ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA INFORMÁTICA

Grado en ingeniería informática

Design of an artificial language for Human-Computer interaction.

Diseño de un lenguaje artificial para la interacción Hombre-Máquina.

Realizador por

Jorge García González

Tutorizado por

Francisco J. Vico Vela

Departamento

Lenguajes y Ciencias de la Computación

UNIVERSIDAD DE MÁLAGA

MÁLAGA, Noviembre 2016

Fecha defensa:

El Secretario del Tribunal

Resumen

Este trabajo tiene como objetivo señalar las motivaciones para la búsqueda y creación de un lenguaje artificial de propósito general óptimo con el objetivo práctico a largo plazo de poder realizar una comunicación entre hombre y máquina eficaz y en ausencia de ambigüedades. Siguiendo esa idea, nos planteamos cómo podría estudiarse qué elementos básicos presentes en un lenguaje son realmente necesarios y cuales son simplemente innecesarios o incluso contraproducentes. Dado el objetivo de optimalidad que motiva el trabajo, será necesario intentar plantearse el estudio de los lenguajes de propósito general de manera constructiva (desde la ausencia de los mismos, componiendolos progresivamente).

Este trabajo no tiene como objetivo la comprensión del funcionamiento de un idioma (lenguaje natural de propósito general) concreto, por tanto no intentaremos profundizar sobre las complejidades de las capacidades humanas para utilizar lenguajes así como no someteremos los lenguajes resultantes a coincidir con las características de ningún lenguaje natural concreto. Pese a ello, sí que partiremos como objeto de estudio de los tres elementos básicos en la sintaxis de los lenguajes naturales (sujeto, verbo y objeto).

Propondremos un problema simple de agrupamiento en el que intervengan varios agentes que dispongan de información parcial que podrá ser compartida entre los susodichos mediante lenguajes no ambiguos de distinta complejidad. Definiremos varios lenguajes con sintaxis, semántica y uso de complejidades crecientes, describiremos tanto el problema como las herramientas para evaluar los resultados contrastando la efectividad de cada lenguaje con el objetivo de tener unos resultados que puedan respaldar empíricamente qué lenguaje resulta mejor para resolver el problema.

Palabras claves

Lenguajes, lenguajes artificiales, agrupamiento, sistema multiagente.

Abstract

This project has as objective to note motivations to search and develop an optimal general purpose artificial language with effective and not ambiguous human-machine communication objective in long term. Following this idea, we will ask ourselves how could be studied which languages basic elements are actually necessaries and which ones are unnecessary or even counter-productive. Given the optimality objective that motivates this works, it will be necessary to face general purpose language from a constructive way (from the absence of language, creating them progressively).

This work does not aim to face an specific natural language understanding so we will not study human language capacity complexities and we will not subdue our resulting languages to match with any natural language. With that, we will use as study subject three basics elements in natural languages (subject, verb and object).

We will propose an enough simple clustering problem performed by agents with partial information that will be shared between them by the use of non ambiguous languages with increasing complex syntax, semantic and use. We will also define some mathematical tools to evaluate results to contrast each language effectiveness to obtain results to empirically support what language is better to face proposed problem.

Keywords

Languages, artificial languages, clustering, multiagent system.

Contents

1 Introduction

This work is divided in three main sections. This, first section, serves as introduction to provide context, motivation and approach to the whole work. Second section is dedicated to expose our proposed problem, models to resolve it, mathematical formalizations and evaluation tools. Third section is dedicated to experiments description and results. To end, there are conclusions about the project, references and an appendix with code details.

1.1 Language problem from computer science point of view

Natural Language Processing (NLP from now on) is one of the main goals in Artificial Intelligence (AI from now on) because of two main reasons: the first is mainly practical, the communication between humans and machines can change the way we interact with computers and allow a channel of communication much more natural and intuitive for us; the second main reason is a matter of cognitive knowledge, through the study of this problem we can try to understand an issue so complex as the human language mental function.

Regarding to the first goal, in the last years we have seen impressive advances. In this direction, statistical NLP is giving goods results, this reformulation from the traditional non-statistical approach is characterized by simpler approximations in front of deep analysis of languages and machine-learning methods based on large and annotated bodies of text (Nadkarni et al., 2011).The best example should be the famous IBM's Watson (Gliozzo et al., 2013) and its participation in 'Jeopardy!' show. We also can see advances in symbolic AI applied to NLP such as the use of George Lakoff's natural logic to infer from a text in natural language (MacCartney & Manning, 2007), for example.

The relation between our usual notion of intelligence, AI and NLP is so old that language is a fundamental part of Turing test. There is no exact definition of intelligence but intuitively is normal to guess that "intelligence", whatever it may be, must imply the ability of use a language as we, humans, do, because we are "intelligent" and we have that ability. Of course, this is an inference that, in absence of a formal definition of intelligence, has little basis but that absence makes so important NLP as reference point like Chess or Go until they were overtaken.

If we observe the usual approach to language analysis and processing, it is often a Top-Down one: it is an attempt that tries to deal with a specific language from its actual complex state or with the language problem in abstract trying to explain all its properties. An important example in this case could be Hans Kamp's Discourse Representation Theory (Geurts et al., 2007) or Montague Grammar (Janssen, 2011).

The problem with this approach is that languages are complex compositions resulted from thousands of years of evolution submitted to culturals, biologicals, historicals and socials factors. For this reason, each natural language could have “burden” in his formation: unnecessary complexity that complicate language analysis and processing.

This is a lack to those two main objectives we exposed previously (NLP and understanding language). NLP is a matter that requires a lot of information much of which is not in the language itself. For example, the matter of give meaning to a word (define a concept in the end). How could a machine understand a word like pain if it doesn’t feel it? Yes. We have definitions like “Highly unpleasant physical sensation caused by illness or injury” from Oxford Dictionary but what allow us to understand the word pain is we can feel it and we guess the others beings feel it like we do.

It would be comparable with the “Flatland, romance of many dimensions” metaphor. In that book, it is proposed a two dimensional world inhabited by two dimensional entities. If a sphere cross this word, entities could not actually understand the sphere, only a succession of circles, because there is only two spatial dimensions for them. Another comparison would be our understanding of quantum physics. Yes, we have mathematical equations to define it but we have little capacity to imagine it. It is not the world we sense, so it is at least unintuitive and maybe we cannot in fact totally understand it. The last comparison and maybe the best would be to question ourselves about the color of an infrared light band. We simply can not answer. We know infrared frequencies exist but we can not link a colour to them since our eyes cone cells lead only with some light frequencies, we only know colours related to a little spectrum of frequencies and others colours seems simply unimaginables.

In this line, can we actually expect a machine to understand all information we associate to words like chair, dog, earth, or liquid? We can make statistical analysis and obtain responses to phrases with that words or even logical definitions and inferences, but maybe we cannot give a machine without other human-like capabilities all information about a word and, without that information, it would be impossible to make a like-human NLP.

In the line of language analysis: can we actually expect to understand language ability without a formalized progression in that understanding? As we have complexities hierarchies to analyse problems and algorithms, it would be useful a reference formalized basis to compare natural languages.

1.2 What are languages for?

This should be the first question to approach language problem. What is the utility of languages? There is no agreement among expert about how the language ability is developed in humans (Pinker, 2003) but, if evolution has kept and reinforced it, according to darwinist evolution, it should be useful. In humans, each natural language is a cultural product based on brain ability, in other species there are also communication systems codified in his behaviour but in all cases the objective is to share information.

Information is a necessary resource to all cognitive beings, due to, without it, there is no sense in cognition itself. Information is our main resource to know how to interact with our world in order to survive. All cognitive beings have at least a mean to gather information directly, what usually is called perception. Our eyes gather light information with the purpose of image composition by our brain, our ears gather sound, our nose smells, our tongue tastes, all our body is full of nerves to gather all kind of information about ourselves and what surrounds us but it seems not enough. To optimize theirs chances, a lot of species (humans included) exchange information between individuals so each of them can obtain that information with no direct gathering or processing. We can save resources (time and energy) by learn information from others and, in human case (as in other simians), it allow the birth of culture (extremely rudimentary in case of simians, like the utilization of some tools shared between parents and children). Information exchange is the basis of civilization.

If we all had to rely only on our senses to gather information, it would be impossible to achieve a lot of advanced conclusions (derived information) as we do or to coordinate ourselves in order to reach many of our goals. If ants or bees were not able to share where food is, it would be impossible for them to gather enough to survive as they do. More beings implies a greater information exchange necessity. Information sharing allows societies to exists and survive.

Of course, there are a lot of means to share differents kinds of information. Simple sounds (like a simple shout with no words) can transmit information like danger, alert or joy between human as well as between other kinds of animals (canines, felines, birds, all them have sounds with meaning), a human baby can not speak yet but his crying also has information (as simple and ambiguous as "maybe i'm hungry, maybe I'm thirsty, maybe I'm painful or maybe I'm just bored" but information). We can barely call it language due to its lack of structure and even with structure, some disciplines (psicolingüistics for example) could limit the word language to human natural languages, referring to others as communication systems. We will see some examples of these alternative communication systems later. In any case, we will refer as language to anyone with syntax structure.

We have argued that languages serve as means to exchange information but they are not the only ones. As we have seen, there are examples of simple sounds with no syntax elements to accomplish that task. How much complexity is necessary in a language? Instead of answer this question, another one could be answered before: are languages actually necessary? As we have said previously, we need our natural languages to build our civilizations but this lead to question if we truly need civilizations and go nowhere.

Languages are a matter of information exchange and yes, it seems we (humans) have a few instinctive reaction to associate some information to certain sounds(cry, laugh,...), faces (smile, frown, ...), even postures (straight, huddled, ...). The problem is that simple information exchange is not enough to face our tasks. There is no necessity of language or some language complexity per se, there is necessary language to allow a group of agents to perform a task. Our human tasks (problems to resolve) require much more information than we can exchange with those simple actions and there lies the necessity of complex languages.

Language and task is a not an extrange relation. From formal language point of view is well known the relation between Chomsky's syntax hierarchy and Automata Theory but we will carry out a different approach. We are interested in study the minimal language required to perform a certain task by a group of entities with partial information so we can obtain knowledge about language elements over a basis and use this knowledge to study the simplest general purpose language creation.

This artificial general purpose language should be complex enough to serve the objective of sharing information about our world but it could lack of defects and problems natural languages carry due to their develop throughout history.

1.3 Natural human language basic structure and animal examples

Words

The first component of language is mental lexicon or dictionary, a finite set of memorized words. Each of which is an arbitrary sign or string of signs that serves as connection, shared by the language users, between an arbitrary sound signal and a concept. In words of Steven Pinker: "The word duck does not look like a duck, walk like a duck, or quack like a duck, but I can use it to convey the idea of a duck because we all have learned the same connection between the sound and the meaning" (Pinker, 2003). Due to this arbitrary relation between words and meanings, we will not pay much attention to lexicon.

Grammar

The second main component of the language is the combination of words into greater structures. The rules for this combination are what we usually call grammar. It's important to note that, from linguistic point of view, grammar includes syntax as much as morphology (or even phonology according to some schools), while formal grammar used in computer science or mathematics is only a set of syntax rules that represents a set of strings. The reason why formal sciences ignore morphology is because it is explained by syntax. Morphology is syntax applied to the formation of words so we will ignore morphology. We will focus on written language so we will also ignore phonology.

For these reasons, our main objective when we speak about language will be mainly the syntax of that language.

It is also important to point out in linguistic and computer science the existence of recursive rules in grammar allows systems to generate an infinite set of sentences (strings from computer science point of view) based on a finite set of words (basic symbols). As well as syntax is semantically compositional: we can obtain the meaning of the whole from knowing the meaning of each element involved and following the rules that combine them. Since we can obtain an infinite set of different sentences and we can obtain the meanings of them, we can express an infinite set of meanings through language. This, and the ability to create new words, allow us to have a general purpose language.

1.3.1 Some alternative languages examples

It is interesting to observe other kinds of communication systems in animal kingdom to compare how sharing information necessity has been covered. We need to note again that, from a linguistic point of view, these systems are not languages but from computer science point of view they are.

The first example is the well known waggle dance. Waggle dance is a term used to refer a particular succession of movements made by honey bees to mark the direction and distance of a food source. The communication consists in a characteristic dance performed by a bee. The dance angle in relation to the hive and the sun is translated by other bees to a direction and a distance where food is (Grüter & Balbuena, 2008).

