J. Expt. Theor. Artif. Intell. 7(1995)347–359

Teams in cooperating grammar systems

LILA KARI*, ALEXANDRU MATEESCU*, GHEORGHE PĂUN† and ARTO SALOMAA*

*Academy of Finland and University of Turku, Department of Mathematics, SF-20500 Turku, Finland
†Institute of Mathematics of the Romanian Academy of Sciences, PO Box 1-764, 70700 Bucuresti, Romania

Abstract. We consider grammar systems in which several components are active at the same moment (a *team* of components is working). The power of such mechanisms is investigated and it is found that in many cases the team feature increases the generative capacity of grammar systems. In the so-called *t*-mode of derivation (a team works as much as it can) it is found that the team size does not induce an infinite hierarchy of languages. However, the family obtained in this case is a full abstract family of languages properly including ET0L.

Keywords: context-free grammars, the blackboard model, grammar systems, ETOL systems, abstract families of languages

Received 16 April 1993; revision received 6 November 1993; revision accepted 13 January 1994

1. Introduction

The cooperating distributed (CD for short) grammar systems in the form we consider here were introduced in Csuhaj-Varju and Dassow (1990) with motivations from artificial intelligence; earlier versions appear in Atanasiu and Mitrana (1989) and Meersman and Rozenberg (1978), motivated by questions related to two-level grammars and regulated rewriting, respectively.

Roughly speaking, such a system consists of several ordinary grammars (any type of rewriting device can be used) working by turns on the same sentential form; at each moment, one component is active, the others are waiting; one starts from a common axiom and one continues working until a terminal word is obtained.

Systems of agents cooperating in solving a common task according to a similar protocol can be met in various areas of artificial intelligence, cognitive psychology and computer science. Following Csuhaj-Varju and Dassow (1990), Csuhaj-Varju and Kelemen (1989), Dassow and Kelemen (1991), Kelemen (1991) and Chapter 1 of Csuhaj-Varju *et al.* (1994), we briefly emphasize here the adequacy of such a model to the so-called blackboard architecture for problem solving systems (see the survey of Nii 1989). More discussion about this relationship and about motivations for CD grammar systems arising from other branches of artificial intelligence can be found in the quoted papers.

The blackboard model consists of three basic parts:

- the knowledge sources, necessary to solve a given problem;
- the global database, represented by the blackboard, describing the current state of the problem-solving process and to which the knowledge sources can operate;
- the control of the order in which the knowledge sources work.

One starts having on the blackboard the initial formulation of the problem; the knowledge sources contribute to the problem solving by changing the current state of the blackboard. During this process, they communicate with each other only through the blackboard content.

The following natural analogy with systems of grammars was already pointed out in Csuhaj-Varju and Kelemen (1989); the knowledge sources correspond to grammars, the current state of the blackboard to a sentential form; the rewriting of a nonterminal can be interpreted as a step of processing the information contained in the current description of the problem, a solution of the problem corresponds to a terminal word. The control can be ensured by some mechanism which determines the order of grammars enabling.

In fact, in the study of CD grammar systems it seems to be more appealing to not consider an external control, but to work according to an opportunistic strategy: a component takes the current sentential form when it can do at least one rewriting and leaves it when a precise condition is fulfilled. The most important such condition is that corresponding to the maximal competence rule: a component works as long as it can. This is called the t-mode of derivation in Csuhaj-Varju and Dassow (1990). Other investigated stopping conditions ask for a component to work a given number of steps, at least, or at most a given number of steps, or arbitrarily many steps.

Two results for the t-mode of derivation are particularly interesting (Csuhaj-Varju and Dassow 1990, Csuhaj-Varju *et al.* 1994): (i) systems consisting of two context-free components can not generate non-context-free languages but three components suffice, systems with an arbitrary number of components can be simulated by systems with three components only; (ii) the family of languages obtained in this way equals the family of languages generated by extended tabled Lindenmayer systems without interaction (ETOL systems), one of the most important (comprehensive, rich in mathematical properties, well studied) in the theory of Lindenmayer languages.

Thus, the cooperation is useful as generative power and (in the t-mode) a comprehensive family of languages is obtained. In view of the discussed relation with the blackboard architectures and with other topics related to complex systems, and also natural from a mathematical point of view, it is of interest to increase (as much as possible) the power of CD grammar systems (using as simple as possible extra features). If a graph is considered as external control or certain predicates depending on the current sentential form are used as start and stop conditions, then in general the family of programmed languages is obtained (Csuhaj-Varju *et al.* 1990), which includes properly the ETOL family (Dassow and Păun 1989). What about not considering such new elements of the model, but increasing the degree of cooperation? A natural possibility is to imagine systems in which more components are active at the same time (a *team*), namely a prescribed number of them, non-deterministically chosen. Such a mixed system, sequential in essence but also involving some degree of parallelism, is probably the most common (and the most efficient) mode of work of complex problem solvers, met for instance, at the level of the brain activity. On the other hand, formally, a team CD grammar system can be also seen as intermediate between the CD grammar systems as described above and the parallel communicating grammar systems introduced in Păun and Sântean (1989) as a grammatical model of parallel (synchronized) computing. In such systems, all components are simultaneously active (but they work on separate sentential forms and cooperate by direct communication, sending on request the current string from one to another).

We investigate here mainly the t-derivation mode (a team works until exhausting its competence, that is until no further symbol from the left sides of its rules is present in the current sentential form, hence no rule of any member of the team can be used). The results are both interesting from the 'practical' point of view—the team feature increases strictly the power of CD grammar systems, hence the cooperation by means of teams is useful—and surprisingly—the size of teams induces no hierarchy, teams of two members suffice (half of this second statement is proved here, the other half appears in Csuhaj-Varju and Păun 1993).

