Analyzing P Systems Structure: Working, Predictions, and Some Linguistic Suggestions*

Gemma BEL ENGUIX

Research Group on Mathematical Linguistics
Rovira i Virgili University
Pl. Imperial Tàrraco, 1, 43005 Tarragona, Spain
and
Department of Computer Science, University of Milan-Bicocca
Via Bicocca degli Arcimboldi, 8, 20126 Milan, Italy
E–mail:gbe@astor.urv.es

Abstract

The aim of this paper is to go more deeply into some mathematical properties of P systems and suggest some possible applications in the field of sociolinguistics. Membrane systems are, in this paper, the theoretical framework, mathematics is a tool for analysis and sociolinguistics is the final goal. By means of this interdisciplinary approach, that takes elements from Computer Science, Biology and Linguistics, we try to start a new and very simple line of research in Linguistics.

1 Introduction

Membrane systems, introduced by [6] are a powerful and increasingly spread model of computation. Their flexibility and intuitive functioning makes them very suitable for applications, not only to Computer Science, but also for computing *real life events*, like interaction between societies, or language evolution. The present paper introduces the study of a membrane system in a given state, trying to establish if it is possible to predict the development of the computation, and applying the results to the study of interaction between languages in societies. Therefore, the chief goals of the paper can be formulated as follows:

- To describe several types of structural relations between membranes (section 2).
- To define the way the system works and the symbol-objects spread (section 3).

*This research has been supported by Marie Curie Fellowship of the European Community programme *Human Potential (IHP)* under contract number HPMF-CT-2002-01582.

- To give a formula for predicting, looking at one state of the computation, what of the elements will spread to the whole system (section 3.1).
- To make some suggestions on how to apply these concepts to interaction between languages, with special attention to sociolinguistics (section 4).

For adjusting working of P systems to our approach, some special features have been introduced, in a way that the main traits of membrane systems described in the paper are:

- They do not generate new symbol-objects. We deal just with the spread capacity of the existent ones.
- They work with no specific rules for each membrane, using just general principles, in the same way societies evolve following general "darwinian" tendencies rather than specific rules.
- The goal of every symbol is to spread to every membrane in the system, achieving what we call over-saturation.

2 Relations between membranes

In this section we examine three possible types of structural relations between membranes in a system. The way the membranes are related to the others is important in the moment they have to interact, and also in the configuration of the communication we are going to deal with later. There are mainly three types of relations: *nesting*, *sibling* and *command*, that we explain in the sequel.

1. Nesting. Given two membranes M_1, M_2, it is said M_2 to be nested in M_1 when it is inside M_1. The outer membrane M_1 is called *parent membrane* and the inner membrane M_2 is called *nested membrane*. It is denoted $M_2 \subset M_1$: $[_1\ [_2\ ]_2\ ]_1$. The notation $\subset M_1$ refers to every membrane nested in M_1.

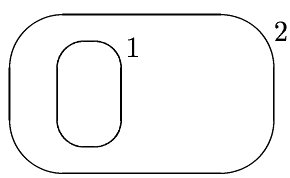

Figure 1: Nesting

Degree of nesting refers to the number of membranes between the nested one and the parent one. The degree of nesting is obtained by subtracting the depth of the parent membrane M_p to the depth of the nested membrane M_n. This is: $deg(M_n \subset M_p) = depth(M_n) - depth(M_p)$.

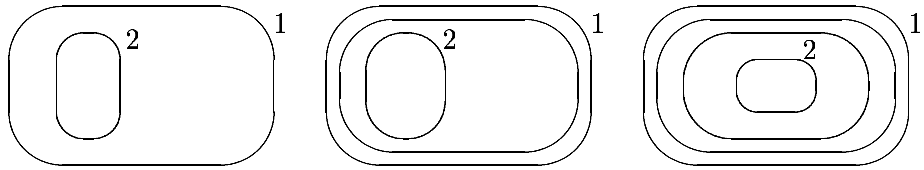

Figure 2: $M_2 \subset M_1$ with degree 1, 2 and 3

2. Sibling. Two membranes M_n, M_m are related by sibling, if they satisfy:

i. they are adjacent or nested in adjacent membranes, and

ii. they have the same depth.