This is one of the more complex non-primate communication system known and is an example of how nature has created an strategy to transfer information between individuals to perform a task with better chances, in this case, get food.

A second example of communication system is stigmergy (Marsh & Olof, 2007). This system is a mechanism of indirect coordination performed by some social insects (ants, for example) using the environment to leave a trace that will affect a future action of the same individual or another one. In the case of ants, they leave a trace of pheromone on their way back to the nest when they have found some food. By the repetition of trace left, the probability of following that trace to the food increases. More ants find food at some place (what could mean a greater amount of food), more probability to go to that point. Ant stigmergy advantages over pathfinding are well

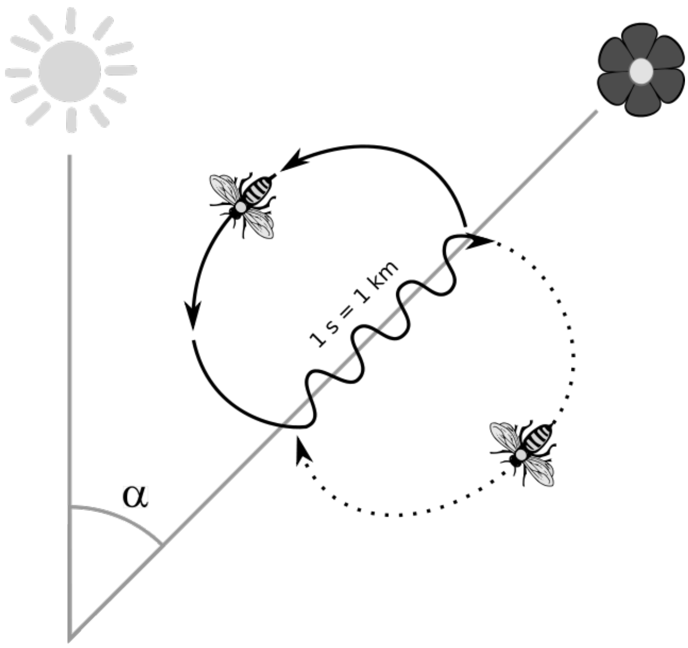

Figure 1: Waggle dance example. The direction to the source is indicated by the direction in which the bee moves in relation to the hive and the Sun. Image obtained from Wikipedia Commons under Creative Commons License.

studied. In this language, the information about the route is implicitly shared between individuals. In this case, is notorious the error known as Ant Mill: a group of ants lose the pheromonal track and start following among them, forming a circle until they die exhausted (Delsuc, 2003). A consequence of information lacking causes leads to what we could call a misunderstanding.

1.4 Language emergency in a machine approach

Previously, we have established a relation between languages and problems and we will define one to contrast language features utilities. Nevertheless, to resolve a problem is not only important the language. It is also important the use of that language, so problem and language must be related between them by an algorithm.

The same language used to resolve a problem, with different use (algorithm) can be useless to resolve the same problem. So, when we speak of a language complex enough to resolve a problem we always speak of that language used by an appropriate algorithm.

For that reason, linked to a problem and a related minimal language would be an algorithm. If we observe the natural approach, we can see that language and algorithm are developed simultaneously. Changes in the use imply changes in the language and vice versa. This would be the evolutionary approach that would lead to language emergency in a machine to resolve a problem.

But this is not enough. This work is born under the perspective of multiple problems that evolutionary approach has lead. This try-error development is extremely complex does not ensure a minimal language. To get one or various minimal languages (and theirs associated uses) would be needed a mathematically theory about languages strong enough to sustain the demonstration of a minimal result.

Unfortunately, that theory does not exist. Any computer scientist could be tempted to think about Formal Language theory but that theory includes only syntax and no semantic to relate that syntax with the world of use of the language. What we will try here will be to shed light on common language elements uses and implications to set the basis to the possible development of that theory.

As we have discussed previously, we have related problems and languages by mean of algorithms. So, the first to be able to shed light on languages is to look for a problem where languages would be useful, complex enough to allow observation of language features utility but simple enough to have no elements disrupting that features and analysis. The aim now is a problem where language features could be appreciated with no disturb to evaluate if they are actually useful, what is their utility and if they can be discarded.

We will test elements in basic SVO (Subject Verb Object) syntax structure. SVO (or any other order between that three elements) is the common basic structure in many natural languages like Chinese, English, Estonian, French, Ganda, Italian, Japanese, Modern Hebrew, Polish, Portuguese, Quiche, Romanian, Serbo-Croatian, Slovene, Spanish, Swahili, Thai, Vietnamese, Zulu and others (Meyer, 2009).

2 Proposed clustering problem

2.1 Informal description

Suppose you are locked in a rail mounted cabin. You are not able to see what is outside the cabin but you know there are others cabins at the same rail. In the cabin, there is a monitor with the distance between you and previous cabin in rail, a button to order the cabin to advance an only step in rail, a keyboard to write a sentence to previous cabin or to next cabin and a monitor to read what cabins at your sides have sent to you. You and other cabins must perform a clustering task by placing adjacents cabins in the same cluster.

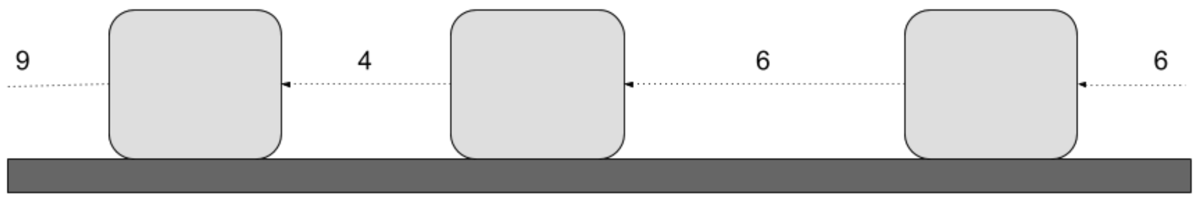

Figure 2: Cabin exterior distribution example

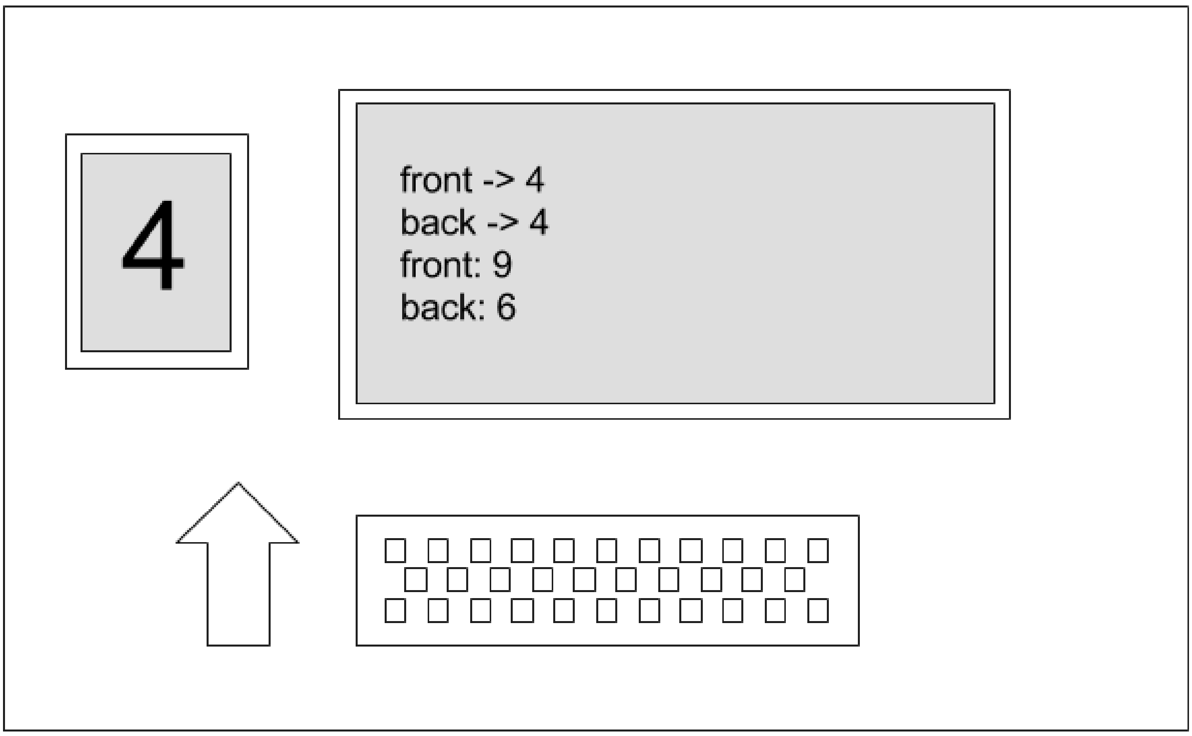

Figure 3: Cabin interior interface example

2.2 Formal description

- A finite one dimensional tape with size l divided into cells (possibles positions) identified by a natural number $(1, 2, 3, ..., l)$ so that the first cell is at the right of the last cell, forming a ring of cells.
- A set E of m entities (agents) $E = e_1, e_2, ..., e_m$ definedas follows
  - $e_i = (\rho_{i,t}, n_i, \Omega_{i,t}, \sigma_{i,t}, \delta_{i,t}, M_i)$ where
    - $\rho_{i,t}$ is e_i cell position (1 or 2,...or l) at iteration t. $\rho_{i,t} \neq \rho_{j,t}, i \neq j$, two entities cannot share cell.
    - n_i is an identification number. This number can or cannot be the same that entities index (e_1 could have $n_i = 2$) and are not assigned in order necessarily.
    - $\Omega_{i,t} = (\alpha_{i,t}, \beta_{i,t})$ is the representation of the world known by e_i. $\alpha_{i,t}$ and $\beta_{i,t}$ are vectors with the world information e_i has gathered till iteration t.
    - $\sigma_{i,t} = [(\phi_{i,t,k}, o_{i,t,k})]$ is a list of pairs of coming into sentences ($\phi_{i,t,k}$) and the side it comes from ($o_{i,t,k}$) at iteration t.
    - $\delta_{i,t}$ is the distance between e_i and the previous entity in the tape at iteration t.
    - M_i is entity's model of behavior. $M_i = M_j \ \forall i, j$ if all entities share the same model of behavior.
- A language (L) whose syntax, lexicon and interpretation is known by all entities.
- A cut distance (C) as dissimilarity reference.
- Each entity e_i at iteration t is able to:
  - "observe": To know the distance between it and entity at its left in ring $\delta_{i,t}$.
  - "move": To change its position for the position at its left if that cell is empty.
    $\begin{cases} \rho_{i,t+1} \leftarrow \rho_{i,t} - 1 \Leftrightarrow \rho_{i,t} \neq 1, \rho_{j,t} \neq \rho_{i,t} - 1 \ \forall j \ 1 \leq j \leq m \\ \rho_{i,t+1} \leftarrow l \Leftrightarrow \rho_{i,t} = 1, \rho_{j,t} \neq l \ \forall j \ 1 \leq j \leq m \end{cases}$
  - "say to previous or next entity": To assign previous or back entity input sentence list a sentence made with L by e_i.
  - "listen": To interpret and extract information from $\sigma_{i,t}$ to Ω_i.
- Each iteration of the problem implies an action turn for each entity.
- The final absolute positions of entities have no relevance, only final adjacency of entities matter.
- Index number is not known by entities (i.e. e_1 does not know its index is 1).
- C is known by all entities.

- The interpretation from each sentence $\phi_{i,t,k}$ must defined only by the sentence itself, speaker $o_{i,t,k}$ and receiver previous information Ω_i.

- Clustering result would be $A(E, [\rho_{1,t}, ..., \rho_{m,t}])$ with t as final iteration and A defined below.