Besides the generative capacity of team CD grammar systems we investigate also closure properties of the family obtained in the t-mode. Studying operations with languages seems, at first sight, a purely theoretical question. However, representing a complex problem as a 'product' of simpler problems amounts exactly to starting from simpler languages describing the simpler problems and mixing them by using appropriate operations. The operations to be used in a particular case is a matter of practice. Putting together by union, renaming by morphisms and inverse morphisms, selecting subsets by intersection with regular languages are only a few possibilities. In language theory the operations are systematically studied in the frame of AFL theory (abstract family of languages—families closed under six basic operations: union, concatenation, morphisms, inverse morphisms, Kleene closure and intersection by regular sets). The result we obtain here is quite pleasant from this point of view: the family generated by team CD grammar systems in the t-mode is a full AFL.

2. Definitions

The reader is referred to Salomaa (1973) for basic elements of formal language theory and to Csuhaj-Varju *et al.* (1994) for details about grammar systems.

For an alphabet V, we denote by V^* the free monoid generated by V under the operation of concatenation; λ is the empty string and $|x|$ is the length of $x \in V^*$. The families of context-free and of ET0L languages are denoted by *CF*, *ET0L*, respectively; *EDT0L* is the family of deterministic ET0L languages.

A CD grammar system (of degree n, $n \geq 1$) is a construct

$$\Gamma = (N, T, P_1, P_2, \ldots, P_n, S),$$

where N is a (nonterminal) alphabet, T is a (terminal) alphabet disjoint from N, $S \in N$ and P_i are finite sets of context-free rules over $N \cup T$, $1 \leq i \leq n$.

For given P_i, the direct derivation $\Rightarrow_{P_i}$ is defined in the usual way; we denote by $\Rightarrow_{P_i}^{=k}$, $\Rightarrow_{P_i}^{\leq k}$, $\Rightarrow_{P_i}^{\geq k}$, $\Rightarrow_{P_i}^{*}$, $\Rightarrow_{P_i}^{t}$ a derivation in P_i consisting of exactly k steps, at most k, at least k steps, $k \geq 1$, of any number of steps and as long as possible, respectively ($x \Rightarrow_{P_i}^{t} y$ means that $x \Rightarrow_{P_i}^{*} y$ and there is no z such that $y \Rightarrow_{P_i} z$).

For $f \in \{*, t\} \cup \{\leq k, =k, \geq k \mid k \geq 1\}$ we denote by $L_f(\Gamma)$ the language generated by Γ in the f mode, that is

$$L_f(\Gamma) = \{x \in T^* \mid S \Rightarrow_{P_{i_1}}^{f} x_1 \Rightarrow_{P_{i_2}}^{f} \ldots \Rightarrow_{P_{i_m}}^{f} = x,$$
$$1 \leq i_j \leq n,\ 1 \leq j \leq m,\ m \geq 1\}$$

and by $CD(f)$ the family of such languages. (Note that we do not distinguish here between systems with λ-free and with arbitrary components.)

Given a grammar system $\Gamma = (N, T, P_1, \ldots, P_n, w)$ with components $N, T, P_1, \ldots, P_n$ as above but with a string axiom $w \in (N \cup T)^*$ instead of a symbol $S \in N$, and given a natural number $s \geq 1$, a subset $Q = \{P_{i_1}, \ldots, P_{i_s}\}$ of $\{P_1, \ldots, P_n\}$ is called an s-team. For such an s-team Q and for $x, y \in (N \cup T)^*$, we write

$$x \Rightarrow_Q y \text{ iff } x = x_1 A_1 x_2 \ldots A_s x_{s+1},\ y = x_1 y_1 x_2 \ldots y_s x_{s+1},$$
$$x_j \in (N \cup T)^*,\ 1 \leq j \leq s+1,\ A_j \to y_j \in P_{i_j},\ 1 \leq j \leq s$$

(in a direct derivation step, each member of the team uses one of its rules, in parallel with the other components of the team; as Q is a set, no order of the components is assumed, hence the rewritten symbols $A_1, \ldots, A_s$ can appear in x in any order).

Then the relations $\Rightarrow_Q^{=k}$, $\Rightarrow_Q^{\leq k}$, $\Rightarrow_Q^{\geq k}$, $\Rightarrow_Q^{*}$, $\Rightarrow_Q^{t}$, $k \geq 1$, can be defined as above, with the clarification that in the t case the derivation is correct when no more rules of any of the team components can be used (for instance, the team $Q = \{\{A \to a\}, \{B \to b\}\}$ can rewrite correctly $AABB$ in two steps, $AABB \Rightarrow_Q aAbB \Rightarrow_Q aabb$, hence we have $AABB \Rightarrow_Q^t aabb$, but AAB cannot be rewritten in the t-mode: after $AAB \Rightarrow_Q aAb$, the team cannot work, but A is still present and it can be rewritten according to the first component of the team).

For given $s \geq 1$ and f as above, we denote by $L_f(\Gamma, s)$ the language generated by Γ in the f-mode, with s-teams. The family of these languages is denoted by $TCD_s(f)$; moreover, we put $TCD(f) = \bigcup_{s \geq 1} TCD_s(f)$. (Note that we allow λ-rules in the systems we deal with.)

3. Examples

We shall consider two examples, both for illustrating the previous definitions and the way in which the teams work, and for proving the power of team CD grammar systems of various types.

$$\Gamma_1 = (\{A, B, A', B'\}, \{a, b, c\}, P_1, P_2, P_3, P_4, AB),$$
$$P_1 = \{A \to aA'b, A \to ab\},$$
$$P_2 = \{B \to cB', B \to c\},$$
$$P_3 = \{A' \to A\},$$
$$P_4 = \{B' \to B\}.$$

We have

$$L_t(\Gamma_1, 2) = L_{=1}(\Gamma_1, 2) = L_{\geq 1}(\Gamma_1, 2) = L_*(\Gamma_1, 2) = L_{\leq k}(\Gamma_1, 2)$$
$$= \{a^n b^n c^n \mid n \geq 1\},\ k \geq 1.$$

Indeed, we start from a string with two nonterminals (AB); at any moment of time, if a string has only one nonterminal, then no team can work. Therefore the rules $A \to ab$, $B \to c$ can be used only at the same time, in the last step of a derivation. On the other hand, the only possible teams are $Q_1 = \{P_1, P_2\}$ and $Q_2 = \{P_3, P_4\}$. In conclusion, the possible derivations are

$$AB \Rightarrow_{Q_1}^f aA'bcB' \Rightarrow_{Q_2}^f aAbcB \Rightarrow_{Q_1}^f a^2A'b^2c^2B' \Rightarrow_{Q_2}^f a^2Ab^2c^2B \Rightarrow_{Q_1}^f$$
$$\ldots \Rightarrow_{Q_2}^f a^nAb^nc^nB \Rightarrow_{Q_1}^f a^{n+1}b^{n+1}c^{n+1},$$

for $n \geq 0, f \in \{=1, \geq 1, *, t\} \cup \{\leq k \mid k \geq 1\}$.