Sibling is denoted $M_n \approx M_m$. In a membrane system drawn as $[_0\ [_1\ [_2\ ]_2\ ]_1\ [_3\ [_4\ ]_4\ ]_3\ ]_0$, $M_1 \approx M_3$ and $M_2 \approx M_4$. The notation $\approx M_n$ refers to every sibling membrane for M_n.

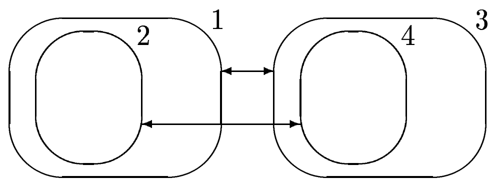

Figure 3: Sibling

Degree of sibling refers to the proximity of two membranes related by sibling. For obtaining the degree of sibling, we establish the following:

- Two sibling membranes, $M_n \approx M_m$, are called of degree 0 when they have the same strict parent membrane.
- For two sibling membranes $M_n \approx M_m$, which are not of degree 0, we obtain the degree of sibling, by subtracting the depth of M_n, M_n (they have the same by definition) minus the depth of M_i, M_j, being M_i, M_j two membranes which satisfy: a) $M_i \approx M_j$ with degree 0, b) $M_n \subset M_i$, $M_m \subset M_j$.

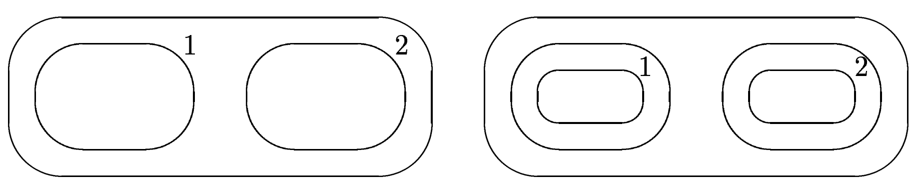

Figure 4: $M_1 \approx M_2$ with degree 0 and 1

3. Command. Given two membranes M_n, M_m, M_n commands M_m iff:

i. they are not nested,

ii. both are nested in a membrane M_j,

iii. $deg(M_n \subset M_j) = 1$, $deg(M_m \subset M_j) > 1$

Command is denoted $M_n \triangleleft M_m$. In the system $[_0\ [_1\ [_2\ ]_2\ ]_1\ [_3\ [_4\ ]_4\ ]_3\ ]_0$, $M_1 \triangleleft M_4$ and $M_3 \triangleleft M_2$. The notation $\triangleright M_n$ refers to every membrane commanded by M_n.

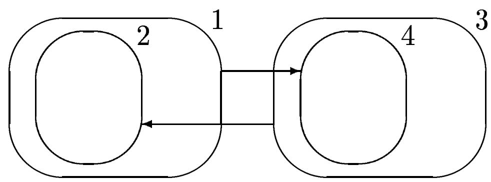

Figure 5: Command

Degree of command is the depth of the commanded membrane respect to the commander one. For obtaining the degree of command between two commanded membranes, we have to subtract the depth of the commanded one minus the depth of the commander. This is: $deg(M_n \triangleleft M_p) = depth(M_p) - depth(M_n)$.

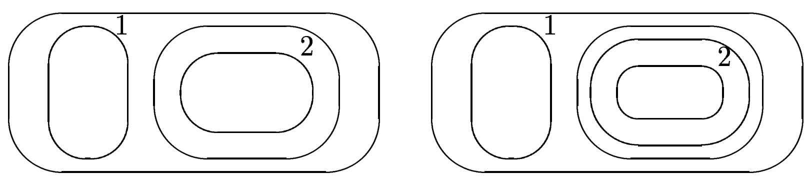

Figure 6: $M_1 \triangleleft M_2$ with degree 1 and 2

3 Definitions and basic concepts

The basic concepts introduced in the present section describe the structure of P system and explain how the properties described above generate activity in replication of symbols in the system. These concepts are gathered in the following topics:

- Structural features of a membrane system.
- Spreading depth and wideness of a symbol in a system.
- Edges and paths.
- Generation and replication of symbols.