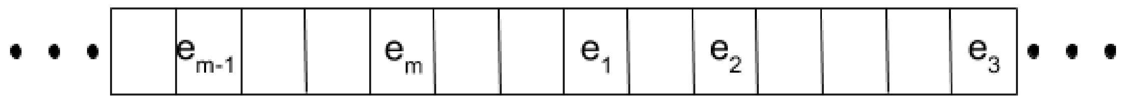

Figure 4: Ring tape example

We will define function Δ as distance between two cells b_1 and b_2 in ring:

$$\Delta(b_1, b_2) = \begin{cases} min(b_2 - b_1, l - b_2 + b_1) & b_1 \leq b_2 \\ min(b_1 - b_2, l - b_1 + b_2) & b_2 < b_1 \end{cases} \tag{1}$$

Given an entities set E, a position vector p and P as power set, adjacency clustering would be a function $A(E, p) = S$ where $\forall c_z \in S, c_z \subseteq E$ and defined as follows:

$$A(E, p) = \forall c_z \in P(E) - \{\varnothing\}, c_z \in S\ iff \tag{2}$$

- An entity with no entities in its sides cells. c_z is a cluster formed by an only entity. $|c_z| = 1, e_i \in c_z, \forall e_j \in E, i \neq j, \Delta(p(i), p(j)) > 1$

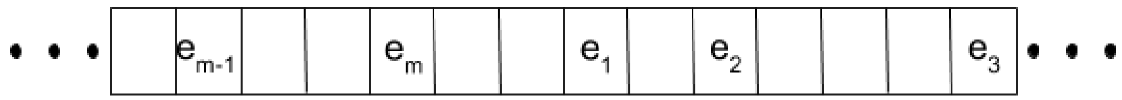

Figure 5: $\{e_{m-1}\}, \{e_m\}, \{e_1\}, \{e_2\}, \{e_3\} \in S$

- There are two entities in c_z and distance between them is 1. c_z is a cluster formed by two entities. c_z is not in a greater cluster c_y.
  $\forall e_i\ e_j \in c_z, i \neq j, \Delta(p(i), p(j)) = 1, \forall c_y \in S / c_z \neq c_y : c_z \nsubseteq c_y$

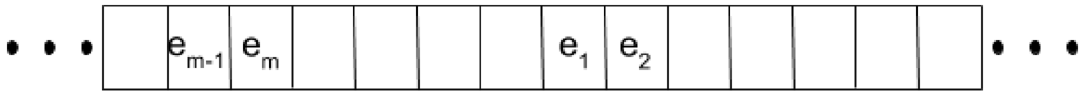

Figure 6: $\{e_{m-1}, e_m\}, \{e_1, e_2\} \in S$

- There are two entities in c_z, the distance between them is greater than 1 and cells between them are filled by other entities. c_z is a cluster with more than 2 consecutive entities.c_z is not in a greater cluster c_y.
  $\forall e_i\ e_j \in c_z, i < j, \Delta(p(i), p(j)) > 2, \forall k / i \leq k \leq j : e_k \in c_z$
  $\forall c_y \in S / c_z \neq c_y : c_z \nsubseteq c_y$

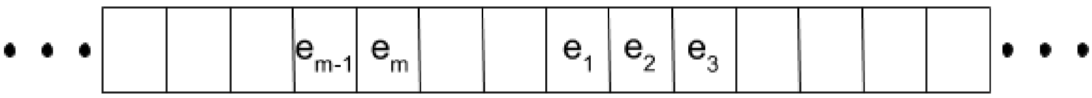

Figure 7: $\{e_{m-1}, e_m\}, \{e_1, e_2, e_3\} \in S$

- There are at least two entities in c_z, the distance between them is greater than 1 and cells between them are filled by other entities. c_z is a cluster with consecutive entities at ring end and ring beginning. c_z is not in a greater cluster c_y.
  $\forall e_i\ e_j \in c_z, i < j, \Delta(p(i), p(j)) > 2, \forall k / i \leq k \leq m \ or\ 1 \leq k \leq j, e_k \in c_z$
  $\forall c_y \in S / c_z \neq c_y : c_z \nsubseteq c_y$

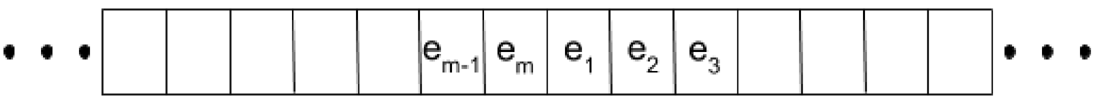

Figure 8: $\{e_m, e_1, e_2\} \notin S$ since $\{e_{m-1}, e_m, e_1, e_2, e_3\} \in S$

To simplify, we will establish the next equalities:
$p_i = \rho_{i,1}$, e_i original position.
$p = [p_1, ..., p_m]$, original positions vector.
$d_i = \delta_{i,1}$, original distance between e_i and its previous entity position.
$d = [d_i, ..., d_m]$, orginal distances vector.

The objective is the performing of a clustering problem by entities from an initial distribution of them in the ring through joining them with only the cut distance as external information and by means of using L to share the partial information that each e_i knows. The complexity of the language should allow a better or a worse result in the task. We will discuss some subjects about the proposed problem:

Does the absolute positions really matter? No. As we have defined, the final absolute positions of entities has no relevance, only adjacency has. If we have a ring with m entities with a determined distribution, the only information these entities have and share is about distance between them and their own identification. They only can know its relative position in relation to other entities and final adjacency would be the same if we move all entities a random number of positions left or right because relative positions would remain the same. d and not p is what really defines the initial configuration. From p, d is infered.

Is a limitation to allow only movement in one direction? Actually, there is no need to allow a two-direction movement, so it is a matter of simplicity. Suppose a ring with m entities. If e_i moves to right in the ring (assuming the cell at its right is empty), its position would be change to $\rho_{i,t} + 1$ (or 1 if $\rho_{i,t} = l$, we will explain the example with an entity is not at extreme cells or extreme index for simplicity) . Applying what we defined previously, what matter is $\delta_{i,t+1}$ would have been increased by 1 with respect to $\delta_{i,t}$ and $\delta_{i+1,t+1}$ would have been decreased by 1 with respect to $\delta_{i+1,t}$. It would be exactly the same if all entities but e_i moves to left so, to move to right e_1 only has to make all other entities move to left by mean of the language. Of course, it would be required a more complex language to achieve the same with a one-directional movement but in this case, that is an interesting feature in problem since the objective is to test languages.

Is actually necessary to allow two-side communication? We allow entities to share information with the entity at its left and its right. It would be simpler to allow only one-side communication but it would imply an important problem: suppose the same problem previously defined but with only one-side communication (i.e: each entity can exclusively “say” sentences to the entity as his right). Information gathering by each entity would be highly limited to one side (opposite to the side he speaks, left in this example) until the information would go over all the ring. So, for example, to know the information from its right, this information should came from left and the entity should be able to recognise it as information from its right. With no information from one side, the ability to perform a clustering task is very limited (absence of comparison) so the entity wouldn't know if move or not. To have information from two sides would imply a language complex enough to gather information from the whole world and it could limit problem utility to observe differences between languages below that point of complexity.

Is actually necessary for entities to know cut distance (C)? Clustering problem can be approached by two main sides: the first is to define the number of clusters we want to obtain (usually referred as classification, k-means algorithm would be the better example) and the second is to define a cut distance to demarcate the maximum similarity between two elements in the same cluster (with hierarchical clustering algorithm as great example). The selection of a cut distance over a number of objective clusters is not random. To classify a set of entities in k clusters there must be at least k entities or the problem has not sense, so the value k implies more external information about the world than C.

Other option would be to use an optimization index to minimize (like Davies-Bouldin Index, for example), this would change an absolute number by an equation to optimize. It would move our problem away from classical clustering approaches like well-studied hierarchical and k-means clustering, so we could use none of them as reference. This would complicate both programming and evaluation with no apparent language analysis improvement so it has been rejected.

Is truly important the lexicon? No. As we discussed previously, lexicon is a vital part of a language (without words, it would be no spoken or written language) but it is arbitrary.

2.3 Simulations approach

Our aim is to show how, by changing L, the resolution of the clustering problem can be improved. This is a bit ambiguous objective since given a clustering problem there is no absolute “best” clustering. Differents algorithms with differents distance functions and differing cut distances would produce differing results. So, our strategy will be to use a reference clustering algorithm to emulate its results.

We will define four increasing complexity languages (L), each one of them with an associated entities behaviour model (M). These behaviours will be focused on the utilization of information gained by the use of language. Through this, we will focus outcome improvements on language capacity to observe each language possibilities.

Each model will try to create the closest mental image to the original tape and then perform reference clustering algorithm by itself to deduce what other entities must be in its same cluster and use that information to know if move or not.

2.3.1 Evaluation

Each cluster c is a set of entities and there are m entities, so we can describe a bijective function between each cluster c an a vector v with m binary positions with $v(i) = 1 \Leftrightarrow e_i \in c$ and $v(i) = 0 \Leftrightarrow e_i \notin c$. In example, with $m = 5$ entities in ring, $c = \{e_1, e_3, e_4\}$, $v = [1, 0, 1, 1, 0]$.

Given two clusters c_1 and c_2 formed by entities and equivalent to vectors v_1 and v_2 respectively, we'll use as distance between c_1 and c_2 the well known Hamming distance:

$$H(c_1, c_2) = \sum_{v_1(i) \neq v_2(i)} 1 \equiv |c_1 \cup c_2| - |c_1 \cap c_2| = |c_1| + |c_2| - 2\,|c_1 \cap c_2| \tag{3}$$

Given two set of clusters S_1 and S_2 over the same total entities E with $|E| = m$, distance function H as defined previously and function min as the usual minimum function, we will define the distance between S_1 and S_2 as follows:

$$D(S_1, S_2) = \sum_{c_1 \in S_1} \underset{c_2 \in S_2}{min}(H(c_1, c_2)) + \sum_{c_2 \in S_2} \underset{c_1 \in S_1}{min}(H(c_2, c_1)) \tag{4}$$

Function D is minimal Hamming distance for each cluster from S_1 over S_2 plus vice versa.

Extreme cases:

$$0 < D(S_1, S_2) < (m+1)(m-1) \tag{5}$$

I) Sets are entirely different:
S_1 contains an unique cluster with all entities: $c_1 \in S_1, c_1 = E, |S_1| = 1$
S_1 contains m clusters, each one with an unique entity: $\forall e \in E,\ \{e\} \in S_2, |S_2| = m$

$$D(S_1, S_2) = (m-1) + m(m-1) = (m+1)(m-1) \tag{6}$$

Since $\forall c_2 \in S_2, c_2 \subseteq c_1,\ |c_2| = 1 \Rightarrow H(c_1, c_2) = H(c_2, c_1) = m - 1$

$$\sum_{c_1 \in S_1} \underset{c_2 \in S_2}{min}(H(c_1, c_2)) = \underset{c_2 \in S_2}{min}(H(c_1, c_2)) = m - 1 \tag{7}$$

And

$$\sum_{c_2 \in S_2} \underset{c_1 \in S_1}{min}(H(c_2, c_1)) = \sum_{c_2 \in S_2} m - 1 = m(m-1) \tag{8}$$

II) Sets are the same:

$$S_1 = S_2 \Rightarrow \forall c_1 \in S_1, c_1 \in S_2, \forall c_2 \in S_2, c_2 \in S_1 \Rightarrow D(S_1, S_2) = 0 \tag{9}$$

Example:

$$S_1 = \{\{e_1, e_2\}, \{e_3, e_4\}, \{e_5, e_6\}\}$$
$$S_2 = \{\{e_1, e_2, e_3, e_4\}, \{e_5, e_6\}\}$$
$$D(S_1, S_2) = (2 + 2 + 0) + (2 + 0) = 6$$

Given proposed clustering problem, original positions vector p, cut distance C, language L, behavior model M, function D and reference clustering algorithm R. With S_1 as R clustering result applied to p and C and S_2 as proposed clustering problem result applied with with p, C, M and L. The evaluation of S_2 respect to S_1 would be $D(S_1, S_2)$ so the lower, the better L and M combination.

Another option we will use is to evaluate the contrast between possibles clusters set is to suppose a third cluster set S_3 as the result of adjacency clusters A applied to p (the result if M is simply to do not move). S_3 would be the initial point and implies an adjacency clustering that could be even better than S_2 since M could break formed from the beginning clusters. This evaluation would be $D(S_1, S_3) - D(S_1, S_2)$ and it represents how has M improved clustering in relation to original distribution p.

As reference algorithm R to clustering performing by entities and compare results has been chosen classical agglomerative hierarchical clustering algorithm (AHC from now on). In this algorithm, from a m-sized positions vector p, a ring size l and a cut distance C, we start with m clusters (each entity is a cluster initially) and construct iteratively greater clusters by joining clusters whose distances are lower than C (Hastie et al. 2008).

