Grammar Systems: A Short Survey

Erzsébet Csuhaï-Varjú
Computer and Automation Research Institute
Hungarian Academy of Sciences
H-1111 Budapest, Kende u.13-17.
csuhaï@sztaki.hu

1 Introduction

“Grammar Systems is a recent field of formal language theory providing syntactic models and frameworks for describing and studying (the behavior of) multi-agent systems at symbolic level.” A few years ago, readers interested in the area often found this or similar sentences in papers published about grammar systems. One decade after the publication of the monograph “Grammar systems: A grammatical approach for distribution and cooperation” [24] and after more than fifteen years research in the field, we can state that Grammar Systems has become a well-recognized field of formal language theory.

Several scientific areas have inspired and influenced the developments: distributed and decentralized artificial intelligence, distributed and parallel computing, artificial life, molecular computing, robotics, ecology, sociology, etc. Computer networks, parallel and distributed computer architectures, distributed and cooperative text processing, natural language processing are candidates for possible applications. However, the relation is two-way: ideas or concepts of grammar systems theory have also inspired and influenced the development of some unconventional computational models in contemporary computer science or research areas being on the boundary of computer science.

But what is a grammar, and what does it mean “system” in this term? The notion grammar refers to any kind of language determining devices, it can be a usual Chomsky grammar, it can be an automaton, but it can also be a splicing system, that is, a rewriting system from molecular computing.

A grammar system consists of several (a finite number) of language determining devices (language processors) which jointly change a common symbolic environment (usually, a string or a finite set of strings) by applying language theoretic operations to it. The symbolic environment can be shared by the components of the system or it can be given in the form of a collection of separated sub-environments, each belonging to a language processor. At any moment of time, the state of the system is represented by the current string describing the environment (the collection of strings of the sub-environments). The functioning of the system is realized by changes of its states. Depending on the variant of multi-agent systems which is represented by the actual grammar system, in addition to performing derivation steps,
the language processors communicate with each other. Usually, this is done by exchange of strings which can be data (for example, sentential forms in derivation) or programs (productions or coded form of some operation).

The behavior of the grammar system can be characterized in different manners. It can be represented by the set of sequences of environmental states following each other starting from an initial state or by the set of all states of the environment or that of a sub-environment which originate from the initial state and satisfy certain criteria. The second case defines the language of the system.

According to the traditional approach in formal language theory, one language is generated by one grammar (produced by one language determining mechanism), while according to the non-standard approach provided by Grammar Systems, generation or accepting is performed by several grammars, in cooperation, by a distributed system of language processors. Thus, grammar systems are both syntactic models of multi-agent systems and distributed models of language.

Major problems in the theory can be formulated as follows: to learn whether or not distribution and cooperation enhance the computational power of language processors, to know whether or not grammar systems decrease the complexity of language specification comparing to the language specification given by single grammars, and furthermore, to learn whether or not phenomena characteristic for multi-agent systems (cooperation, distribution, communication, parallelism, emergent behavior, etc.) can be expressed and formalized by tools of grammar systems theory, and if the answer to the last question is positive, which are the proper formalizations.

Although this list of problems is a short one, these items cover a long-long list of subproblems, among them we find questions closely related to basic problems of formal language theory.

To give a picture about the research directions in the area, without the aim of completeness, we list some important frameworks and models and we call the attention to some new problems or research areas which are - according to our opinion - worth studying. The interested reader can find detailed information on grammar systems theory in the monograph [24] and in the book chapter [54], or in the articles listed in the on-line bibliography [43].

For more details concerning the motivations, the background, and connections to the theory of multi-agent systems, artificial intelligence, and artificial life the reader is advised to consult [74, 75, 76, 77, 78, 79].

For possible applications of the theory in linguistics, for connections to natural language generation and modeling the reader is referred to [18, 84, 69, 70, 71].

