Lexical Functions for Ants Based Semantic Analysis

Didier Schwab, Mathieu Lafourcade

Lexical Functions for Ants Based Semantic Analysis

Didier Schwab
Computer-Aided Translation Unit (UTMK)
School of Computer Sciences, Universiti Sains Malaysia,
Penang, Malaysia
Email: didier@cs.usm.my

Mathieu Lafourcade
Équipe TAL
Université Montpellier II-LIRMM,
Montpellier, France
Email: lafourcade@lirmm.fr

***Index Terms*—semantic analysis, lexical functions, conceptual vectors, lexico-semantic networks, Ant colony algorithms**

***Abstract*—Semantic analysis (SA) is a central operation in natural language processing. We can consider it as the resolution of 5 problems: lexical ambiguity, references, prepositional attachments, interpretative paths and lexical functions instanciation. In this article, we show the importance of this last and explain why these tasks should be simultaneously carried out using thematic (conceptual vectors) and lexical (semantico-lexical network) information. We present an ant colony model which fulfill these criteria. We show the feasability of our approach using a small corpus and the contribution of lexical functions for solving the problem. This ant colony model offers new and interesting research perspectives.**

Many Natural Language Processing applications, like automatic summarization, information retrieval or machinal translation, can take advantage of semantic analysis (SA) which consists of, among other things, computing a thematic representation for the whole text and for its subparts. In our case, thematic information is computed as conceptual vectors which represent ideas and provide a quick estimation if texts, paragraphs, sentences or words are in the same semantic field, i.e. if they share ideas or not. At least five main problems should be solved during a SA. (1) *lexical ambiguities* (2) *references* i.e. resolving anaphora and identity referencing ; (3) *prepositional attachments* i.e. to find the syntactic head to which a prepositional phrase is linked ; (4) *interpretation paths* which concerns the resolution of compatible ambiguities; (5) the most important for us in this article, *instanciation of lexical function* (LF).

LFs model typical relations between terms and include synonymy, the different types of antonymies, intensification ("*strong fear*", "*heavy rain*") or the typical relation of instrument (‹*knife*› is the typical instrument of ‹*to cut*›, ‹*shovel*› of ‹*to dig*›). In this article, we show that we need lexical functions to model world knowledge ("*Napoleon was an emperor*") or language knowledge (‹*destiny*› is synonym of ‹*fate*›) and the central role they play both in SA while contributing to the resolution of ambiguities mentioned earlier and also adressing specific problems of individual applications. We will see that their detection in texts require thematic and lexical information. Thematic information is handled using conceptual vectors which allows us to describe ideas contained in any textual segment (document, paragraph, sentence, phrase, ...) . Lexical information is addressed using a lexical network. Thus, our objective is to solve the five phenomena using a semantic lexical base whose lexical objects are linked to each others by typical relations and associated with conceptual vectors describing ideas they convey.

Usually, resolution of these phenomena are done separatly. Thus, anaphora resolution, prepositional attchment problem and especially lexical disambiguation are independently studied. However, this is not the approach we adopt here. Instead, our work is based on the reasonable assumption that these ambiguities are often interdependent and that it would be advantageous to undertake these tasks in a holistic way.

A way to holisticly deal with these various problems is to use a technique resulting from the distributed artificial intelligence, meta-heuristic of *ant colony algorithms*. Inspired by the collective behavior of biological ants, these algorithms are used to resolve difficult problems, in particular those related to graphs (TSP, partitionings, ...) and are used in operational research or to solve network routings problems. Ant colony algorithms are used in a different way for SA. It is not a method among others to solve a problem but rather a method which allows the simultaneous and interdependent resolution of these various tasks. Each ant caste corresponds to a heuristic which helps to solve a particular problem (in the model presented, detection of a particular lexical function) and has a behaviour influenced in part by the other ant activities. The environment is made up of both the text morpho-syntactic tree and a lexical network which contains typical relations between terms. We have one nest for each word meaning (acceptions) which competes during resource foraging. Ants build bridges between compatible acceptions which can be considered as sentence interpretations. We demonstrate the efficiency of this approach in order to solve SA problems.