As similarity function we will use the distance between clusters centroids. It is important to consider that position l is just next to position 1 and the distance between them are 1, as the distance between position $l-1$ and 1 is 3 and so on. Given two centroids c_1 and c_2 distance between centroids in a ring is previously defined function Δ.

2.4 Model designing general approach

All these algorithms but the first one (this will be explained in his own section) will have an associated language to use and will follow the same general scheme to perform clustering task. Each entity commence with empty mental imagen $\Omega_{i,1}$ and its objective is to get the mental image nearest to original vector position p through the sharing of distances. To achieve that, they follow the next phases:

- Information exchange: when entity e_i uses language to obtain and distribute information in order to add information to its own mental image as much as help other entities to do the same.
- Information use: entity e_i uses gathered information to decide if it should move or not. To perform this task, entity will apply the same clustering algorithm to its mental image of the world that will be used to compare the final result (AHC, as we have explained previously).
- Movement: e_i moves or not according to previous phase information process.

During information exchange phase, e_i will propagate information about last entity at its front it knows to its back and information about last entity at its back it knows to its front.

The inflection point change from the first phase to the second is what we call enough information evaluation that is, as sais its own name, an evaluation made by each entity to decide if it has enough information to decide if moves.

2.5 Model 1: Without a language

The first stage of language to resolve the problem must be the absence of any kind of communication. In this stage, entities will only gather information by observation so the only information each of them are able to gather is the distance between them and the entity at its left (δ_i).

This algorithm is not able to follow the general scheme we previously talk of because entities can not perform the first phase beyond gather the information they directly can "see". So, entity must work only with the distance between it, the one at its left and cut distance (C). Mental image will be only the last observation and the number of steps it has advanced to know how far it is from its initial position.

$$\Omega_{i,t} = ([\delta_{i,t}], k)$$

Model 1 for each e_i at iteration t acts as follows:

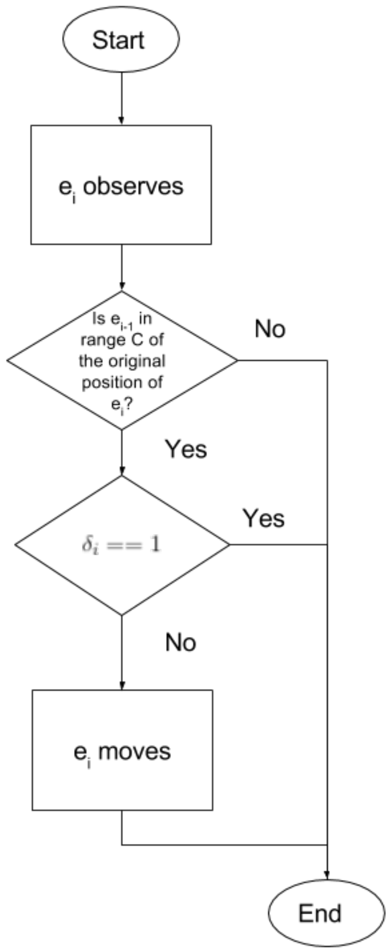

Figure 9: Model 1 flowchart

Example:

As can be seen, each entity will move only if the entity it can see is in range of C from its original position. Given position vector $p = [1, 3, 4, 7, 9]$ with cut distance $C = 4$ and ring length 10, the system would evolve as follows:

Table I: Model 1 evolution example

t	$\rho_{1,t}$	$\delta_{1,t}$	$\rho_{2,t}$	$\delta_{2,t}$	$\rho_{3,t}$	$\delta_{3,t}$	$\rho_{4,t}$	$\delta_{4,t}$	$\rho_{5,t}$	$\delta_{5,t}$
1	1	2	3	2	4	1	7	3	9	2
2	10	2	2	2	4	2	6	2	8	2
3	9	2	1	2	3	2	5	2	7	2
4	9	2	1	2	2	1	5	3	7	2

At third iteration only e_4 moves since its previous position is at distance 2 and it has only stepped once. At fourth iteration no one should move so it is final iteration.

Model relevant details:

- No information exchange.
- Ensures finalization

2.6 Model 2: With a simple language formed only by numbers

The second stage of language to resolve the problem is a very simple language allows only to communicate a number. The syntax is as simple as follows:

$S \to n$, with n as any natural number.

To interpret the syntax, we should remember that each sentence received by an entity is related to the side it has been sended ($\sigma_{i,k,t} = (\phi_{i,k,t}, o_{i,k,t})$). To mark sides we have selected words $front$ and $back$. The first to indicate the side entity "sees" (left from above point of view) and $back$ to represent the other side. So each sentence received by e_i would be a pair $(\phi_{i,t,k}, o_{i,t,k})$ with the following interpretation:

$(n, front) \Rightarrow d_{i-1} = n$
$(n, back) \Rightarrow d_{i+1} = n$

With this simple language, each entity e_i is able to only refer its own original distance d_i and can only obtain its relative distances to e_{e-2},e_{i-1} and e_{i+1} due to information interpretation restriction rule (*The interpretation from each sentence must defined only by the sentence itself, transmitter and receiver*). The first phase of our algorithm will consist in gather that information (one iteration will be enough). Once gathered, the enough information condition will be satisfied and the hierarchical algorithm will be applied to the mental image ($\Omega_{i,2}$) formed by obtained information by each e_i during first iteration.

This information will be structured in a relative distances vector where e_i will save its world vision. Due to the absence of external information about world disposition, e_i does not know its own absolute position. So, the information it obtains can only make sense as relative to its own position. In e_i world mental image, its position is 0, the position of e_{i-1} is $-d_i$ and so the information is structured like follows:

$$\Omega_{i,2} = ([d_i, d_{i-1}], [d_{i+1}])$$

Translated to the following vector:

$$v_i = d_{i-1} + d_i + 1 + [-d_{i-1} - d_i, -d_i, 0, d_{i+1}]$$

Information use performed by each e_i will consist in apply AHC to v_i and check if e_i is the first element of any obtained cluster. If e_i is the first, e_i will not move. If e_i is not its cluster first entity, e_i will maintain itself closest to its cluster immediate left companion.

The algorithm for each e_i acts as follows:

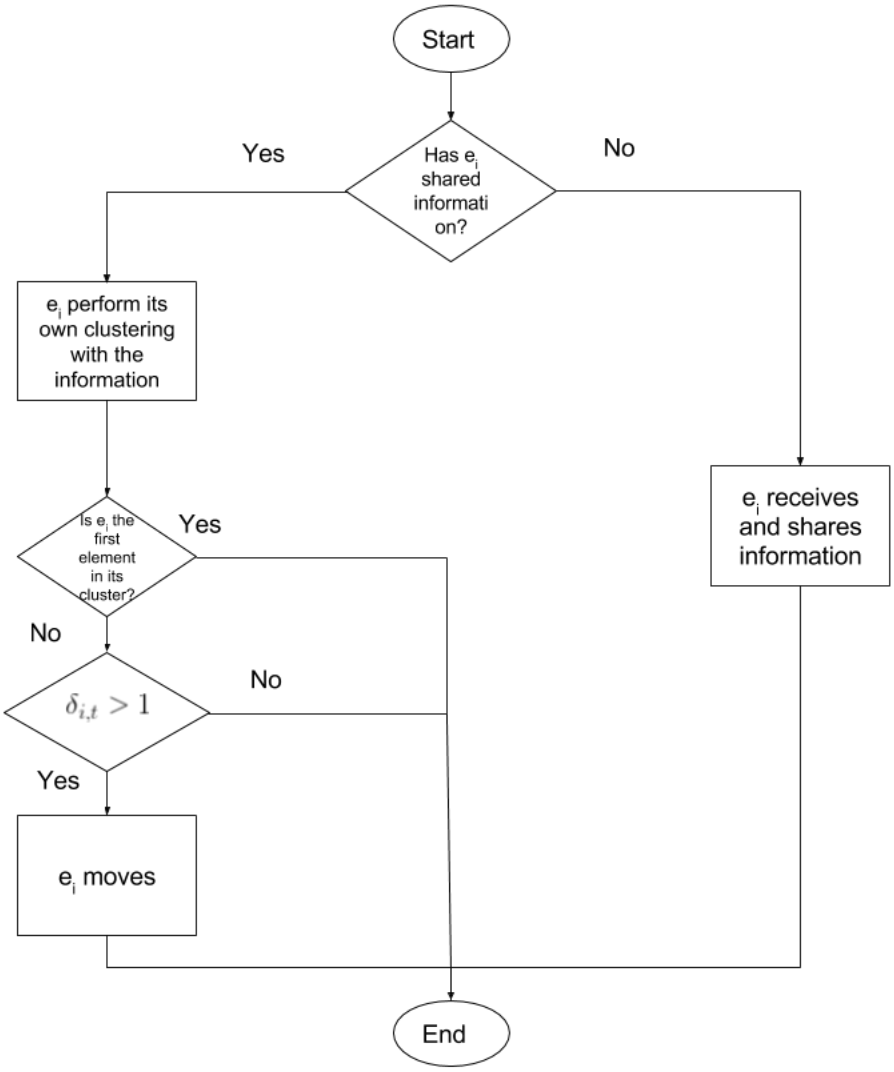

Figure 10: Model 2 flowchart

Model relevant details:

- Little information exchange about distances, no information exchange about identifications.
- With C large enough, all entities deduce they share cluster with other threes they know so all entities try to catch the previous entity and they start an infinite loop similar to Ant Mill problem commented previously.

2.7 Model 3: With language formed by pronouns and numbers

The third stage in language development will be a syntax where noun has been added. This change has the aim to compensate the previous language limitation. The syntax is like follows:

$S \rightarrow A\ n$ with n as a natural number.
$A \rightarrow f\ A$ with f as word referring to front.
$A \rightarrow f$ with f as word referrng to front.
$A \rightarrow b\ A$ with b as word referring to back.
$A \rightarrow b$ with b as word referring to back.

We should note, as we have previously used, the word $front$ is used to denote the previous entity and $back$ to denote the next one.

Pairs $(\phi_{i,t,k}, o_{i,t,k})$ interpretation by e_i will introduce a data $r_{z,j}$ defined as distance between e_z and e_j. This data $r_{z,j}$ needs some processing and previous data to be translated to d_j data by e_i. Specifically, $r_{z,j}$ has the following properties:

i $r_{z,j} = r_{j,z}$ - Symmetric relation. Distances are non negatives.

ii $r_{i,j} = r_{i,z} + r_{z,j}$ with $i < z < j$ or $i > z > j$ - The distance between e_i and e_j is defined by the addition of distance between e_i and e_z and distance between e_z and e_j if e_z is situated between e_i and e_j.

iii $r_{i,j} = |r_{i,z} - r_{z,j}|$ with $z < i < j$ or $z > i > j$ - The distance between e_i and e_j is defined by absolute value of difference of distances between e_i and e_z and distance between e_z and e_j if e_i is situated between e_z and e_j . The value must be absolute by (i).

iv $r_{j-1,j+1} = d_j + d_{j+1}$ - Total distance is composed by distances between two entities.

v $d_j = r_{j-1,j}$ - Distance between e_{j-1} and e_j is d_j as defined in problem formalization.

With the following pair $(\phi_{i,t,k}, o_{i,t,k})$ interpretation by e_i:

$(front\ n,\ back) \Rightarrow r_{i,i+1} = n \equiv d_{i+1} = n$ by (v).
$(front\ n,\ front) \Rightarrow r_{i-1,i-2} = n \equiv d_{i-1} = n$ by (v).
$(back\ n,\ back) \Rightarrow r_{i+1,i+2} = n \equiv d_{i+2} = n$ by (v).
$(back\ n,\ front) \Rightarrow r_{i-1,i} = n \equiv d_i = n$ by (v). This use has not sense since gathers directly d_i.
$(front\ P\ n,\ back) \Rightarrow r_{i+1,j-1} = n$, with $(P\ n,\ back) \Rightarrow r_{i+1,j} = n$.
$(front\ P\ n,\ front) \Rightarrow r_{i-1,j-1} = n$, with $(P\ n,\ front) \Rightarrow r_{i-1,j} = n$.
$(back\ P\ n,\ back) \Rightarrow r_{i+1,j+1} = n$, with $(P\ n,\ back) \Rightarrow r_{i+1,j} = n$.
$(back\ P\ n,\ front) \Rightarrow r_{i-1,j+1} = n$, with $(P\ n,\ front) \Rightarrow r_{i-1,j} = n$.