2 Models, Results, and Problems

2.1 Cooperating Distributed Grammar Systems

Grammar systems theory started in 1988 by introducing cooperating distributed grammar systems (CD grammar systems) for modeling syntactic aspects of the blackboard model of problem solving [22]. The concept and the formal analogy between CD grammar systems and blackboard architectures were discussed and developed further in [27]. Other main inspirations to formalize the concept was the
Grammar systems: A short survey

demand to develop a mathematically well-founded, but sufficiently simple tool for
language support of distributed text processing, distributed document creation, and
tele-conferencing. We should note, however, that the term “cooperating grammars”
was first introduced in [88] as a notion for extending the two-level substitution
mechanisms of grammars to a multi-level concept and modeling concurrent operat-
ing systems. A notion, related to cooperating grammar systems, called the modular
grammar, based on the concept of modularity and motivated by regulated rewriting,
was introduced in [4].

A cooperating distributed grammar system is a finite set of (usually generative)
grammars which cooperate in deriving words of a common language. At any mo-
ment of time, there is exactly one sentential form in generation. The component
grammars generate the string by turns, under a cooperation protocol, called the
derivation mode. In this model, the cooperating grammars represent independent
cooperating problem solving agents which jointly solve a problem by modifying the
contents of a global database, called the blackboard which is used for storing in-
formation on the problem solving process. In blackboard architectures the agents
communicate with each other only through the blackboard, that is, there is no direct
communication among them. According to the grammatical framework, the actual
contents of blackboard is represented by the sentential form and the generated words
correspond to the solutions of the problem.

Most of the investigations in the theory of CD grammar systems have focused
on studying the question whether or not cooperation enhances the computational
power of the generative capacity of the individual grammars, and, if the answer is
positive, how much simple grammars are able to obtain this capacity. Cooperation
protocols are decisive factors from this point of view. By the original motivation, the
distributed problem solving, most of the studied protocol variants are based on the
so-called competence of the grammars, that is, on the capability of the component
grammar to perform a derivation step on the actual sentential form. This is usually
formalized as the number of different nonterminals (letters) the grammar is able
to rewrite in the given string (mainly context-free CD grammar systems have been
examined). Most of these protocols are defined via start/stop conditions for the
components: these conditions prescribe the competence level of the grammar at
starting the derivation and either determine its competence level at the end of its
the work or prescribe a number of derivation steps the grammar has to perform in
succession (with or without having a certain competence level).

Another important goal of the theory has been to develop protocols where the
grammars demonstrate fair behavior, in the sense that their contribution to the
generation is almost the same (the number of the performed derivation steps or the
number of activation is almost the same for any grammar).

The achieved results demonstrate that CD grammar systems even with simple
components and simple protocols form powerful computational devices. Large lan-
guage classes (the ETOL language class, the class of programmed languages with
appearance checking) can be described in terms of systems of a small number of
very simple cooperating language determining devices which use relatively very sim-
ple, competence-based cooperation protocols, most of them even a protocol requiring
fair behavior.

143
For results about the generative capacity of CD grammar systems using the so-called basic derivation modes, the reader is advised to consult [22, 24, 54, 105, 103]. Other characteristics of the basic, generative model and its different variants have been discussed in various articles. Without the aim of completeness, we mention some of these topics: determinism in CD grammar systems [49, 92], comparisons of systems with components using homogeneous (the same) and hybrid (different for each grammar) cooperation strategies [93, 103, 60, 61], variants of competence-based cooperation protocols [22, 51, 24, 9, 88, 11], similarity of the components [106, 94], hierarchies, priorities, stratification among them [95, 6, 25], fairness in cooperation [46]. CD grammar systems with components working with dynamical start/stop (context) conditions were examined in [23], while CD grammar systems controlled by graphs were considered in [44, 24].

Classical language theoretic questions as descriptional complexity [53, 106] and decidability properties [91] have also been studied.

The original model was introduced for generative grammars, but the idea of the concept has been extended and applied to other computational mechanisms: accepting grammars [58, 59], array grammars [45], limited OL (Lindenmayer) systems [125], automata [48, 82, 32], tree processing devices [63], etc.

Further, powerful extensions of the original notion, as CD grammar systems with registers were studied in [52, 123], and CD grammar systems with communication aided by a generalized sequential machine (a “translator”) was the topic of article [124]. Several papers investigated properties of stream X-machines by Eilenberg based on cooperating distributed grammar systems (see [5] for the basic notions and results), and probabilistic CD grammar systems were examined in [3].