- Densities of symbols.
- Working of replication in the system.

Structural features

1 The *degree of a system* is the number of membranes it has (cf. [6]). It is denoted by $degree(\mu)$.

2 The *depth* of a membrane M_n, denoted by $depth(M_n)$, is the number of parental membranes of M_n, + 1.

3 The set of all membranes in the system with the same depth constitutes a *level*. The level of a membrane, denoted by $lv(M)$, is equal to $depth(M) - 1$; this is the number of parental membranes of it.

4 No symbol exists in level 0.

5 A *terminal nesting membrane* is a membrane that does not contain any nested membrane.

6 A *terminal commanded membrane* is a membrane that cannot command any other membrane. The skin membrane is not terminal in command even if it cannot command.

7 A *terminal sibling membrane* is a membrane that does not have any sibling membrane, except the skin membrane.

8 The *system depth*, denoted by $depth(\mu)$, is the maximal depth of membranes belonging to μ (cf. [6]).

9 The *system wideness*, denoted $w(\mu)$, is the number of membranes M in the system so that $depth(M) = 2$.

10 The deepest level in the system is denoted by $dl(\mu)$, and it is also the number of levels the system has.

11 $degree\ (depth(n))$ is the number of membranes in the system whose depth is n.

12 $degree\ (level(n))$ refers to the number of membranes in the system which are in level n.

Spreading depth and wideness

13 The *spreading wideness* of an element x in a system is the number of membranes in the system where a symbol is present. We denote it by $\mu sw(x)$. When $\mu sw(x) = degree(\mu)$, it is said that the spreading is *complete* and the symbol is present in every membrane of the system.

14 The *spreading wideness in level n of a symbol x* is the number of membranes in level n where the symbol is located. Here, i is the total number of membranes in level n in the system, and it is represented by *sw(x) in level n over i*. If $sw = i$ in a given level, then the spreading wideness is said to be *full* in that level.

15 The *spreading depth* of a symbol x (denoted $\mu sd(x)$) is the level of the deepest membrane where the symbol is placed.

16 The *continued spreading depth* of a symbol x is the deepest level n in the system so that from level 1 to n we have $sw(x) \neq 0$. It is denoted by $csd(x)$.

17 The *full spreading depth* is the deepest level n of the system where a symbol x has the property that from level 1 to n, $sw(x)$ is *full*. It is denoted by $fsd(x)$.

18 When the *full spreading depth* for a symbol x is $depth(\mu) - 1$, then the spreading is called *complete* (see definition 13). If the spreading is complete, then a copy of x is present in every membrane of the system.

Edges and paths

19 *Edges* are ways to connect two membranes in the system, considering the relations of *nesting*, *sibling*, and *command*. There are three types of edges:

- *Nesting edge.* It is the one linking two membranes related by nesting of degree 1.
- *Sibling edge.* It is the one linking two membranes related by sibling of degree 0.
- *Command edge.* It is the one linking two membranes related by command of degree 1.

A connection between two membranes which are not related by nesting, sibling or command is not an edge.

20 A *path* is a set of connected edges. It is a graph where membranes are vertices with or without symbol-objects.

21 Every vertex in a path has degree 2, except for two of them, called initial and terminal, which have degree 1.

22 The most external membrane in a path which is a vertex of degree 1 is called *initial vertex.* The deepest membrane in a path which is a vertex of degree 1 is called *terminal vertex.*

23 If we label the vertices in a path with $v_1, v_2, ..., v_n$, where v_1 is the initial vertex and v_n the terminal one, then they must observe the following condition $lv(v_1) \geq lv(v_2) \geq ... \geq lv(v_n)$.