I. Semantic Analysis (SA)

Five semantic phenomena can be solved during a SA:

(1) **Lexical Ambiguity** : Words can have several meanings. This well-known phenomenon leads to one of the most important problems in NLP, lexical disambiguation (also often called Word Sense Disambiguation). It involves selecting the most appropriate acception of each word in the text. We define an acception as a particular meaning of a lexical item acknowledged and recognized by usage. It is a semantic unit acceptable in a given language. For example, we can consider that ‹*mouse*› has three acceptions: the nouns for the ‹*computer device*› and for the ‹*rodent*› and the verb for the ‹*hunt*› of the animal. Contrary to lexical items, acceptions

are thus monosemantic. WSD is certainly a widely studied problem in SA [11]. For MT, it is essential to know which particular meaning is used in the source text because their translations are often different. For example, the English word ‹*river*› can be translated in French as ‹*fleuve*› or ‹*rivière*›. In information retrieval, it helps to eliminate documents which contain only inappropriate senses of a word according to the request, thereby increasing recall and precision.

(2) **References** : They are two types: (1)*Anaphora* is the phenomena whereby a pronoun is properly related to another element of the text. For example, in "*The cat climbed onto the seat, then it began to sleep.*", "it" refers to "cat" and not to "seat". Anaphoric resolution in MT is important as it associates pronouns to content nouns. Indeed, genders often vary according to the language. Thus, anaphoric resolution can help to translate the word which supports it. Therefore, in French, "it" can be translated either as "il" (masculine), as here in our case, or "elle" (feminine) whereas in German it could be either "er", "sie" or "es" since German has three genders. Note that in German the pronoun would be ă"sie" (feminine) and not masculine, as in French ("*Die Katze klaetterte auf den Sitz und (sie) begann dann zu schlafen*"). (2) *Identity* stands when two words in a text are references to the same entity such as "cat" and "animal" in the sentences "*The cat climbed onto the chair. The animal began to sleep.*".

(3) **Prepositional attachment** concerns finding the dependence link between a prepositional phrase and a syntactic head (verb, noun, adjective) [9]. In "*He sees the girl with a telescope.*" the prepositional phrase "*with a telescope*" can be attached to the nominal phrase "*the girl*" or to the verbal phrase "see". This is crucial in MT especially for a language like English where prepositions considerably modify verb meaning. In "*The man took a ferry across the river.*", the most logical attachment for ‹*across*› should be the verb ‹*to take*›. We then have for French "*L'homme traversa la rivière en ferry.*". The attachment to ‹*ferry*› gives another meaning and then gives as a translation "*L'homme pris un ferry à travers la rivière.*".

(4) **Interpretation path** : due to other ambiguities, a sentence can have several interpretations. Such ambiguities occur often especially if the text is short since there is less available information. [17] presents discussions and examples on this phenomenon. As an example, "*The sentence is too long.*" can be interpreted as a phrase with a non-trivial length or as a condemnation with a non-trivial duration.

(5) **instanciations of Lexical functions for analysis** which is a central point of this article and is presented now.

II. Lexical Functions

A. Lexical and World Knowledge

The existence of a distinction between lexical knowledge (LK) and world knowledge (WK) has been the subject of a great debate particularly since the beginning of the 1980's. According to John Haiman [10], there is no difference between the two, while Wierzbicka [22] argues that they are completely different. An interesting review can be found in Kornél Bangha's PhD. thesis [1] about the status of lexical knowledge versus world knowledge in the process of interpretation. Here, we adopt an intermediary stand close to his one's. We consider that knowledge can be divided into three categories: (1) *WK which are not directly lexicalised*. Thus they are not LK. For example, someone can know some facts of geography (Where is New York?), of history (How did JFK die?) or of everyday life (What is the color of a horse?). However, these information are not lexicalised and can be expressed only through statements; (2) *WK which are directly lexicalised*. As an example, the sentence "*During monsoon season, Penang has heavy rain*" is the representation in the real world of the amount of rainfall in Penang during Monsoon lexicalised thanks to ‹*heavy*›; (3) *some LK which can't be considered as lexicalisation of WK*. This is the case for grammatical gender in languages like French or German. Thus, the French lexical items ‹*voiture*› (‹*car*›) and ‹*mare*› (‹*pool*›) are feminine that does not correspond to any information on the objects.