Note that this language is not context-free, although $CD(*) = CD(=1) =$

$CD(\geq 1) = CD(\leq k) = CF$, $k \geq 1$ (Csuhaj-Varju and Dassow (1989)). Moreover, as each rule in Γ_1 can be considered terminal if we separate the nonterminal and the terminal alphabets on components, the system Γ_1 can be considered a colony, in the sense of Kelemen and Kelemenova (1992). (A colony is a grammar system with regular components generating finite languages and rewriting only the occurrences of the axioms in the current string.) As the colonies with $(=1)$-derivation characterize the context-free languages and colonies with t-derivation characterize the 1-restricted ET0L languages in the sense of Kleijn and Rozenberg (1980) (the result is proved in Kelemenova and Csuhaj-Varju 1992), which in turn cannot generate the language $\{a^n b^n c^n \mid n \geq 1\}$, the previous example shows that in the case of colonies the team feature enlarges the generative power. As we shall see, this is the case also for general grammar systems for many derivation modes.

Let us now examine a more sophisticated example:

$$\Gamma_2 = (\{A, B, A', B'\}, \{a, b, c\}, P_1, P_2, P_3, P_4, P_5, P_6, AB),$$

$$P_1 = \{A \rightarrow A'A'\},$$

$$P_2 = \{B \rightarrow aB, B \rightarrow bB, B \rightarrow aB', B \rightarrow bB'\},$$

$$P_3 = \{A' \rightarrow AA\},$$

$$P_4 = \{B' \rightarrow aB', B' \rightarrow bB', B' \rightarrow aB, B' \rightarrow bB\},$$

$$P_5 = \{A \rightarrow c, B \rightarrow B, B' \rightarrow B'\},$$

$$P_6 = \{B \rightarrow cB, B \rightarrow c, A \rightarrow A, A' \rightarrow A'\}.$$

We obtain

$$L_t(\Gamma_2, 2) = \{c^{2^{2n}} w c^{2^{2n}} \mid w \in \{a, b\}^*, |w| = 2^{2n} - 1, n \geq 0\}.$$

Indeed, P_5 and P_6 cannot be in a team with components P_1, P_2, P_3, P_4, due to the presence of the 'trap rules' $A \rightarrow A, A' \rightarrow A', B \rightarrow B, B' \rightarrow B'$: at least one of these symbols is present, hence the derivation cannot be correctly finished in the t-mode of derivation. By using P_5, P_6 together, we end the derivation (by introducing the same number of c occurrences on the left side and on the right side of the generated string). Starting from AB, besides P_5, P_6, we can use only the team $\{P_1, P_2\}$, which leads to $A'A'aB'$ or to $A'A'bB'$. Now only P_3, P_4 can be applied and we get $A^4 xB$, $x \in \{a, b\}^*, |x| = 3$. The process can be iterated, for $i \geq 0$ times, thus obtaining $A^{2^{2i}} wB$, $w \in \{a, b\}^*, |w| = 2^{2i} - 1$. Using now the team $\{P_5, P_6\}$ we get $c^{2^{2i}} w c^{2^{2i}}$. (Please note, besides the role of the 'trap-rules' of the form $X \rightarrow X$, the way in which the last team produces the same number of symbols c both from $A^{2^{2i}}$ and from the single B, by synchronizing the rewriting in the two components of the team.)

4. On the generative power

Theorem 1: $CD(f) = TCD_1(f) \subset TCD_2(f), f \in \{*, =1, \geq 1\} \cup \{\leq k \mid k \geq 1\}$.

Proof: We have already pointed out that $CF = CD(f), f$ as above. The equality $CD(f) = TCD_1(f)$ is obvious from definitions and the first example in Section 3 shows that $TCD_2(f) - CF \neq \emptyset$. Therefore it remains to prove that $CF \subseteq TCD_2(f)$. We shall prove the more general relation $CF \subseteq TCD_s(f)$, $s \geq 2$.

Take $G = (N, T, S, P)$ a context-free grammar and construct the system

$$\Gamma_s = (N \cup \{A_i \mid 1 \leq i \leq s-1\}, T, P_1, P_2, \ldots, P_{s-1}, P_s, SA_1A_2 \ldots A_{s-1}),$$

with

$$P_i = \{A_i \to A_i, A_i \to \lambda\},\ 1 \leq i \leq s-1,$$

$$P_s = P.$$

The only possible s-team consists of all components $P_1, \dots, P_s$, hence in $P_1, \dots, P_{s-1}$ we have to use the rules $A_i \to A_i$, $1 \leq i \leq s-1$, respectively, and in P_s any rule of P, excepting the last step, when in P_i, $1 \leq i \leq s-1$, the rules $A_i \to \lambda$ are used, together with a terminal rule in $P_s = P$. The equality $L(G) = L_f(\Gamma_s)$ follows. □

It is plausible to expect that the size of the team leads to a hierarchy of language families generated by various classes of grammar systems. Surprisingly enough, for the t-mode of derivation the result is different: teams of size two suffice. The cases of other modes of derivation remains *open*. We want to emphasize that our situation differs from analogous situations in language theory: it is by no means obvious that an increase in the size of teams implies an increase in the language family.