The reader can observe that the dominating research areas in the theory of CD grammar systems are the generative capacity and the size complexity of these constructs.

As topics for future research, it would be worth studying the applicability of these systems in solving well-known algorithmic problems, and to investigate cooperation complexity and communication complexity of CD grammar systems. Although steps have already been made to formulate proper notions for cooperation complexity or collective complexity, the problem and the area still needs further elaboration. Similarly the notions of competence and fairness need further elaboration: how these concepts relate to other notions used in the theory of agents or in concurrency theory? Handling incomplete information can also be of particular interest in the theory. Suppose that the problem solving agent has incomplete information on the contents of the blackboard. Has this fact some influence on the problem solving capacity of the problem solving community? This problem can be interpreted in a very nice manner in CD grammar systems theory: some parts of the sentential form are not accessible for the active grammar. How much is the minimal amount of information the agent must have to effectively take part in the problem solving? Another interesting and promising fields of research would be examining CD grammar systems working on multisets of strings and studying how concepts of evolutionary computing can be interpreted in terms of CD grammar systems. Finally, a natural, but yet not explored area is the investigation of the connections between CD grammar systems and games.
While CD grammar systems are sequential computing devices, team grammar systems with simultaneous actions of certain grammars (teams) on the sentential form, introduce parallelism in the model [73]. These systems, among other things, demonstrate the equivalent computational power of programmed grammars with appearance checking and freely chosen grammar teams with a very limited number of components (pairs of grammars) which work with an important variant of competence-based derivations [108]. For further important results and different variants the reader is advised to consult [7, 8, 86].

Team grammar systems provide a wide area of open problems as well. What about the sequence of teams following each other in the course of the generation? Are there problems which can be solved with teams working periodically, that is, with periodical sequences of teams that follow each other in the course of the derivation? What about problems where these team sequences are aperiodical? What about the dynamism of the size, the simplicity of the components, and behavior of teams that are formed according to the actual competence of the grammars? How concepts of evolutionary computing can be implemented in this area? And, last but not least a general question: are team grammar systems more effective in problem solving than CD grammar systems (for example, according to their derivation complexity) and, if this is the case, when?

2.2 Colonies and Eco-Grammar Systems

Colonies are important concepts in grammar systems theory that describe language classes in terms of collections of very simple, purely reactive, situated agents with emergent behavior [80, 81]. In this model the agents are represented by very simple regular grammars (each grammar generates a finite language) which generate a common sentential form. The basic variant of colonies determines the context-free language class, while the more sophisticated models (competition among the agents, agents with point mutation operations) represent computational tools with considerably enhanced computational power [47, 87]. The interested reader can find several other interesting models and results in the articles referred in the bibliography [43].

For future research, further elaboration of notion of emergence in terms of formal language theory would be fundamental in the area. In addition, this field appears to be particularly suitable for finding applications, especially in robotics.

Eco-grammar systems form a language theoretic framework for modeling ecosystems: developing agents, represented by L systems, in a dynamically changing population, interact with each other and with their shared evolving symbolic environment (See [28] and [29] for basic information). These constructions were motivated by artificial life and provide tools for describing life-like phenomena (birth, death, hibernation, overpopulation, pollution, etc.) in terms of formal grammars and languages. For detailed information on the results obtained in this area and in the theory of colonies the reader to the articles listed in the on-line annotated bibliography [43]. While according to the original concept of eco-grammar systems the agents perform local actions on the environment, so-called conditional tabled eco-grammar systems are models with indirect interaction between the agents and the environment. This subfield has been extensively explored, see [33, 34, 114, 115, 116, 117] for further
information.

Eco-grammar systems are very promising from the point of view of future investigations. Firstly, the relation of the different variants of eco-grammar systems and that of evolutionary computing are of worth analyzing, but the general concept of eco-grammar systems can be related to several other scientific areas: modeling economy, modeling society, modeling the World Wide Web. Valuable research has started in modeling culture and pragmatics in [70] and investigations of the relations between dynamical systems and eco-grammar systems have been started in [14].