If $r_{i,j}$ and $r_{j,z}$ are data known by e_i, e_i obtains $r_{i,z}$ by (ii). So, if e_i knows its own distance to e_j and the distance between e_j and e_z, it can obtain its own distance to e_z.

This language allows entity e_i to express any distance d_j by transmitting enough information to be deduced by other entity.

Model 3 e_i mental image at iteration $t > 1$ with m entities in ring:

$$\Omega_{i,t} = (\alpha_{i,t}, \beta_{i,t})$$
$$\alpha_{i,t} = [d_{k+1}], min(k), k \equiv i - a - 1(mod\ m), a = 0, ..., t-1.$$
$$\beta_{i,t} = [d_{k+1}], min(k), k \equiv i + a - 1(mod\ m), a = 1, ..., t-1.$$

Guess a ring defined by the following vector of positions $p = [1, 4]$ and ring size $l = 5$. With $t = 4$, $d_1 = 2$, $d_2 = 3$ mental images would be like follows:

$$\Omega_{1,4} = ([2,3,2,3],[3,2,3])$$
$$\Omega_{2,4} = ([3,2,3,2],[2,3,2])$$

And vector v_i to process would be obtained from last iteration $\Omega_i = (\alpha_i, \beta_i)$ with a as α_i size and b as β_i size:

$$v_i = -1u_{i,1}(1) + 1 + [u_{i,1}, 0, u_{i,2}]$$
$$\forall k \in 1,..,a : u_{i,1}(k) = -1 \sum_{j=1}^{a-k+1} \alpha_i(j), \forall k \in 1,..,b : u_{i,2}(k) = \sum_{j=1}^{k} \beta_i(k)$$

So with previous example, $\alpha_1 = [-10, -7, -5, -2]$, $\beta_1 = [3, 5, 8]$, $\alpha_2 = [-10, -8, -5, -3]$, $\beta_2 = [2, 5, 7]$:

$$v_1 = 11 + [-10, -7, -5, -2, 0, 3, 5, 8] = [1, 4, 6, 9, 11, 14, 16, 19]$$
$$v_2 = 11 + [-10, -8, -5, -3, 0, 2, 5, 7] = [1, 3, 6, 8, 11, 13, 16, 18]$$

We must note that $\Omega_{i,t}$ is subjective, k index value existence in $\alpha_{i,t}$ is interpreted as d_{i-k+1} and in $\beta_{i,t}$ is interpreted as d_{i+k} by e_i but this does not even ensure e_{i-k+1} or e_{i+k} existence. This will be Model 3 greatest problem.

If an entity e_i knows about two distances d_j and d_k equal or greater than C and with $j \leq i$ and $i < k$, e_i knows about a distance at its front and a distance at its back that are equals or greater than cut distance so it is sure that e_{j-1} and e_j would be in different clusters like e_k and e_{k+1}. This is a totally valid condition about enough information but there is no certainty about the existence of d_j and d_k in the ring so there is no certainty about stop. In most cases, C is greater than any d_i so this condition will not be satisfied. Each e_i will face its own Halting problem so there is no more solution than force a upper limit. To Add a value T to force information gathering stop.

Model 3 algorithm for each at iteration acts as follows:

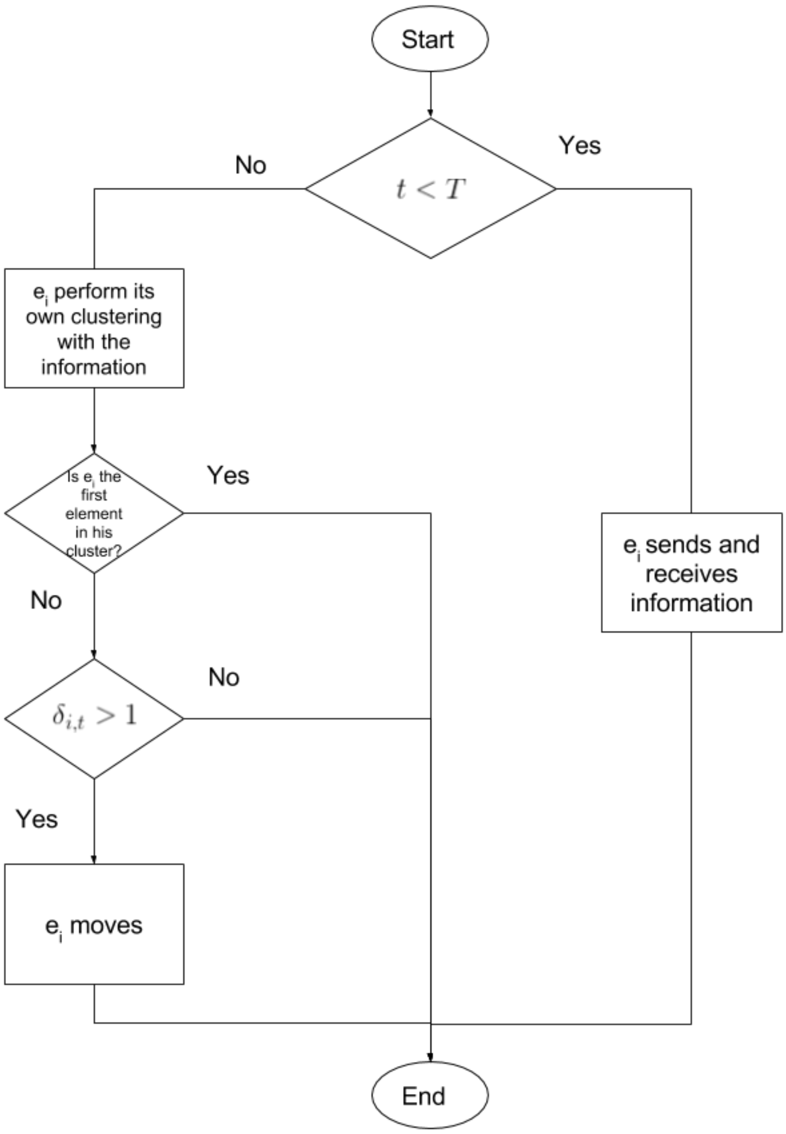

Figure 11: Model 3 flowchart

By using this language, each one of our entities is able to build a potentially infinite world mental image. This could appear to be enough in order to development a complete ring mental image by e_i but it is not the case. Problem rules imply e_i does not know how many entities in the ring are so, it cannot know when stop to gather information.

Guess a ring defined by the following vector of positions $p = [1,4]$ and ring size $l = 5$ and $C = 4$. From e_1 (position 1) point of view, i is its index so $d_i = 2$. From e_2 (position 4) point of view, j is its index so $d_j = 3$. Evolution of their information step by step would be like follows:

Table II: Model 3 evolution example

Action	e_1 knows	e_1 says to e_2	e_2 knows	e_2 says to e_1
They observe.	$d_i = 2$	-	$d_j = 3$	-
They share information.	$d_i = 2$	($front$ 2,$front$),($front$ 2,$back$)	$d_j = 3$	($front$ 3,$front$),($front$ 3,$back$)
They process information.	$d_i = 2$, $d_{i+1} = 3$, $d_{i-1} = 3$	-	$d_j = 3$, $d_{j+1} = 2$, $d_{j-1} = 2$	-
They share information.	$d_i = 2$, $d_{i+1} = 3$, $d_{i-1} = 3$	($front$ $front$ 5,$front$),($back$ $back$ 5,$back$)	$d_j = 3$, $d_{j+1} = 2$, $d_{j-1} = 2$	($front$ $front$ 5,$front$),($back$ $back$ 5,$back$)
They process information.	$d_i = 2$, $d_{i+1} = 3$, $d_{i-1} = 3$, $d_{i+2} = 5 - d_{i+1} = 2$, $d_{i-2} = 5 - d_{i-1} = 2$	-	$d_j = 3$, $d_{j+1} = 2$, $d_{j-1} = 2$, $d_{j+2} = 5 - d_{j+1} = 3$, $d_{j-2} = 5 - d_{j-1} = 3$	-

As can be seen in the table, e_1 and e_2 are feeding back each other and creating an erroneous ring mental image, this is due to the fact that they do not really know who is at their left or right. After three gathering information iterations they think there are at least 5 entities in the ring. This is an analogous situation to Ant Mill previously commented. As ants, entities are sharing information in circles. They do not actually have any information that indicates when they are repeating once and again the same. For that reason. To know if they have enough information to stop gathering and start using it is a complex matter. Using this language, entities are not able to gather information to know absolute references. It could be used an statistical approach and observe patterns repetitions to estimate a confidence threshold. We have prefered to evade an stochastic approach so we will not look for patterns due this not provide absolutely reliable information and it does not focus on language. Is impossible to create a language with given restrictions to allow enough information exchange to avoid this problem? No. It is actually possible to create a language like that, the following is an example:

$S \rightarrow D\ n$ with n as a natural number.
$S \rightarrow I\ n$ with n as a natural number.
$D \rightarrow fd\ D$ with fd as word referring to front entity distance.
$D \rightarrow fd$ with fd as word referring to front entity distance.
$D \rightarrow bd\ D$ with bd as word referring back entity distance.
$D \rightarrow bd$ with bd as word referring back entity distance.
$I \rightarrow fi\ I$ with fi as word referring to front entity identification.
$I \rightarrow fi$ with fi as word referring to front entity identification.
$I \rightarrow bi\ I$ with bi as word referring back entity identification.
$I \rightarrow bi$ with bi as word referring back entity identification.
$I \rightarrow m$ with m as word referring itself identification.

With $frontdistance$ as fd, $backdistance$ as bd, $frontidentification$ as fi, $backidentification$ as bi, me as m, the following pair $(\phi_{i,t,k}, o_{i,t,k})$ interpretation by e_i:

$(frontdistance\ n,\ back) \Rightarrow r_{i,i+1} = n \equiv d_{i+1} = n$ by (v).
$(frontdistance\ n,\ front) \Rightarrow r_{i-1,i-2} = n \equiv d_{i-1} = n$ by (v).
$(backdistance\ n,\ back) \Rightarrow r_{i+1,i+2} = n \equiv d_{i+2} = n$ by (v).
$(backdistance\ n,\ front) \Rightarrow r_{i-1,i} = n \equiv d_i = n$ by (v). This use has not sense since e_i gathers directly d_i.
$(frontdistance\ P\ n,\ back) \Rightarrow r_{i+1,j-1} = n$ with $(P\ n,\ front) \Rightarrow r_{i-1,j} = n$.
$(frontdistance\ P\ n,\ front) \Rightarrow r_{i-1,j-1} = n$ with $(P\ n,\ front) \Rightarrow r_{i-1,j} = n$.
$(backdistance\ P\ n,\ back) \Rightarrow r_{i+1,j+1} = n$ with $(P\ n,\ back) \Rightarrow r_{i+1,j} = n$.
$(backdistance\ P\ n,\ front) \Rightarrow r_{i-1,j+1} = n$ with $(P\ n,\ back) \Rightarrow r_{i+1,j} = n$.
$(frontidentification\ P\ n,\ back) \Rightarrow n_i = n$ This use has not much sense since e_i knows n_i.
$(frontidentification\ P\ n,\ front) \Rightarrow n_{i-2} = n$.
$(backidentification\ P\ n,\ back) \Rightarrow n_{i+2} = n$.
$(backidentification\ P\ n,\ front) \Rightarrow n_i = n$ This use has not much sense since e_i knows n_i.
$(frontidentification\ P\ n,\ L) \Rightarrow n_{j-1} = n$ with $(P\ n,\ L) \Rightarrow n_j = n$.
$(backidentification\ P\ n,\ L) \Rightarrow n_{j+1} = n$ with $(P\ n,\ L) \Rightarrow n_j = n$
$(me\ n,\ front) \Rightarrow n_{i-1} = n$.
$(me\ n,\ back) \Rightarrow n_{i+1} = n$.

With a language like that, an entity is able to exchange information about distance as well as about identification. By knowing entities identifications, entities can perform a match between what is in front of them and at their back and mount a complete ring image. We will expose in detail how this information can be used in the next stage. In this stage, we will not use this kind of language due to its expansion problem.

The "trick" of joining in a lexical element two semantic elements as "who" is related to the information we receive (pronouns) and "what" is that information meaning, is not acceptable if we want to expand a language efficiently. It implies the necessity of a lexical element for each pronoun-meaning combination. For example, if we have x pronouns elements and y possible meanings for a number n, we would need $x * y$ lexical elements. This supports language meaning on lexicon much more than on syntax. Although from a theoretical point of view, this would be acceptable, this is not the natural evolve way and it confront one of our objectives: efficiently, since it grows disproportionality in space.