Theorem 2: $TCD_s(t) \subseteq TCD_{s+1}(t)$, $s \geq 1$.

Proof: For a CD grammar system $\Gamma = (N, T, P_1, \dots, P_n, w)$ and for given $s \geq 1$ construct the CD grammar system Γ' having the terminal alphabet T, the nonterminal alphabet

$$N' = N \cup \{A_1, \dots, A_n, A'_1, \dots, A'_n, A''_1, \dots, A''_n, B, B', B''\},$$

the axiom string

$$w' = wA_1A_2 \dots A_nB,$$

and the next components:

1. $P_0 = \{B \to B, B \to B'\}$,
   $P'_0 = \{B' \to B', B' \to B''\}$,
   $P''_0 = \{B'' \to B'', B'' \to B\}$,
   $P'''_0 = \{B \to \lambda\} \cup \{X \to X \mid X \in N\} \cup$
   $\cup \{A_i \to A_i, A'_i \to A'_i, A''_i \to A''_i \mid 1 \leq i \leq n\}$;

2. $P'_i = \{A_i \to \lambda, A''_i \to \lambda\} \cup \{X \to X \mid X \in N\} \cup \{B'' \to B''\}$,
   $P''_i = \{A_i \to A''_i\} \cup \{X \to X \mid X \in N\} \cup \{B'' \to B''\}$,
   $P'''_i = \{A''_i \to A_i\} \cup \{X \to X \mid X \in N\} \cup \{B'' \to B''\}$,
   for all $i = 1, 2, \dots, n$;

3. $P_{i,M} = \{A_i \to A'_i\} \cup$
   $\cup \{A_j \to A_j \mid j \in M\} \cup \{A'_j \to A'_j \mid j \in \{1, 2, \dots, n\} - M\} \cup$
   $\cup \{B \to B, B'' \to B''\}$, $i \in M$,
   $P'_{i,M} = P_i \cup \{A'_j \to A'_j \mid j \in \{1, 2, \dots, n\} - M\} \cup$
   $\cup \{A_j \to A_j \mid j \in M\} \cup \{B \to B, B' \to B'\}$, $i \in M$,
   $P''_{i,M} = \{A'_i \to A_i\} \cup$
   $\cup \{A'_j \to A'_j \mid j \in M\} \cup \{A_j \to A_j \mid j \in \{1, 2, \dots, n\} - M\} \cup$
   $\cup \{B' \to B', B'' \to B''\}$, $i \in M$,
   for $M \subseteq \{1, 2, \dots, n\}$, $\mathrm{card}(M) = s$.

Let us examine the possible derivations in Γ' by teams of $s+1$ components.

Every nonterminal string generated by Γ' is of the form $w''z\alpha$, $w'' \in (T \cup N)^*$,

$z \in \{A_1, \ldots, A_n, A'_1, \ldots, A'_n, A''_1, \ldots, A''_n\}^*$, $\alpha \in \{B, B', B''\}$. (The nonterminal B is removed only in the last step of a derivation, using the component P'''_0.) Initially we have $w' = wA_1 \ldots A_nB$. Therefore, every team can contain at most one of the components P_0, P'_0, P''_0.

No pair of components (P_{i,M_1}, P'_{j,M_2}), (P_{i,M_1}, P''_{j,M_2}), $(P'_{i,M_1}, P''_{j,M_2})$ can belong to the same team, because all rules $B \to B$, $B' \to B'$, $B'' \to B''$ appear in all such pairs, irrespective of the values of i, j, M_1, M_2, hence the derivations cannot be correctly finished (at least one of B, B', B'' is present).

Similarly, P_0 cannot be in a team with any $P'_{i,M}$ (P_0 requests the presence of B, which will be replaced by B', whereas $P'_{i,M}$ contains the rule $B' \to B'$), or with any $P''_{i,M}$; P'_0 cannot be in a team with any $P'_{i,M}$ or $P''_{i,M}$ and P''_0 cannot be in a team with any of $P_{i,M}$ or $P'_{i,M}$.

No component $P'_{i,M}$, $P''_{i,M}$ can be in a team with P'''_0 (due to the rules $A_i \to A_i$, $A'_i \to A'_i$ in P'''_0); if $P'_{i,M}$ or $P''_{i,M}$ is together with a component P'_j, P''_j, P'''_j, then the string generated from w is already terminal, hence the sentential form belongs to $T^*\{A_1, \ldots, A_n, A'_1, \ldots, A'_n, A''_1, \ldots, A''_n\}^*\{B, B', B''\}$ and only modifications of the auxiliary symbols are possible.

No component $P'_{i,M}$ can be in a team with P'_j, P''_j, P'''_j because all the rules $B \to B$, $B' \to B'$, $B'' \to B''$ are present and one of the symbols B, B', B'' appears in the sentential form. If $P'_{i,M}$ is used together with P'''_0, then no symbol A_j, A'_j, A''_j is present, which implies that the components P'_0, P''_0, P'''_0 were used, hence no symbol $X \in N$ has been present at that time, hence $P'_{i,M}$ cannot be used, a contradiction. Therefore, no component in group 2 can be in a team with a component in group 3.

Assume now that a team can contain $s+1$ components from one class $P_{i,M}, P'_{i,M}, P''_{i,M}$, $1 \le i \le n$, $M \subseteq \{1, 2, \ldots, n\}$. As card $(M) = s$, at least two components with different Ms appear—say P_{i,M_1}, P_{j,M_2}, $M_1 \ne M_2$. Then let $A_k \in M_1 - M_2$, $A_l \in M_2 - M_1$; P_{i,M_1} contains the rules $A_k \to A_k$, $A'_l \to A'_l$, and P_{j,M_2} contains the rules $A'_k \to A'_k$, $A_l \to A_l$. As either A_k or A'_k, as well as either A_l or A'_l is present in the obtained string, the derivation cannot be ended, hence the team is illegal.