2.3 Network of language processors

Networks of language processors are one of the basic areas in the theory of grammar systems. In these constructs, language processors, that is, grammars or other language determining devices are located in nodes of a network (a virtual graph). Each processor works on its own sentential form (on its own collection of sentential forms) and informs the others about its activity by communicating strings which can be data and/or programs. Rewriting and communication take place alternately. The system functions (usually) in a synchronized manner.

It is easy to observe the difference between CD grammar systems and these architectures: while in the first case the grammars generate one common string, in the second case each of the components operates on its own string. There are several important different models in the area, “networks of language processors” as a collective term was introduced in [36, 19, 37], where basic characteristics and basic variants of the framework were discussed and several research directions were proposed.

Parallel communicating grammar systems (or PC grammar systems, for short), the first models in the area of networks of language processors, form a very important, highly elaborated field, according to the original definition with Chomsky grammars at the nodes. These are networks of grammars with components communicating strings by emerging requests.

The concept was introduced in [110] (and continued in [111, 99]), as a grammatical representation of the so-called “classroom model” of problem solving, which is a modification of the blackboard architecture. It consists of several agents which jointly solve a problem. Each agent has its own “notebook” containing the description of a particular subproblem of the given problem and each agent operates only on its own “notebook” and one distinguished agent, the master, operates on the “blackboard”. This agent has the description of the whole problem and it decides when the problem is solved. Moreover, the agents communicate with each other by requests concerning the contents of their “notebooks.” According to this model, the agents can be considered as pupils in a classroom and the master is their classroom leader or teacher.

According to the grammatical model, each agent is represented by a grammar which generates its own sentential form. In each time unit, every grammar performs a rewriting step on its own sentential form (modifies its own notebook) and communication is done by requests through so-called query symbols, one different symbol referring to each grammar in the system. When a query symbol appears in the sen-
tential form of a grammar, then the rewriting process stops and one or more communication steps are performed by replacing all occurrences of the query symbols with the current sentential forms of the queried component grammars, supposing that these strings are query-free. When no more query symbol is present in any of the sentential forms, then the rewriting process starts again. In so-called returning systems after communicating its current sentential form the queried component returns to its start symbol and begins to generate a new string. In non-returning systems the components continue the rewriting of their sentential form. The language (the set of problem solutions) of the system is the set of terminal words which appear as sentential forms at the master.

Parallel communicating grammar systems, as cooperating distributed grammar systems, have been the subject of detailed study during the years. The investigations concentrated on the generative power of these systems and on examining how this power is influenced by changes in the basic characteristics of the system: the way of communication and synchronization among the components and their way of functioning. The results, as in the case of CD grammar systems, demonstrate that PC grammar systems with very simple component grammars and with bounded size are convenient tools for describing large language classes. For example, non-context-free context-sensitive languages can be generated by PC grammar systems with two regular components, and returning PC grammar systems with a few number of context-free component grammars are sufficient to generate any recursively enumerable language. For information concerning the generative power of PC grammar systems the reader is referred to [24, 54] and the articles listed in [43]. For details about the generative capacity and size of systems with regular or right-linear components consult [110, 122, 72], results on systems with context-free components are presented in [39, 83], and systems with context-sensitive components are discussed in [113, 62].

Normal forms for context-free PC grammar systems are presented in [40, 42], and normal forms for systems with right-linear components are given in [56].

The very important feature of context-free (linear, regular) PC grammar systems that the returning systems simulate the non-returning systems was pointed out in [55, 120]. Synchronization problems, that is, PC grammar systems with additional synchronization mechanisms or unsynchronized systems, were studied in [98, 102, 40]. Different ways of communication and their impact on the generative power were examined in [40], and different ways of defining the language of the system (so-called popular and competitive systems) were proposed and discussed in [119]. An important aspect, namely the case of incomplete information communication, where the grammars communicate subwords of their sentential form by request, was studied in [41].

Communication complexity problems of PC grammar systems were investigated in [96, 97, 64, 65, 66, 68], their computational complexity aspects were discussed in [1, 16, 15], while results concerning the descriptonal complexity of these constructs were presented in [100, 24]. Some decision problems concerning PC grammar systems were examined in [118].