24 The *degree of a path* is the number of edges it has.

25 It may exist a path with just one edge. It is called a *minimal path.*

26 A path with more than one edge is called a *multiple path.*

27 If every edge of a multiple path belongs to the same type, then the path is called *monotonic*.

28 If not every edge of a multiple path belongs to the same type, then the path is called *complex*.

29 There are three types of monotonic paths:

- A *nesting path* is a path where every edge is a nesting one.
- A *sibling path* is a path where every edge is a sibling one.
- A *command path* is a path where every edge is a command one.

30 When the initial vertex of a monotonic path is in level 1 and the terminal one is in a terminal membrane with level > 1, then it is called a *complete path*.

31 There are two types of complete paths:

- A *nesting complete path* is the one going from a level 1 membrane to a nested terminal membrane.
- A *command complete path* is the one going from a level 1 membrane to a command terminal membrane.

32 A path connecting every sibling membrane of degree 0 is called a *ring*.

Replication and generation

33 In linguistics, as well as in genetics, two steps are necessary for a symbol to spread: 1) generation, 2) replication. Generation happens only once, whereas replication can be applied an arbitrary number of times.

34 An element can be generated anywhere in the system. In this paper we do not deal with mechanisms of generation.

35 The *primary occurrence* of a symbol in a system is the one which does not exist by replication, but by generation.

36 If a system has just one copy of an element, then such a symbol is called *unitary*. A unitary symbol is considered to be *primary*.

37 A primary occurrence of a symbol can be calculated in a system. It is the initial node of a spreading route.

38 When in a ring the primary occurrence cannot be calculated, this is, by convention, the vertex (membrane) labelled with the lowest number.

39 A *spreading route* is a path where every membrane (vertex) has a copy of the same symbol.

40 A *spreading route* must be maximal, that is, must connect as many membranes as possible.

41 A replicated symbol of the system may configure several spreading routes.

42 If two spreading routes of the same symbol share at least one node, then they are *connected.*

43 If a route for a replicated symbol in a system does not share any node with the other spreading routes of the same symbol, then this is a *disconnected* spreading route.

44 If no path representing a spreading route can be drawn from an element, then it is an *isolated symbol.*

45 If a spreading route is a minimal path, then it is a *minimal spreading route.* This is called *simple replication.*

46 If a spreading route is a multiple path, then it is a *multiple spreading route* and the process is called *multiple replication.*

47 If a multiple spreading route is a monotonic path then it is a *monotonic route* and the process is called *monotonic replication.*

48 If a multiple route is a complex path, then it is a *complex route* and the process is called *complex replication.*

49 When a monotonic route is a nesting path, the route is called *nesting spreading route.* When a nested spreading is a complete path, it is called *complete nesting spreading route.*

50 When a monotonic route is a command path, the route is called *command spreading route.* When a nested spreading is a complete path, it is called *complete command spreading route.*

51 When a monotonic route is a sibling path, the route is called *sibling spreading route.* When a nested spreading is a complete ring, it is called *ring spreading route.*

52 The union of all spreading routes of a symbol, forms a *spreading tree.* The parental node is the primary element.

Density

53 The *density* of a symbol in a *membrane* M_n is the number of copies this symbol has in the membrane. It is denoted by $dens(x_n)$.

54 The *density* of a symbol *in a level* n of the system is the average of the density of the symbol in every membrane of the level. It is denoted by $dens(x)$ *in level* n.

55 The *deepest density* of a symbol x is the density of this symbol in the deepest level of the system. It is denoted by $ddens(x)$.

56 The *maximal density* of a symbol x is the highest density of the symbol in the system. It is denoted by $max\ dens(x)$.

57 The *minimal density* of a symbol x is the lowest density of the symbol in the system. It is denoted by $min\ dens(x)$.

58 The *maximal level density* of a symbol x, denoted by $max\ ldens(x)$, is the maximal density of the element in the levels of the system.