In conclusion, the only teams containing the components in group 3 are $Q_M = \{P_{i,M} \mid i \in M\} \cup \{P_0\}$, $Q'_M = \{P'_{i,M} \mid i \in M\} \cup \{P'_0\}$, $Q''_M = \{P''_i, M \mid i \in M\} \cup \{P''_0\}$, for $M \subseteq \{1, 2, \ldots, n\}$, card$(M) = s$.

By Q_m we can transform $wA_1 \ldots A_nB$ to a string having the symbols A_i with $i \in M$ replaced by A'_i and B replaced by B'. No team $Q_{M'}$, $Q''_{M'}$ can now be used (B and B'' are not present), hence we have to continue by a team $Q'_{M'}$. If $M' \ne M$, this is not allowed (for $i \in M - M'$, A'_i is present in the string and the rule $A'_i \to A'_i$ in the components of $Q'_{M'}$). Therefore we have to use exactly the components $P'_{i,M}$ with $i \in M$, hence we use the components P_i, $i \in M$, as in Γ. Now, exactly as above, only Q''_M can be used and we return to a string of the form $zA_1 \ldots A_nB$, $z \in (N \cup T)^*$. Every such step simulates a derivation in Γ using an s-team and, conversely, every derivation in Γ can be simulated in Γ' by using the associated $(s+1)$-teams Q_M, Q'_M, Q''_M.

When we obtain a string $zA_1 \ldots A_nB$ with $z \in T^*$ we can use teams of components in group 2 plus P'''_0 for finishing the derivation. Indeed, assume $n + 1 = k(s+1) + t$, $k \ge 1$, $0 \le t \le s$. If $t = 0$, then using k suitable teams of components P'''_0, P'_i, $1 \le i \le n$, we can erase all symbols $A_1, \ldots, A_n, B$. If $k \ne 0$, then we use $k - 1$ such teams. For the remaining $s + 1 + t$ symbols (including B) we use first t teams consisting of one component P'_i and s components P''_j, P'''_r for suitable i, j, r (all i, j, r different). Thus we obtain a string containing exactly $s + 1$ nonterminals A_i, A''_i and B. Using a team composed of suitable components P'_i and P'''_0 we can end the derivation.

In conclusion, $L_t(\Gamma, s) = L_t(\Gamma', s+1)$, which proves the inclusion in the theorem. □

A counterpart of this result has been proved in Csuhaj-Varju and Păun (1993): $TCD_s(t) \subseteq TCD_2(t)$, $s \geq 2$. Consequently, for $s \geq 2$ we have $TCDF_s(t) = TCD_2(t)$, the team size induces a hierarchy with at most two levels $TCD_1(t) \subseteq TCD_2(t)$.

In order to prove the main result of this section, showing that the team feature increases the power of grammar systems also for the t-mode of derivation, we need the next two results from Rozenberg and Salomaa (1980).

Theorem V.2.10: *Let V_1, V_2 be two disjoint alphabets. Let $L_1 \subseteq V_1^+$, $L_2 \subseteq V_2^+$ and let f be a surjective function from L_1 onto L_2. Let $L = \{wf(w) \mid w \in L_1\}$.*

(i) If L is an ETOL language, then L_2 is an EDTOL language.

(ii) If L is an ETOL language and f is bijective, then L_1 is also an EDTOL language.

Corollary IV.3.4: *Let V be a finite alphabet with card(V) ≥ 2. Let k be a positive integer larger than 1. Then neither $\{w \in V^* \mid |w| = k^n$ for some $n \geq 0\}$ nor $\{w \in V^* \mid |w| = n^k$ for some $n \geq 0\}$ are EDTOL languages.*

Theorem 3: $CD(t) = TCD_1(t) \subset TCD_2(t)$.

Proof: The equality follows from definitions, the inclusion from the previous theorem. In order to prove its properness we consider the following grammar system, similar to some extent with the second example in Section 3:

$$\Gamma = (\{A, B, A', B'\}, \{a, b, c, d, e\}, P_1, P_2, P_3, P_4, P_5, P_6, AB),$$

$$P_1 = \{A \to A'A'\},$$

$$P_2 = \{B \to aBc, B \to bBd, B \to aB'c, B \to bB'd\},$$

$$P_3 = \{A' \to AA\},$$

$$P_4 = \{B' \to aB'c, B' \to bB'd, B' \to aBc, B' \to bBd\},$$

$$P_5 = \{A \to e, A' \to e, B \to B, B' \to B'\},$$

$$P_6 = \{B \to B, B \to ac, B \to bd, B' \to B', B' \to ac, B' \to bd, A \to A, A' \to A'\}.$$

We obtain

$$L_t(\Gamma, 2) = \{e^{2^n} wf(w) \mid n \geq 0, w \in \{a, b\}^+, |w| = 2^n\},$$

where $f(w) = h(\mathrm{mi}(w))$, with mi denoting the mirror image and h the morphism defined by $h(a) = c$, $h(b) = d$.

Indeed, when using P_5, no symbol B, B' can be present at the end of that step, when using P_6 no symbol A, A' can remain in the current sentential form. Therefore P_5, P_6 can participate in a team only together, namely at the last step of a derivation. Now P_1 cannot be in the same team with P_3, and P_2 cannot be in the same team with P_4. Starting from a string containing only symbols A and B (initially we have AB), besides the team $\{P_5, P_6\}$, only the team $Q_1 = \{P_1, P_3\}$ is applicable, leading to a string containing only symbols A' and B'. Now only $Q_2 = \{P_3, P_4\}$ can be applied, leading again to a string containing the nonterminals A, B. In such a cycle the number of occurrences of A is doubled and for each use of a rule $A \to A'A'$ or $A' \to AA$, one rule in P_2 or in P_3 is used; therefore the number of symbols a, b and c, d, respectively, equals the number of occurrences of A, A' subtracted by one. Finally, every A, A' is replaced by e, but in this time rules $B \to B$, $B' \to B'$ are used in P_6, excepting the last moment when one of $B \to ac$, $B \to bd$, $B' \to ac$, $B' \to bd$ is used. In this way, the number of symbols a, b (hence also c, d) equals the number of occurrences of e, hence we have the mentioned language $L_t(\Gamma, 2)$.