In that language there is a lot of redundant information. A sentence would be, for example, like the following:

$frontidentification\ frontidentification\ frontidentification\ 4.$

In that sentence is repeated three times 4 has indentification meaning, so there is not only a problem with lexico size, also a redundancy problem.

We should note that, $frontidentification$ could seem an easy to remember word since it is a composed word but words are arbitrary signs. We chose composed words to simplify remember its meaning for the example.

Model relevant details:

- Complete information exchange about distances, no information exchange about identification.
- With C greater than all d_i, it starts an infinite exchange information loop and need and extra iteration limit to avoid it.
- Once information exchange has finished, it does not ensure finalization since it is teorically possible all entities tries to reach the previous one.
- Its erroneos information gathering affects mental image vector and can lead e_i to incorrect conclusions.

2.8 Model 4: With language formed by pronouns, verbs and numbers

Language syntax is like follows:

$S \rightarrow A\ B\ n$ with n as a natural number.
$A \rightarrow f\ A$ with f as a word referring to front (pronoun).
$A \rightarrow f$ with f as a word referring to front (pronoun).
$A \rightarrow b\ A$ with b as word referrig to back (pronoun).
$A \rightarrow b$ with b as a word referring to back (pronoun).
$A \rightarrow m$ with m as word referring to the entity itself (pronoun).
$B \rightarrow d$ with d as a word referring to distance (verb).
$B \rightarrow i$ with i as a word referring to identification (verb).

The word me will be used as m, $identification$ as i and $distance$ as d.

With the following pair $(\phi_{i,t,k}, o_{i,t,k})$ interpretation by e_i:

$(front\ distance\ n,\ back) \Rightarrow r_{i,i+1} = n \equiv d_{i+1} = n$ by (v).
$(front\ distance\ n,\ front) \Rightarrow r_{i-1,i-2} = n \equiv d_{i-1} = n$ by (v).
$(back\ distance\ n,\ back) \Rightarrow r_{i+1,i+2} = n \equiv d_{i+2} = n$ by (v).
$(back\ distance\ n,\ front) \Rightarrow r_{i-1,i} = n \equiv d_i = n$ by (v). This use has not much sense since e_i gathers directly d_i.
$(me\ distance\ n,\ back)$ is not defined.
$(me\ distance\ n,\ front)$ is not defined.
$(front\ identification\ n,\ back) \Rightarrow n_i = n$ This use has not sense since e_i knows n_i.
$(front\ identification\ n,\ front) \Rightarrow n_{i-2} = n$.
$(back\ identication\ n,\ back) \Rightarrow n_{i+2} = n$.
$(back\ identification\ n,\ front) \Rightarrow n_i = n$. This use has not sense since e_i gathers directly n_i.
$(me\ identification\ n,\ back) \Rightarrow n_{i-1} = n$.
$(me\ identification\ n,\ front) \Rightarrow n_{i+1} = n$.
$(front\ P\ n,\ back) \Rightarrow r_{i+1,j-1} = n$ with $(P\ n,\ back) \Rightarrow r_{i+1,j}$.
$(front\ P\ n,\ front) \Rightarrow r_{i-1,j-1} = n$ with $(P\ n,\ front) \Rightarrow r_{i-1,j}$.
$(back\ P\ n,\ back) \Rightarrow r_{i+1,j+1} = n$ with $(P\ n,\ back) \Rightarrow r_{i+1,j}$.
$(back\ P\ n,\ front) \Rightarrow r_{i-1,j+1} = n$ with $(P\ n,\ front) \Rightarrow r_{i-1,j}$.
$(front\ P\ n,\ L) \Rightarrow n_{j-1} = n$ with $(P\ n,\ L) \Rightarrow n_j = n$.
$(back\ P\ n,\ L) \Rightarrow n_{j+1} = n$ with $(P\ n,\ L) \Rightarrow n_j = n$.

Model 4 e_i mental image at iteration $t = 2n+1, n \in \mathbb{N}$ with m entities in ring:

$$\Omega_{i,t} = (\alpha_{i,t}, \beta_{i,t})$$
$$\alpha_{i,t} = \begin{bmatrix} d_{k+1} \\ n_{q+1} \end{bmatrix} min(k), k \equiv i - a - 1 \ (mod\ m)$$
$$min(q), q \equiv k - 1 \ (mod\ m) \begin{cases} a = 0, ..., min(n, \frac{m}{2}) & m \equiv 0 \ (mod\ 2) \\ a = 0, ..., min(n, \frac{m+1}{2}) & m \equiv 1 \ (mod\ 2) \end{cases}$$
$$\beta_{i,t} = \begin{bmatrix} d_{k+1} \\ n_{k+1} \end{bmatrix} min(k), k \equiv i + a - 1 \ (mod\ m) \begin{cases} a = 1, ..., min(n, \frac{m}{2}) & m \equiv 0 \ (mod\ 2) \\ a = 1, ..., min(n, \frac{m+1}{2}) & m \equiv 1 \ (mod\ 2) \end{cases}$$

Note Model 4 mental images are defined only with odds t greater than 1. It is because of each e_i alternates the sending of distance and identification information so there are t where there is no n_j exchange and it is useless to cross data.

Guess a ring defined by the following vector of positions $p = [1, 5]$ and ring size $l = 9$. With $t = 3$, $d_1 = 5$, $d_2 = 4$, $n_1 = 15$ and $n_2 = 23$, mental images would be like follows:

$$\Omega_{1,3} = (\begin{bmatrix} 5 & 4 \\ 23 & 15 \end{bmatrix}, \begin{bmatrix} 4 \\ 23 \end{bmatrix})\ \Omega_{2,3} = (\begin{bmatrix} 4 & 5 \\ 15 & 23 \end{bmatrix}, \begin{bmatrix} 5 \\ 15 \end{bmatrix})$$

So each e_i have enough information to look for a common n_j in $\alpha_{i,3}$ and $\beta_{i,3}$. That is enough information to deduce there is a ring and implies knowledge about all original distances in vector d. From vector d is trivial to obtain ring size l.

When e_i reach this point, it will process surplus information in $\Omega_{i,t}$ (delete columns on $\alpha_{i,t}$ or $\beta_{i,t}$ with repeated $n_{i,j}$). With α_i and β_i as distances vectors with no repeated information and a as α_i length and b as β_i length they would create a deduced positions vector v_i in a l size ring like this:

$$v_i = -1 u_{i,1}(1) + 1 + [u_{i,1}, 0, u_{i,2}]$$
$$\forall k \in 1, .., a : u_{i,1}(k) = -1 \sum_{j=1}^{a-k+1} \alpha_i(j), \forall k \in 1, .., b : u_{i,2}(k) = \sum_{j=1}^{k} \beta_i(k)$$

So, in previous example, with $\alpha_1 = [5]$, $\beta_1 = []$, $\alpha_2 = [4]$, $\beta_1 = []$:

$$v_1 = 6 + [-5, 0] = [1, 5]$$
$$v_2 = 5 + [-4, 0] = [1, 6]$$

Note that v_1 represent the same relative distances in a $l = 9$ size ring than v_2, so, both are correct but subjectives to e_i original relative position.

The algorithm for each e_i acts as follows:

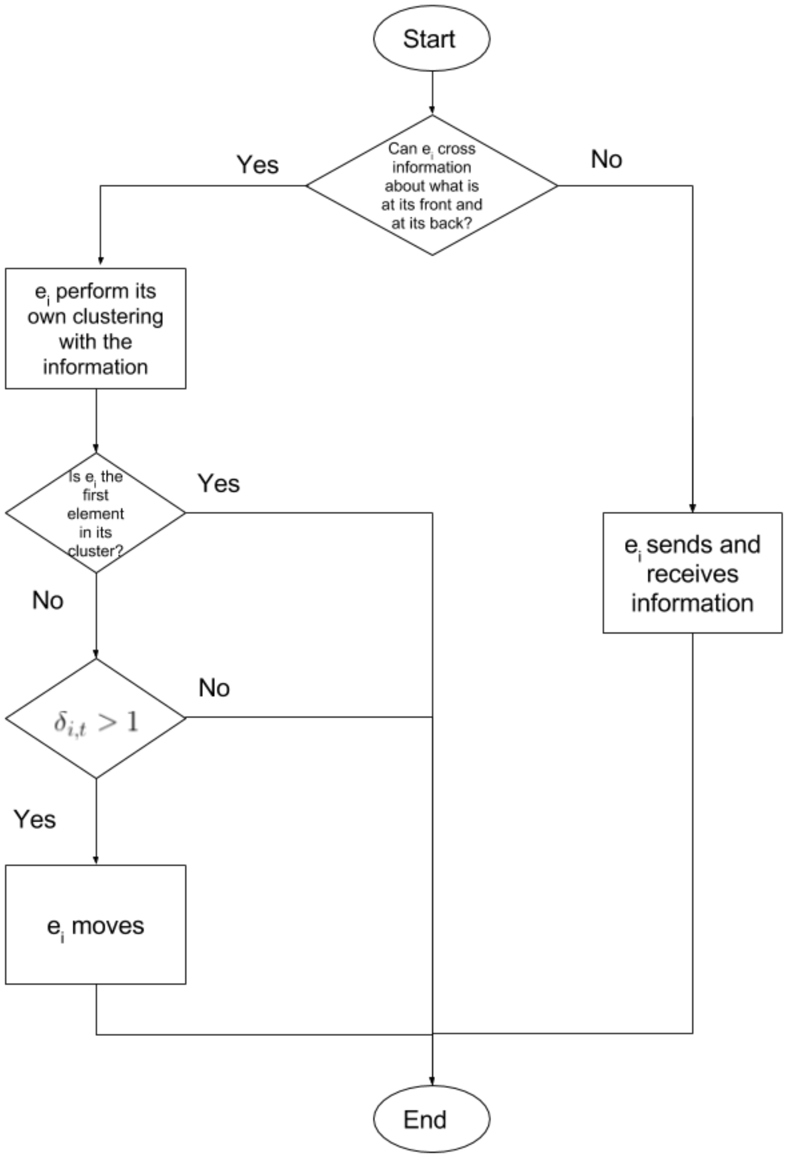

Figure 12: Model 4 flowchart

Model 4 will rely on identify a common entity in $\alpha_{i,t}$ and $\beta_{i,t}$ as explained previously. An example: Guess a ring defined by the following vector of positions $p = [1, 4]$ and ring length $l = 5$, $n_1 = 1$ and $n_2 = 2$.

From e_1 (position 1) point of view, i is its index so $d_i = 2$. From e_2 (position 4) point of view, j is its index so $d_j = 3$.

The evolution of their information step by step would be like follows:

Table III: Model 4 evolution example

Action	e_1 knows	e_1 says to e_2	e_2 knows	e_2 says to e_1
They observe and share information.	$d_i = 2$, $n_i = 1$	($front\ distance$ 2,$front$),($front\ distance$ 2,$back$)	$d_j = 3$, $n_j = 2$	($front\ distance$ 3,$front$),($front\ distance$ 3,$back$)
They process received information and share identification information.	$d_i = 2$, $n_i = 1$, $d_{i+1} = 3$, $d_{i-1} = 3$	($me\ identification$ 1, $front$),($me\ identification$ 1, $back$)	$d_j = 3$, $n_j = 2$, $d_{j+1} = 2$, $d_{j-1} = 2$	($me\ identification$ 2, $front$),($me\ identification$ 2, $back$)
They process information and cross it.	$d_i = 2$, $n_i = 1$, $d_{i+1} = 3 = d_{i-1}$, $n_{i+1} = 2 = n_{i-1}$,	-	$d_j = 3$, $n_j = 2$, $d_{j+1} = 2 = d_{j-1}$, $n_{j-1} = 2 = n_{j+1}$	-

Now, each entity has a correct ring mental image and it has been avoided Model 3 Ant Mill problem previously commented.

Model relevant details:

- Complete information exchange about distances and identifications.
- No exchange information infinite loop.
- Once information exchange has finished, it does not ensure finalization since is teorically possible all entities tries to reach the previous one.
- Each e_i mental image is subjetive so two differents entities could infer differents clusters.

