2011, Milan, Italy, July 19-22, 2011. Proceedings*, pages 429–440. Springer Verlag, Berlin Heidelberg, 2011.