Erase now by a morphism the symbol e. If $L_t(\Gamma, 2) \in ET0L = CD(t)$, then the obtained language, $\{wf(w) \mid w \in \{a, b\}^+, |w| = 2^n, n \geq 0\}$ would be in $ET0L$; according to the above quoted Theorem V.2.10 (ii) it follows that $\{w \in \{a, b\}^+ \mid |w| = 2^n, n \geq 0\} \in EDT0L$ (the mapping f is a bijection from $L_1 = \{w \in \{a, b\}^+ \mid |w| = 2^n, n \geq 0\}$ to $L_2 = \{\{w \in \{c, d\}^+ \mid |w| = 2^n, n \geq 0\})$, which contradicts Corollary IV.3.4 in Rozenberg and Salomaa (1980). In conclusion, $L_t(\Gamma, 2) \notin ET0L$. □

Summarizing the previous theorems and the result in Csuhaj-Varju and Păun (1993), we have

$$ET0L = CD(t) = TCD_1(t) \subset TCD_2(t) = TCD_s(t),\ s \geq 2.$$

Note that the system Γ in the preceding proof is λ-free.

Theorem 4: *If Γ is a λ-free CD grammar system and $s \geq 1$, then $L_t(\Gamma, s)$ can be generated by a λ-free matrix grammar with appearance checking.*

Proof: Take $\Gamma = (N, T, P_1, \ldots, P_n, w)$, denote $\mathrm{dom}(P_i) = \{A \mid A \to x \in P_i\}$, $1 \leq i \leq n$, and construct the matrix grammar

$$G = (N', T \cup \{c\}, S, M, F),$$

where

$$N' = \{S, X, Z\} \cup \{A, A' \mid A \in N\} \cup \{[i_1, \ldots, i_s] \mid 1 \leq i_j \leq n,\ 1 \leq j \leq s\},$$

and M contains the following matrices:

1. $(S \to wX)$.
   (We start by introducing the axiom w and the control symbol X.)
2. $(X \to [i_1, \ldots, i_s])$, $[i_1, \ldots, i_s] \in N'$.
   (The symbol X introduces an s-tuple $[i_1, \ldots, i_s]$ which will determine the derivation in the team $\{P_{i_1}, \ldots, P_{i_s}\}$.)
3. $([i_1, \ldots, i_s] \to [i_1, \ldots, i_s], A_1 \to x_1', \ldots, A_s \to x_s')$,
   for $[i_1, \ldots, i_s] \in N'$, $A_j \to x_j \in P_{i_j}$ and x_j' is obtained by replacing all nonterminals $B \in N$ appearing in x_j by primed B', $1 \leq j \leq s$.
   $(A' \to A)$, $A \in N$.
   (The derivation in the s-team $\{P_{i_1}, \ldots, P_{i_s}\}$ is simulated in this way.)
4. $([i_1, \ldots, i_s] \to X, A_1 \to Z, A_1' \to Z, \ldots, A_r \to Z, A_r' \to Z)$,
   for $[i_1, \ldots, i_s] \in N'$, $\{A_1, \ldots, A_r\} = \bigcup_{j=1}^{s} \mathrm{dom}(P_{i_j})$.
   (The derivation in a team $\{P_{i_j}, \ldots, P_{i_s}\}$ is correctly ended when no symbol can be rewritten by a rule of some P_{i_j}, $1 \leq j \leq n$.)
5. $(X \to c)$.

The set F, consisting of rules used in the appearance checking manner, contains all rules of the form $A \to Z$, $A' \to Z$ in group 4 of the rules (Z is a trap-symbol which blocks the derivation if introduced).

From the explanation above it is easy to see that $L(G) = L_t(\Gamma, s)\{c\}$. As the family of languages generated by λ-free matrix grammars with appearance checking is closed under restricted morphisms (see Dassow and Paun 1989) the proof is complete. □

5. Closure properties

Generative capacity in itself does not imply strong closure properties. We have already observed the generative power of team CD grammar systems. Therefore, the following result is particularly pleasing.

Theorem 5: *The family TCD(t) is a full AFL.*

Proof: *Union.* Given $L, L' \in TCD(t)$, in view of Theorem 2 we may assume $L, L' \in TCD_s(t)$ for a given s. Then take two systems, Γ, Γ', with $L_t(\Gamma, s) = L$, $L_t(\Gamma', s) = L'$, $\Gamma = (N, T, P_1, \ldots, P_n, w)$, $\Gamma' = (N', T, P'_1, \ldots, P'_m, w')$. Assume $N \cap N' = \emptyset$ and construct the system

$$\Gamma'' = (N \cup N' \cup \{X_1, \ldots, X_s\}, T, P_1, \ldots, P_n, P'_1, \ldots, P'_m, P''_1, \ldots, P''_s, X_1 \ldots X_s),$$

where P_i, P'_j are the components of Γ, Γ' and

$$P''_1 = \{X_1 \to w, X_1 \to w'\},$$

$$P''_i = \{X_i \to \lambda\}, 2 \le i \le s.$$

As the components of Γ cannot be in the same team with the components of Γ' or with P''_i, $1 \le i \le s$, and by using P''_1 (in a team with $P''_2, \ldots, P''_s$) we obtain either w or w', we have $L_t(\Gamma'', s) = L_t(\Gamma, s) \cup L_t(\Gamma', s)$.