Several other interesting variants were proposed and investigated: PC grammar systems with query words [89], with separated alphabets [90], with signals [112],
with negotiations [107], probabilistic PC grammar systems [2], etc.

An area, with distinguished importance is the theory of parallel communicating L systems, started in [101]. Several interesting results were obtained in this field, among other things it was shown that centralized PC Lindenmayer systems with $D0L$ components can be simulated with $EDT0L$ systems [101], and that systems with $E0L$ and $ET0L$ components are equally powerful both in the returning and in the non-returning case, [121]. For detailed information, the reader should consult [24] and the articles in the on-line annotated bibliography [43].

As in the case of CD grammar systems, the concept of PC grammar systems has been extended to parallel communicating automata systems. For the finite automata models we refer to [85], for basic results on parallel communicating pushdown automata systems consult [31].

For future research, interesting an important topics in the theory of PC grammar systems can be studying PC grammar systems with asynchronously working components, and further variants with incomplete information communication. For example, models where the grammars are allowed to communicate only subwords with a certain fixed length or with a bounded length, or where the grammars split the sentential form into as many pieces as were requested by the other grammars. Another promising directions can be the implementations of well-known distributed algorithms in this framework and the comparison of the efficiency of the different methods.

While parallel communicating grammar systems form a network of Chomsky grammars communicating by the dynamically emerging requests, CCPC grammar systems, that is, parallel communicating grammar systems with communication by command, represent another communication philosophy. In this case, after each rewriting step (after performing a prescribed sequence of rewriting steps) the Chomsky grammars communicate with each other by exchanging information which is represented by their current sentential form or a subword of it. The communicated strings are filtered: each component is associated with a so-called selector (filter) language. A communicated string is accepted by the component if it is an element of its selector language. After communication, a new string is composed from the accepted ones and the grammar begins to generate this word. For more details about the motivations beyond the idea and for the possible different variants the reader is referred to [30]. A basic and surprising result concerning these constructions is, that CCPC grammar systems with three regular grammars and regular filter languages are able to generate any recursively enumerable language [67].

Similar architectures, that is, networks of language processors communicating by command with the aid of filter languages were studied having $0L$ systems as components in [36], called networks of parallel language processors. Test tube distributed systems based on splicing [26] and networks of Watson-Crick $D0L$ systems [21, 38] are also examples for networks of language processors communicating by command, with components inspired by the behavior and properties of DNA, namely DNA recombination and the phenomenon of Watson-Crick complementarity. Investigations in a framework, called networks of evolutionary processors, where the language determining devices are simple rewriting systems with point mutation operations, have recently started with interesting results. (See for the basic notions [17]).
Grammar systems: A short survey

According to these latter, bio-inspired constructs, promising and challenging research areas can be studying the population dynamics of the objects (strings, symbols, etc.) in the network, supposing that the language processors operate with multisets of strings. Phenomena characteristic for real networks, namely overloaded situations at the nodes, black-holes, chaotic-like events, and wave-like phenomena would also worth formalizing and studying.

Since networks of language processors are closely related to membrane systems, bridges between membrane systems theory, that is, P systems theory should also be built.

3 Conclusions

Grammar Systems is a wide and rich area for further investigations, both concerning classical problems of formal language theory and non-standard problems as descriptions of multi-agent systems in terms of distributed systems of grammars. One of the most interesting and promising problem area would be the study of the properties of grammar systems with dynamically changing (evolving, adapting) components. Similarly, open networks of language processors, that is, networks with dynamically changing number of components would be worth studying. A promising research area is the study of grammar systems operating over multisets of strings, especially in the direction of connections to molecular computing, in particular membrane computing [104]. Another challenging field for research is to find connections between game theory and grammar systems theory, especially in the case of CD grammar systems. Obviously, classical language theoretic problems, that is, decidability questions, communication complexity and computational complexity issues are of interest. Grammar Systems theory opens a new research area in descriptional complexity as well, namely it gives the possibility to develop the concept of “collective complexity” and to compare the behavior of the corresponding measures to the behavior of the well-known complexity measures based on single grammars in the case of different language classes.

References


149


Grammar systems: A short survey


Grammar systems: A short survey