3 Experiments

3.1 Experiments Details

We will use Fisher's iris flower data set to perform our experiments. Specifically vector sepal length (usually first column from data set) rounded (distances must be naturals) as initial distance vector d.

$$d(i) = round(iris(i,1)), i = 1, ..., 150$$
$$l = \sum_{i=1}^{150} d(i) = 886$$

Since no model ensures finalization, we have chosen $t = 3l$ with l as ring size as top iteration limit. Model 3 exchange information iteration limit has been defined as $T = 1.5l$. Since results change with dissimilarity value (cut distance C), we will apply each model from $C = 5$ to $C = 150$ with an incrementation of 5.

It is important to note that all distances in this vector are lower than 8 since this will affect model 2 performance and it implies model 3 will need given T for all but the first test.

$$d(i) < 8, i = 1, ..., 150$$

Project implementation was made with GNU Octave over an Ubuntu 16.04 LTS 64-bits operating system. Computer CPU is intel Core i5-6300HQ, 2.30GHz four cores. As code editor has been used Sublime-Text 2 and as code repository was used Mercurial-Bitbucket combination to milestone saves, Mercurial as code repository manager and Bitbucket as repository cloud manager. It was also used Dropbox as constant cloud backup.

Initial implementation strategy was to implement literally all languages and compute each information exchange. It was too slow (thousands of sentences to create, exchange and parse...) so final approach has been to avoid all literal language implementation. With $\Omega_{i,t}$ evolution defined for each model and known how final $\Omega_{i,t}$ is transformed to be processed as was discussed in corresponding sections, it has been much simplier to create directly transformed vectors working only pure information vectors and matrix.

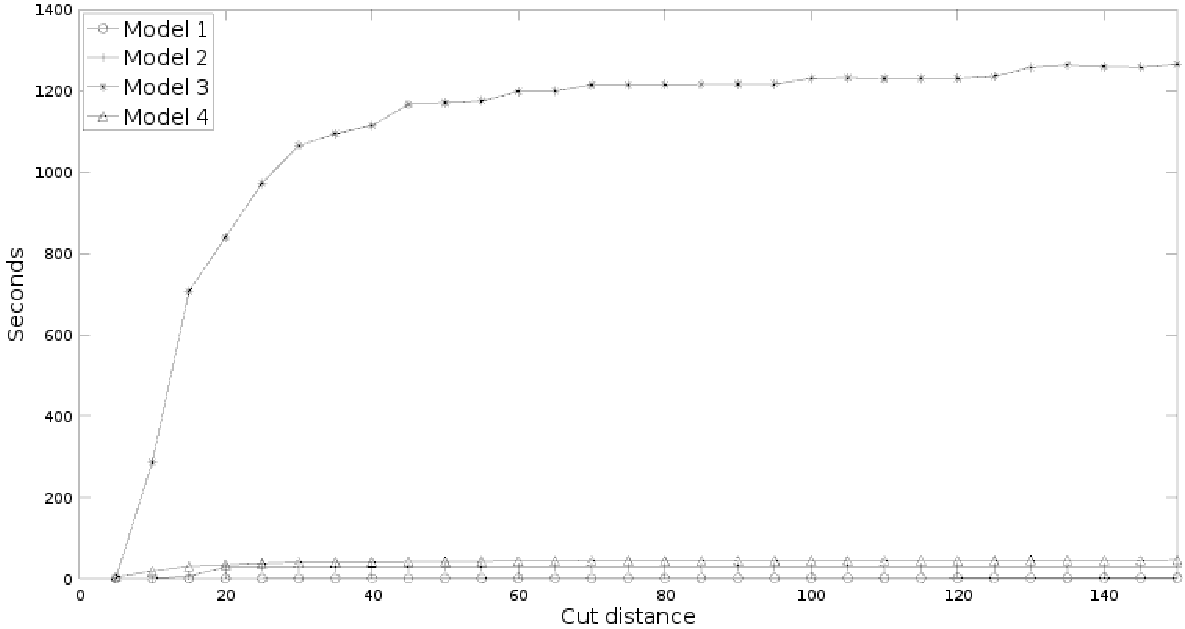

Figure 13: Execution times with C from 5 to 150

Model 1 execution times increase linearly from 0.05 to 1.58 seconds. Model 2 only finishes with three first C values, from that point on, Model 2 times have little variations and are close to 27 seconds. Model 3, as could be expected, is the slowest model. With $C = 5$, Model 3 does not requires to use top limit exchange information limit T but, from that point on, times increase abruptly to increase slowly from $C = 45$ on. Model 4 begins with the higher time but it reaches stability (close to 44 seconds) from $C = 70$ on. Model 1, 3 and 4 processes ends without reach top iteration limit in all cases.

There are two main code bottle necks:

The first bottle neck is to create information vectors for each entity to be processed. Basically, it consists in matrix manipulations and additions so it could be interesting to study if the use of a many-core focus language (like CUDA to use nVidia GPU) could improve performance.

The second bottle neck is information processing. Agglomerative hierarchical cluster is applied to vector infered from each $\Omega_{i,t}$ for each model but 1. So, in the worst case (Model 3) as with Fisher's iris data set there are 150 entities (or agents), hierarchical clustering is applied 150 times. Ring size l is equal to 886. Due to model 3 explained Ant Mill problem, and established iteration top limit $T = 1.5l$, each entity ends with a $1.5 * 886 * 2 = 2658$ length associated mental image vector (each iteration entity receives information from left and right). So, with that model, there are 150 vectors with 2658 positions each one to apply hierarchical algorithm. Those 150 vectors are very similar (each one is the previous desplaced one position) so there are many calculations repeated. The creation of a hierarchical clustering algorithm to compute clustering to many vectors simultaneous with this conditions could improve speed notably.

3.2 Results

Given S_1, S_2, S_3 and S_4 as adjacency clusters set results from Model 1, Model 2, Model 3 and Model 4 respectively, S_{AHC} as cluster set result from reference algorithm Agglomerative Hierarchical Clustering and S_0 as function A aplied to initial position vector p. First we will show total distance D between each model cluster set S_i and reference cluster set S_{AHC} (the lowest, the better since shows distance to reference).

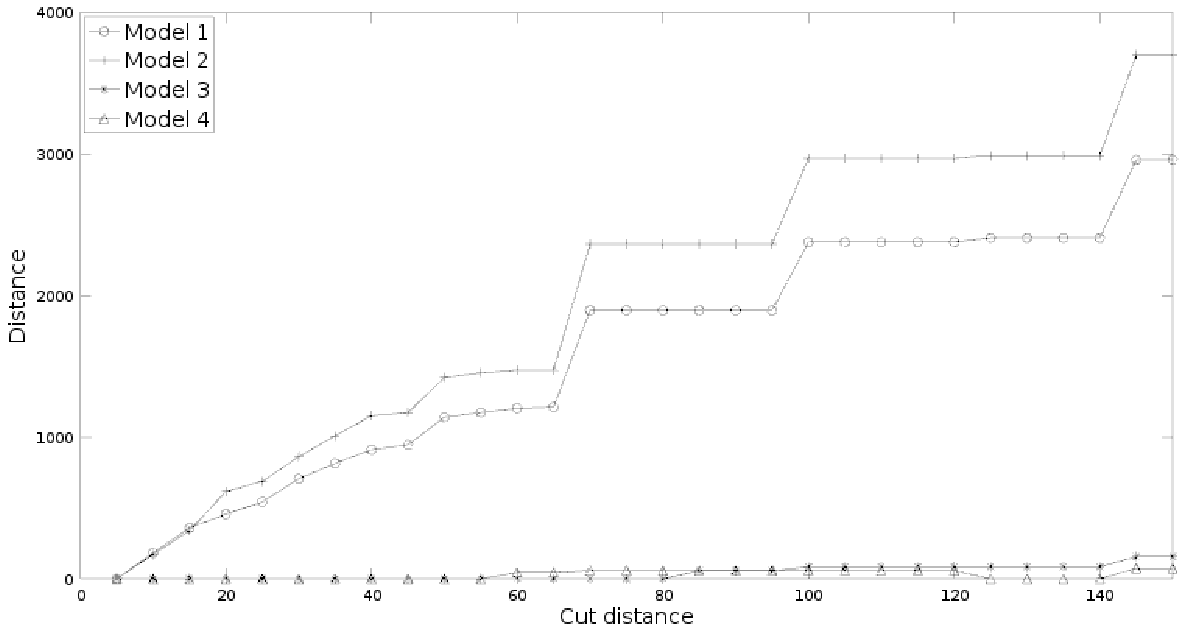

Figure 14: $D(S_i, S_{ACH})$ with C from 5 to 150

Next, we will show how evolves the difference between previously showed values and base value $D(S_0, S_{ACH})$ so it will be much more intuitive how good has been each model in improving adjacency clustering result from original position p adjacency.

This time the higher value the better since it shows the improvement regarding to do nothing.

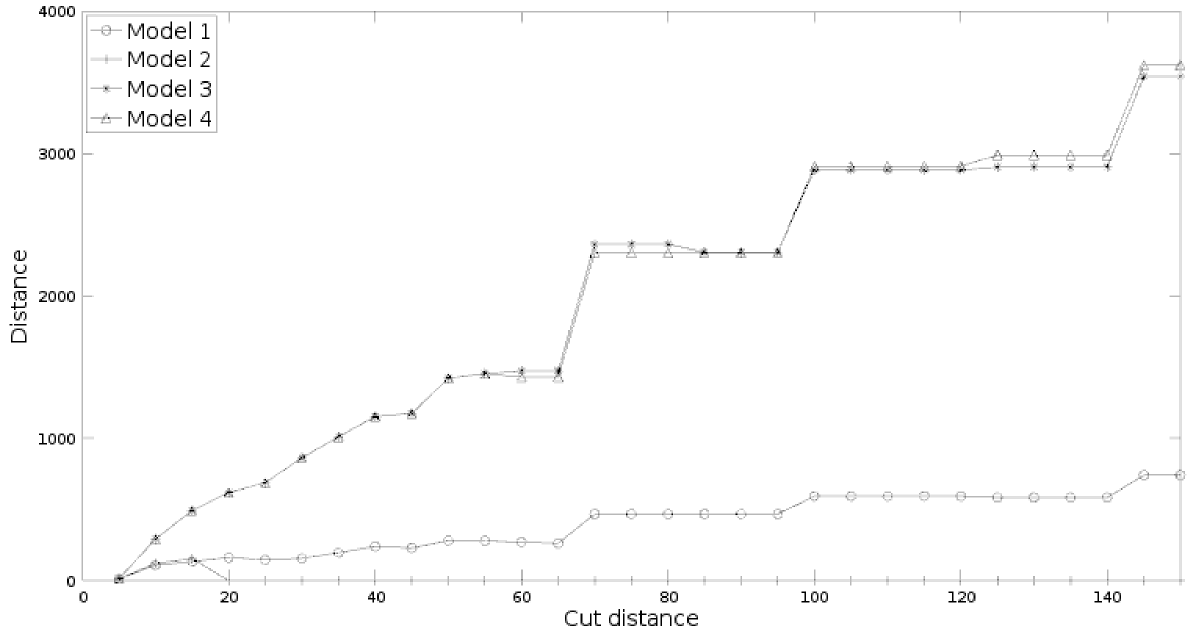

Figure 15: $D(S_0, S_{AHC}) - D(S_i, S_{AHC})$ with C from 5 to 150

Model 1 shows a simple behavior based only in observation with no information exchange efectiveness.

Model 2 shows how a language with a limited information exchange about a part of the problem based on a syntax that allows raw data with no context is worst than the simplest observation and, as we can see in figure 15, it shows results similars to do nothing for most C.

Model 3 shows how a language that allows complete information exchange about a part of problem (it exchange all information about d_j but no information about n_j) based on a SO (Subject, Object) syntax structure, has obviously better results. As figure 14 shows, outcome clusterings are equals or near to those from reference algorithm. We can not forget this model has needed an extra data to limit information exchange iterations (T) to avoid an infinite loop for any tested C but $C = 5$ and his execution times are the worst as figure 13 shows.

Model 4 shows how a language that allows the information exchange about two parts of problem (it allow d_j and n_j exchange) based on a SVO (Subject, Verb, Object) syntax structure has restuls similars to model 3. It is important to note that, unlike model 3, model 4 does not requiere a given exchange information iteration limitation. All this supports language with basic Subject, Verb and Object syntax structure utilization to face proposed problem.

4 Conclusions

Our experiments have involved the empirical study about what elements in usual SVO syntactic structure (Subject, Verb and Object) are useful for proposed problem resolution by a set of agents with partial information.