Concatenation. Start again from Γ, Γ' as above and construct

$$\Gamma'' = (N \cup N' \cup \{X_1, \ldots, X_s\}, T, P_1, \ldots, P_n, P'_1, \ldots, P'_m, P''_1, \ldots, P''_s, wX_1X_2 \ldots X_s),$$

with

$$P''_1 = \{X_1 \to w'\} \cup \{X_j \to X_j \mid 2 \le j \le s\} \cup \{X \to X \mid X \in N\},$$

$$P''_i = \{X_i \to \lambda\} \cup \{X_j \to X_j \mid 1 \le j \le s\} \cup$$
$$\cup \{X \to X \mid X \in N\}, 2 \le i \le s.$$

The components $P''_1, \ldots, P''_s$ can be used only together, after obtaining a string of the form $xX_1 \ldots X_s$, $x \in T^*$, hence the derivations in Γ, Γ' are separated by a step using the team $\{P''_1, \ldots, P''_s\}$. This ensures the equality $L_t(\Gamma'', s) = L_t(\Gamma, s)L_t(\Gamma', s)$.

Morphisms. Obvious (replace each symbol a by $h(a)$ in all rules and in the axiom).

Kleene +. Take $\Gamma = (N, T, P_1, \ldots, P_n, w)$ and construct

$$\Gamma' = (N \cup \{X_1, \ldots, X_s, Y_1, \ldots, Y_s\}, T, P'_1, \ldots, P'_n, P''_1, \ldots, P''_s,$$
$$P'''_1, \ldots, P'''_s, wX_1X_2 \ldots X_s),$$

where

$$P'_i = P_i \cup \{Y_j \to Y_j \mid 1 \le j \le s\}, 1 \le i \le n,$$

$$P_1'' = \{X_1 \to wY_1 \dots Y_s, X_1 \to \lambda\} \cup \{X \to X \mid X \in N\} \cup$$
$$\cup \{X_j \to X_j \mid 1 \le j \le s\},$$
$$P_i'' = \{X_i \to \lambda\} \cup \{X \to X \mid X \in N\} \cup$$
$$\cup \{X_j \to X_j \mid 1 \le j \le s\}, \ 1 \le i \le s,$$
$$P_i''' = \{Y_i \to X_i\} \cup \{Y_j \to Y_j \mid 1 \le j \le s\} \cup$$
$$\cup \{X \to X \mid X \in N\}, \ 1 \le i \le s.$$

No component P_i''' can be used in the same team with a component P_j'' due to the appearance of rules $X_r \to X_r$, $1 \le r \le s$, in P_j'' for all j. We start from a string with s symbols $X_1, \dots, X_s$; if we use less than s components $P_1'', \dots, P_s''$, then the derivation cannot be ended, as all rules $X_j \to X_j$, $1 \le j \le s$, are present. Therefore we pass to a string $zwY_1 \dots Y_s$, $z \in L_t(\Gamma, s)$. Now the only possible team to use is $\{P_1''', \dots, P_s'''\}$, hence we obtain $zwX_1 \dots X_s$ and the process is iterated. When using the rule $X_1 \to \lambda$ in P_1'', the derivation is finished. In conclusion, $L_t(\Gamma', s) = L_t(\Gamma, s)^+$.

Intersection with regular sets. For $\Gamma = (N, T, P_1, \dots, P_n, w)$, $w = w_1A_1w_2 \dots w_kA_kw_{k+1}$, $w_i \in T^*$, $1 \le i \le k+1$, $A_i \in N$, $1 \le i \le k$, and $\rho = (Q, T, q_0, F, \delta)$ a finite automaton, consider all strings of the form

$$w_1(q_1, A_1, q_1')w_2(q_2, A_2, q_2') \dots w_k(q_k, A_k, q_k')w_{k+1},$$

where $q_1 = \delta(q_0, w_1)$, $q_i = \delta(q_{i-1}', w_i)$, $2 \le i \le k$, $\delta(q_k', w_{k+1}) \in F$, for all possible $q_1', \dots, q_k' \in Q$.

For each such string, consider a CD grammar system having this string as the axiom, T as the terminal alphabet, the nonterminal alphabet $Q \times N \times Q$, and the components $P_1', \dots, P_n'$ constructed as follows: each rule $A \to x_1A_1x_2 \dots x_rA_rx_{r+1}$, $x_1 \in T^*$, $1 \le i \le r+1$, $A_i \in N$, $1 \le i \le r$, $r \ge 1$, in P_i is replaced by the set of rules

$$(q, A, q') \to x_1(q_1, A_1, q_1')x_2(q_2, A_2, q_2') \dots x_r(q_r, A_r, q_r')x_{r+1},$$

where $q_1 = \delta(q, x_1)$, $q_i = \delta(q_{i-1}', x_i)$, $2 \le i \le r$, $\delta(q_r', x_{r+1}) = q'$, for all possible $q, q', q_1', \dots, q_r'$ in Q.

Moreover, each rule $A \to x$, $x \in T^*$, is replaced by all rules

$$(q, A, q') \to x, \ q' = \delta(q, x), \ q, q' \in Q.$$

One can see that a derivation in such a new system corresponds to a derivation in Γ, observing also the restriction imposed by ρ (because δ is defined for all pairs $(q, a) \in Q \times T$, all rules $(q, A, q') \to z'$ are present, for all $q, q' \in Q$, hence a derivation in the new system cannot be correctly finished if the corresponding derivation in Γ is not correctly finished). As the family $TCD(t)$ is closed under union, we obtain that $L_t(\Gamma, s) \cap L(\rho)$, which is equal to the union of all languages $L_t(\Gamma', s)$ with Γ' denoting the new systems, is in $TCD(t)$.

Inverse morphisms. The closure under inverse morphisms follows from the closure under intersection with regular sets, substitution by (λ-free) regular sets and (restricted) morphisms; therefore it is enough to prove the closure under substitution with regular sets.