59 The *minimal level density* of a symbol x, denoted by $min\ ldens(x)$, is the minimal density of the element in the levels of the system.

60 *Over-representation*

- A symbol x is over-represented in a membrane M_n if $dens(x_n) > 1$.
- A symbol x is over-represented in a level n for a number i of membranes if: i) $sw(x)\ in\ level\ n\ over\ i = i$, and ii) $dens(x)\ in\ level\ n > 1$.
- A symbol x is over-represented in a complete nesting path if: i) $dens(x) \geq 1$ in every membrane of the path, and ii) at least in one membrane $dens(x) > 1$.
- A symbol x is over-represented in a complete command path if: i) $dens(x) \geq 1$ in every membrane of the path, and ii) at least in one membrane $dens(x) > 1$.
- A symbol x is over-represented in a ring, if: i) $dens(x) \geq 1$ in every membrane, and ii) at least in one membrane $dens(x) > 1$.
- A symbol x is over-represented in a spreading route if at least in one membrane we have $dens(x) > 1$.

Working of the system

61 In every path of the computation, only one element of each membrane can be replicated. The symbol replicated is the one with the highest density in that membrane. Just one copy of the maximal density symbol is created by a membrane in one step.

62 If the highest density is shared by two or more elements, then the membrane does not replicate any symbol.

63 The copy of a replicated symbol spreads to a membrane linked to the one it is by an edge.

64 When a terminal nesting membrane is also a terminal sibling membrane, then copies of symbols remain in the same membrane.

65 Computation proceeds in parallel.

66 Spreading goes from higher to lower levels, except for the case of rule 67.

67 *Rule of saturation.* When an element is present in every membrane nested with degree 1 in a membrane M_m, then it is expanded to M_m, provided that there is no copy of this symbol in M_m. The application of the expansion rule is immediate and it is applied just once. It does not take a step in the computation.

68 The goal of every symbol is to reach a *complete* μsw, that is, to be present in every membrane. When this happens, the system is said to be *oversaturated* by this element.

69 When an element reaches the oversaturation, the system stops evolving.

From the statements above, we think it is possible to predict which element will be the most spread after n steps, and which of the elements will oversaturate that system.

3.1 Predicting oversaturation

There are situations which allow predictions about which elements will reach a complete spreading before other elements. To this aim, some observations are in order::

1. Spreading goes to the more shallow levels to the deepest ones, except for the case of saturation.
2. In general, a symbol located in the shallowest levels spreads in a easier way than a symbol located in the deeper levels.
3. For a symbol, to be alone in a membrane assures that it can be replicated.
4. To get high densities is even more important than to be very spread. A high density has two advantages:
   - it makes easier the replication,
   - it blocks the spreading of the other elements.
5. Rings are paths which help to make a double movement of expansion:
   - an external expansion by saturation,
   - the common inner expansion.
6. To get rings in deep nesting terminal membranes allows just one movement by saturation, but the symbols get blocked and they cannot spread.

Keeping those considerations in mind, we establish some advantageous situations for a symbol to get the oversaturation.

1. If a symbol is located, in a given state, in every nesting terminal string, it will get a complete spread in one step. Recall that the application of rule of saturation does not depend of the density and that it is immediate. By replication, the probabilities this configuration to take place are very few, but when it happens it is direct.

2. A ring or fsw in level 1 with over-representation gives advantage to a symbol.

From here, we will try to give a formula for calculating the *spreading capacity*, which will be denoted by σ. For a symbol x, we establish:

$$\sigma = \sum_{i=1}^{dl(\mu)-1} \frac{dens(x)\ in\ lev\ i}{i} + \frac{sw(x)\ in\ dl(\mu)}{degree\ (dl(\mu))}$$

Usually, the symbol with the highest σ in the state we look at the system, will be the one to get the oversaturation.