Results support basics structures with subject, verb and object presence are justified. If we have needed an structure with this features to face a simple problem like proposed one, it seems logic the conjunctions of problems general information exchanges by means of natural languages are applied, or an hypothetic artificial language between humans and machines would be applied, would need also a language with a basic structure at least as complex as SVO since a general purpose language must work on proposed problem.

It is interesting to think about what Model 2 results could imply. If we consider the evolution of natural languages as an evolutive process from the simplest language to current human languages, a basic language as the proposed one in Model 2 would be a problem to that approach. Hypotheticals humans with a language like that would have needed other means to information exchange or they would have applied the language with greats restrictions or that language would have been counter-productive.

From possible future work perspective, this project is only the first step in the search of an optimal human-machine language. Of course, more complex languages and behaviors models like those proposed could be studied. But, it seems more promising the evolutive generation of behavior models for a given language or even further: evolutive generation of pairs language-behavior with mutual interference to resolve the problem.

Those works not only could have interest from computer science point of view, to face language semantic studies related to problems solved by that language. Also, from cognitive point of view, those works could have important overcome to linguistics and biologicals matters to understand how different species had developed their languages, included human one.

As student, this work has been a challenge. To start, it has allowed me to learn about a problem as important in Artificial Intelligence as Natural language Processing but from an approach different to usual one. I was used to move through already explored study fields, this project showed an exciting perspective with an approach with no found similar in bibliography.

This has forced me to face a problem initially ambiguous and with no acceptable formalization as study if a language is or not optimal. The creation of reference problem as study case to approach the question has teached me a new way to face that kind of problems to, with no developed theory to work, shed light empirically. Problem refining process, its continuous simplification through deleting all unnecessary element and its final mathematical formalization to be used as test subject maybe has been the most instructive thing in the project. To conceive a problem complex enough to be useful but not enough to disturb our objective and obtaining mathematical tools to evaluate results has been actually educational.

Technically, I have improved my ability and knowledge with GNU Octave and using cloud repositories to face a project continuous modifications while maintaining olders versions to return in case of necessity.

5 Conclusiones

Nuestros experimentos han consistido en estudiar empíricamente qué elementos en la usual estructura sintáctica SVO (Sujeto Verbo Objeto) son útiles para la resolución del problema planteado por un conjunto de agentes con información parcial.

Los resultados respaldan que la presencia de estructuras básicas con sujeto, verbo y objeto en los lenguajes están justificadas. Si hemos necesitado una estructura con esas características para un problema tan simple como el propuesto, parece lógico que para la conjunción de problemas al que se aplica el intercambio de información general entre seres humanos mediante los lenguajes naturales o se aplicaría el intercambio de información general entre seres humanos y máquinas mediante un hipotético lenguaje artificial, sea también necesaria una estructura que al menos contenga esos tres elementos ya que un lenguaje de proposito general debería servir para resolver el problema propuesto.

Resulta interesante también reflexionar sobre lo que pueden implicar los resultados del Modelo 2. Si nos planteamos la posibilidad de que los lenguajes naturales se hallan desarrollado evolutivamente desde un lenguaje muy simple hasta los lenguajes actuales, un lenguaje básico como el del Modelo 2 resultaría un escollo para dicho planteamiento. Unos hipotéticos seres humanos con un lenguaje de esas características habrían necesitado de otros medios de intercambio de información o habrían aplicado el lenguaje de manera muy restringida o el lenguaje habría sido simplemente contraproducente.

Desde el punto de vista del posible trabajo futuro, este proyecto es solo un primer paso en la búsqueda del lenguaje óptimo para la comunicación hombre-máquina. Por supuesto, se podrían estudiar lenguajes más complejos asociados a modelos de comportamientos más depurados al igual que se han estudiado los modelos propuestos. Seguramente sería aún más prometedora la generación de modelos de comportamiento mediante algoritmos evolutivos para un lenguaje dado previamente e incluso ir aún más allá: la generación evolutiva del par lenguaje-comportamiento retroalimentados para resolver el problema.

Esos estudios no solo podrían tener interés desde el punto de vista informático de cara a la creación del hipotético lenguaje óptimo sino con vistas al estudio de la semántica de los lenguajes en relación al problema al que se aplican. Además, desde el punto de vista cognitivo, tanto en cuestiones lingüísticas, como biológicas podrían aportar información que permita comprender mejor cómo las especies desarrollan sus lenguajes y, más concretamente, cómo evolucionó esa capacidad en la nuestra.

Como alumno, este trabajo ha sido todo un desafío. Para empezar, me ha permitido aprender sobre un problema tan importante en Inteligencia Artificial como el Procesamiento del Lenguaje Natural pero enfocado desde un punto de vista distinto al usual. Acostumbrado a moverme por campos de estudio ya explorados, este proyecto presentaba la emocionante perspectiva de un enfoque nuevo para el que no se han encontrado similitudes en la bibliografía.

Esto me ha obligado a enfrentarme a un problema inicialmente ambiguo y sin una formalización aceptable como estudiar si un lenguaje es óptimo o no. La creación del problema de referencia como caso de estudio del que partir para abordar la cuestión me ha ilustrado una manera de enfrentar dicha clase de problemas para, en ausencia de una teoría desarrollada con la que trabajar, arrojar luz de manera empírica. El proceso de depurado del problema, su continua simplificación mediante la eliminación de aquello innecesario y su formalización matemática final de cara a ser utilizado como sujeto de pruebas quizás haya sido lo más instructivo de todo. Concebir un problema lo suficientemente complejo como para ser útil pero no lo suficiente como para tener elementos que estorbaran nuestros propósitos y obtener herramientas matemáticas para evaluar los resultados ha sido verdaderamente didáctico.

A nivel técnico, he profundizado en mi conocimiento de GNU Octave y en el uso de repositorios en la nube para enfrentar modificaciones continuas de un proyecto sin perder la posibilidad de volver a estadios anteriores del desarrollo.

6 References

Nadkarni, P. M., Ohno-Machado, L., Chapman, W. W. (2011). Natural language processing: an introduction. *Journal of the American Medical Informatics Association, 18*, pp. 544-551.

Hastie, T., Tibshirani, R., Friedman, J. (2008). Hierarchical Clustering. In T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*(pp.520-527). Stanford, California: Springer.

Meyer, C. F. (2009). Typological clasifications of languages. In C. F. Meyer, *Introducing English Linguistics*(pp.34-38). Cambridge: Cambridge University Press.

Gliozzo, A., Biran, O., Patwardhan, S., McKeown, K. (2013). Semantic Technologies in IBM Watson. *Teaching Natural Language Processing, Fourth Workshop*, pp. 85-92.

MacCartney, B., Manning, C. D. (2007). Natural logic for textual inference (Doctoral Thesis). *ACL Workshop on Textual Entailment and Paraphrasing*, Prague.

Pinker, S. (2003). Language as an adaptation to the cognitive niche. In M. H. Christiansen, S. Kirby (Eds.), *Language evolution: States of the Art*(pp.16-37). New York: Oxford University Press.

Janssen, T. M. V., "Montague Semantics", *The Stanford Encyclopedia of Philosophy* (Spring 2016 Edition), E. N. Zalta (ed.),
URL = <http://plato.stanford.edu/archives/spr2016/entries/montague-semantics/>.

Geurts, B., Beaver, D. I., Maier, E., "Discourse Representation Theory", *The Stanford Encyclopedia of Philosophy* (Spring 2016 Edition), E. N. Zalta (ed.),
URL = <http://plato.stanford.edu/archives/spr2016/entries/discourse-representation-theory/>.

Grüter, C., Balbuena, M. S. Farina, W. M. (2008). Informational conflicts created by the waggle dance. *The Royal Society B, 275*, 1321-1327.

Marsh, L., Onof, C. (2007). Stigmergic epistemology, stigmergic cognition. *Cognitive Systems Research, 9*, pp. 136-149.

Delsuc F. (2003). Army Ants Trapped by Their Evolutionary History. *PLoS Biology, 1*, pp. 155-156.

GNU Octave, Basic documentation manual (Web page). Retrieved throughout 2016. URL=<https://www.gnu.org/software/octave/doc/v4.0.0/>

Mercurial, Documentation (Web page). Retrieved throughout 2016. URL=<https://selenic.com/hg/help/>

7 Appendix A: Implementation Details

Main code files provided with this project are:

```
[finalpositions, finalclusterset] = ringclusteringreduced (model,
        nentity, cutdistance, positions, ringlength, boolshow)
```

This function is the code to perform proposed clustering as model defines. Function uses four matrix (F, Fid , B, Bid) as entities mental image vector so each $\Omega_{i,t}$ is equivalent to combination of i rows from each matrix at iteration t. F and B are used to contain front and fack distances and Fid and Bid to contain front and back identifications.

$finalpositions$ -> Natural numbers vector. Final entities positions.
$finalclusterset$ -> Natural numbers matrix. Cluster set defined by $finalpositions$ as defined in project. Clusters are provided as numbers pairs $[a, b]$ in each row. a is cluster beginning and b cluster final. If $b < a$, cluster includes from a to nentity and from 1 to b.
$model$ -> Natural number from 1 to 4. Model to use M.
$nentity$ -> Natural number. Number of entities. Used only if it is not provided an initial positions vector to obtain a vector based in prime numbers. cutdistance -> Natural number. Dissimilarity value, cut distance C.
$positions$ -> Not empty natural numbers vector. Originals entities positions to use in problem p.
$ringlength$ -> Natural number. Ring size l.
$boolshow$ -> Boolean. True to show each iteration evolution. Used mainly for debugging tasks.

```
value = evaluateclusteringwithhammingdistance (clusterset1,
        clusterset2, elementstocluster)
```

Function to evaluate the distance between two cluster with $elementstocluster$ as number of elements in clustering. Function D implementation. Clusters from the sames set must be disjunct.

$value$ -> Natural number.
$clusterset1$ -> Natural numbers matrix. Clusters are provided as numbers pairs $[a, b]$ in each row. a is cluster beginning and b cluster final. If $b < a$, cluster includes from a to nentity and from 1 to b.
$cluserset2$ -> Natural numbers matrix. Clusters are provided as numbers pairs $[a, b]$ in each row. a is cluster beginning and b cluster final. If $b < a$, cluster includes from a to nentity and from 1 to b.
$numberstocluster$ -> Natural number greater than 0.

```
clusterset = hierarchicalclustering (vectorpositions, cutdistance,
        ringlength)
```

Function to perform agglomerative hierarchical clustering over a vector keeping in mind it is peformed over a ring with $ringlength$ size.

$clusterset$ -> Natural numbers matrix. Clusters are provided as numbers pairs $[a, b]$ in each row. a is cluster beginning and b cluster final. If $b < a$, cluster includes from a to nentity and from 1 to b.
$vectorpositions$ -> natural number vector. Positions vector to perform clustering.
$cutdistance$ -> Natural number. Dissimilarity value, cut distance C.
$elementstocluster$ -> Natural number. Total number of entities in clusters. $ringlength$ -> Natural number. Ring size. If 0, function will act as if there is no ring so, last position in $vectorpositions$ is not next to first one.

```
move = processinformation1 (clusterset, mentalimagepositionvector,
        entitypositionindex, ringlength)
```

Function to infer if an entity must move or not from a set of clusters, a vector representing entity's image of the world, its position in that world and ring size. Entity should move if it is not the first in his assigned cluster or if, being the first and having all his cluster partners nearest as is possible, there is no space between its cluster and the next.

If $ringlength$ is greater than 1 and there is an only cluster, function should not move if is the entity with greater distance between it and the previously. This avoid an infinite loop when performing clustering. If all entities are in the same cluster, no on is the first, so no one would be motionless.

$move$ -> Boolean value. True/1 if entity must move. False/2 if entity must keep its position.
$clusterset$ -> Natural numbers matrix. Clusters are provided as numbers pairs $[a, b]$ in each row. a is cluster beginning and b cluster final. If $b < a$, cluster includes from a to nentity and from 1 to b.
$mentalimagepositionvector$ -> Natural numbers vector. It represents how the entity thinks the world is.
$entitypositionindex$ -> Natural number greater than 0. It represents entity's index in $mentalimagepositionvector$.
$ringlength$ -> Natural number. It represesnts how long entity thinks ring is. If 0, it is not kept in mind ring shape so, last position in $mentalimagepositionvector$ is not treated as next to first and vice versa.