Take $\Gamma = (N, T, P_1, \dots, P_n, w)$ and $\sigma: T^* \to 2^{T^*}$ with $\sigma(a) = L_a$ regular, $L_a = L(G_a)$, for $G_a = (N_a, T, P_a, S_a)$, $a \in T$, regular grammars. We assume $N_a \cap N_b = \emptyset$ for all $a \ne b$ in T and construct the system Γ' with the terminal alphabet T, the nonterminal alphabet

$$N \cup \{A' \mid A \in N\} \cup \bigcup_{a \in T} N_a \cup \{X_1, \ldots, X_s, Y_1, \ldots, Y_s, Y'_1, \ldots, Y'_s\},$$

the axiom string $wX_1 \ldots X_s$ and the following components:

$$P'_i = P_i \cup \{Y_j \to Y_j, Y'_j \to Y'_j \mid 1 \leq j \leq s\}, 1 \leq i \leq n,$$
$$P''_i = \{X_i \to Y_i\} \cup \{X \to X \mid X \in N\} \cup$$
$$\cup \{X_j \to X_j \mid 1 \leq j \leq s\}, 1 \leq i \leq s,$$
$$P'_a = P_a \cup \{A \to A, A \to A' \mid A \in N_a\} \cup$$
$$\cup \{X_i \to X_i \mid 1 \leq i \leq s\}, a \in T,$$
$$P''_a = \{A' \to A', A' \to A \mid A \in N_a\} \cup \{X_i \to X_i \mid 1 \leq i \leq s\}, a \in T,$$
$$P'''_i = \{Y_i \to Y_i, Y_i \to Y'_i\} \cup \{X_j \to X_j \mid 1 \leq j \leq s\}, 1 \leq i \leq s,$$
$$P^{iv}_i = \{Y'_i \to Y'_i, Y'_i \to Y_i\} \cup \{X_j \to X_j \mid 1 \leq j \leq s\}, 1 \leq i \leq s,$$
$$P^{v}_i = \{Y_i \to \lambda\} \cup \{X \to X \mid X \in \bigcup_{a \in T} N_a\} \cup$$
$$\cup \{Y_j \to Y_j, Y'_j \to Y'_j \mid 1 \leq j \leq s\}, 1 \leq i \leq s,$$
$$P^{vi}_i = \{Y'_i \to \lambda\} \cup \{X \to X \mid X \in \bigcup_{a \in T} N_a\} \cup$$
$$\cup \{Y_j \to Y_j, Y'_j \to Y'_j \mid 1 \leq j \leq s\}, 1 \leq i \leq s.$$

The components P'_i cannot be used in the presence of symbols Y_j, Y'_j, none of the other components can be used in the presence of symbols X_j, hence each derivation has two phases, an initial one when we pass from $wX_1 \ldots X_s$ to some $zX_1 \ldots X_s$, $z \in L_t(\Gamma, s)$, and another one when z is replaced by some string in $\sigma(z)$ in the presence of $Y_1, \ldots, Y_s$. The components P''_i can be used only together. The components P'''_i, P^{iv}_i, by passing from Y_i to Y'_i and conversely, are used in order to complete s-teams together with components P'_a and P''_a. Thus all derivations in G_a can be simulated in Γ' too. Finally, the symbols Y_i, Y'_i are removed at the same time, by a team composed of components P^{v}_i, P^{vi}_i. In conclusion, $\sigma(L_t(\Gamma, s)) = L_t(\Gamma', s)$, which completes the proof. □

Acknowledgements

Research supported by the Academy of Finland, grant 11281, and the Alexander von Humboldt Foundation.

References

Atanasiu, A. and Mitrana, V. (1989) The modular grammars. *International Journal of Computer Mathematics*, **30**: 101–122.

Csuhaj-Varju, E. and Dassow, J. (1990) On cooperating distributed grammar systems. *Journal of Information Processing and Cybernetics, EIK*, **26**: 49–63.

Csuhaj-Varju, E., Dassow, J., Kelemen, J. and Păun, Gh. (1994) *Grammar Systems* (London: Gordon and Breach).

Csuhaj-Varju, E. and Kelemen, J. (1989) Cooperating grammar systems: a syntactical framework for the blackboard model of problem solving. In I. Plander (ed.) *Proceedings of Artificial Intelligence and Information, Control Systems, and Robots* (Amsterdam: North-Holland) pp. 121–127.

Csuhaj-Varju, E. and Păun, Gh. (1993) Limiting the team size in cooperating grammar systems. *Bulletin EATCS*, **51**: 198–202.

Dassow, J. and Kelemen, J. (1991) Cooperating distributed grammar systems: a link between formal languages and artificial intelligence. *Bulletin EATCS*, **45**: 131–145.

Dassow, J. and Păun, Gh. (1989) *Regulated Rewriting in Formal Language Theory* (Berlin, Heidelberg: Springer-Verlag).

Kelemen, J. (1991) Syntactical models of distributed cooperative systems. *Journal of Experimental and Theoretical Artificial Intelligence*, **3**: 1–10.

Kelemen, J. and Kelemenova, A. (1992) A subsumption architecture for generative symbol systems. In R. Trappl (ed.) *Cybernetics and Systems Research '92, Proceedings of 11th European Meeting on Cybernetics and Systems Research* (Singapore: World Scientific) pp. 1529–1536.

Kelemenova, A. and Csuhaj-Varju, E. (1992) Languages of colonies. In *2nd International Colloquium on Words, Languages, and Combinatories*, Kyoto, Japan.

Kleijn, H. C. M. and Rozenberg, G. (1980) A study in parallel rewriting systems. *Information and Control*, **44**: 134–163.

Meersman, R. and Rozenberg, G. (1978) Cooperating grammar systems. In *Proceedings of the Mathematical Foundations of Computer Science '78 Symposium, Lecture Notes in Computer Science 64* (Berlin, Heidelberg: Springer-Verlag) pp. 364–376.

Nii, P. H. (1989) Blackboard systems. In A. Barr, P. R. Cohen and E. A. Feigenbaum (eds) *The Handbook of Artificial Intelligence*, vol. 4 (Reading, MA: Addison-Wesley).

Păun, Gh. and Sântean, L. (1989) Parallel communicating grammar systems: the regular case. *Annals of the University of Bucharest, Series Mathematics-Informatics*, **38:** 55–63.

Rozenberg, G. and Salomaa, A. (1980) *The Mathematical Theory of L Systems* (New York: Academic Press).

Salomaa, A. (1973) *Formal Languages* (New York, London: Academic Press).