4 Relevance of the approach for sociolinguistics

In several papers (see [1] and the references therein), we have pointed out that P systems are suitable to deal with different parts of linguistics where the notions of *domains*, *alphabets*, and *membranes* are relevant/natural.

A key idea for dealing with social sciences, sociolinguistics, and language interaction, is the consideration of membranes as social groups. In our approach, such groups can be either users of different languages or different variants of the same language. In the first case, the model is valid for dealing with the contact between languages, in the latter, it leads to linguistic shifts by means of the spread of several particular structures.

The linguistic motivation of the paper is the following: sometimes, languages interact, and words, structures and even sounds travel from one membrane to another one. In the beginning, they are rejected, they are not recognized as a part of that language. However, after some time the domain uses to be modified to accept the novelties, in a way that some words or structures of a powerful (from the social or political point of view) language may be very well represented in other membranes.

In the present approach we have not used the concept of domain. We have considered that any word of any language can be spread to other languages if the situation is advantageous. This is the process we are experimenting today with the globalization, where words travel from languages to languages and from societies to societies surprisingly easily. Domains can be taken into account in the present formalization, with the increasing of complexity but perhaps with a better adequacy to the real word. As a consequence of the lack of domains, only one alphabet has been used. However, its symbols (words, or structures) have been considered to belong to different linguistic systems.

The final result of interaction between languages is usually one the following:

a) there is a simple exchange of elements, and some linguistic units of a system are integrated in another one,

b) bilingualism: connivance between two languages in the same group or territory (membrane),

c) in the most critical cases, some of the languages die when they are completely substituted by others.

The features we have described and analyzed in the systems introduced here are not suitable to study the interaction and the exchange of elements between languages, because domains and several alphabets are necessary in order to do that. But the other situations can be described easily, in the following way:

- The linguistic substitution corresponds to the situation we have described in this paper. When one of the symbols reaches oversaturation, then it substitutes the other symbols in the system. The process is not automatic, but when an element oversatures, the others get blocked, and it is easy to infer that an element which cannot replicate, disappears.

- The bilingualism refers to the situation, not usual but possible, in which no element is able to get the oversaturation. In this sense, we see that every step which is not a final step can be a bilingual state in a membrane and, some times, even n-lingual. However, in most of cases bilingualism is not a final state, but an intermediate state in the process of substitution.

The goal for us is to predict the final state of every symbol of the system by analyzing a given configuration of the computation. To know if the future for a word, structure, slang, or language is the spread or the death is a key point for sociolinguistics, and in the framework of membrane systems, this can be calculated by formal and simple methods.

5 Conclusions

The present paper is an initial approach to describing the working of P systems with no rules, and to the structural analysis of membrane systems. Both topics have been just introduced, and a further treatment should be done to improve the method and formalization. The development of the theory has as a final goal the application to sociolinguistics, and this goal has been only pointed out in this paper. In the near future, the tools proposed here must be improved and applied to more realistic situations – and we plan to return to this topic in a further work.

References

[1] Bel Enguix, G. (2003), Preliminaries about Some Possible Applications of P Systems in Linguistics, in Păun, Gh., Rozenberg, G., Salomaa, A. & Zandron, C., *Membrane Computing*, LNCS 2597, Springer, Berlin.

[2] Calude, C. & Păun, Gh. (2001), *Computing with Cells and Atoms*, Taylor and Francis, London.

[3] Croft, W. (2000), *Explaining Language Change. An Evolutionary Approach*, Longman, Singapore.

[4] Jiménez-López, M.D. (2002), *Using Grammar Systems*, GRLMC Report, Rovira i Virgili University, Tarragona.

[5] McMahon, A.M.S. (1996), *Understanding Language Change*, Cambridge University Press.

[6] Păun, Gh. (2000), Computing with Membranes, *Journal of Computer and System Sciences*, 61: 108-143.

[7] Păun, Gh. (2002), *Membrane Computing. An Introduction*, Springer, Berlin.

[8] The P Systems Web Page: http://psystems.disco.unimib.it