B. LF for Linguistic Knowledge (LFLK)

LFLK are similar to Mel'čuk's LF [15]. They model LFs which correspond to linguistic knowledge. One must be aware of the fact that these functions also represent a state of the world, but this state is represented by a particular, but arbitrary (synchronically) item in the language. Thus, the sentence "*John had a strong fear*" corresponds to the real world situation describing the intense fear experienced by John, and is lexicalised by the *magnitude* LF *Magn* and one of its values, ‹*strong*›. There are two kinds of LFLK, *paradigmatics* which formalise classical semantic relations and *syntagmatics* which formalise collocations, "*combinations of lexical items which prevail on others without any obvious logical reason.*" [18].

In the first category we have: *synonymy* (*Syn*) which characterises different forms with a same meaning which is only given by use and without direct relationship to reality. *Syn*(‹*plane*›)={‹*airplane*›, ‹*aeroplane*›,...}; *antonymies* (*Anti*) which concern items whose semantic features are symetric relatively to an axis. *Anti*(‹*life*›) = {‹*death*›,...}; *Anti*(‹*hot*›) = {‹*cold*›,...} *generics* (*Gener*) which correspond to substitution hypernyms i.e. terms of the hierarchy which are preferred to others as reference by use. To illustrate, we do not say "*The vehicle has landed*" but "*the aircraft has landed*" so Gener(‹*plane*›) = {‹*aircraft*›} but not Gener(‹*plane*›)≠{‹*vehicle*›}. This function is different from hypernymy where Hyper(‹*plane*›)={‹*aircraft*›,‹*vehicle*›}.

In the syntagmatics, we have, *adjectival* LF like intensification (*Magn*) or confirmation (*Ver*). Magn(‹*tea*›)={‹*strong*›}; *Magn*(‹*rain*›)={‹*heavy*›}; *Ver*(‹*agreement*›)={‹*good*›, ‹*positive*›,...}; *collective* Mult(‹*dog*›)={‹*pack*›} and its opposite Sing Sing(‹*rice*›)={‹*grain*›}

C. LF for the World Knowledge (LFWK)

LFWK permit to model knowledge about the world. Among the LFWKs, we have, *hypernymy* (*Hyper*) which is the class hypernymy contrary to *Gener* which is the substitution hypernymy. As we have already mentioned, the world knowledge "*a chair is a seat* " is retranscribed in language by the fact that ‹*seat*› is hypernym of ‹*chair*› which is a LK. Hyper(‹*plane*›)={‹*aircraft*›,‹*vehicle*›,...}; it's opposite relation, *hyponymy*. Hyponymy can be seen as the transcription in

language of the property that a class is subclass of another. Hypo(‹*aircraft*›)={‹*plane*›}, Hypo(‹*vehicle*›)={‹*plane*›,‹*car*›,‹*boat*›}; *instance*(*Inst*) : *Inst*(‹*writer*›)={‹*Ernest Hemingway*›, ‹*Victor Hugo*›, ...}, *Inst*(‹*horse*›) = {‹*Tornado*›, ‹*Black*›,...}; its opposite relation, *Class* : *Class*(‹*Ernest Hemingway*›)={‹*writer*›, ‹*American*›, ...}, *Class*(‹*Black*›) = {‹*horse*›,...}; *meronymy* (*Mero*), the part-of relation and its opposite *holonymy* (*Holo*). *Mero*(‹*plane*›)= {‹*fuselage*›, ‹*wing*›,...}; verbal relations as *instrument* (*Instr*) which links an action to its typical instrument *Instr*(‹*to dig*›)= {‹*pick*›,...} *Instr*(‹*to write*›)= {‹*pen*›, ‹*keyboard*›,...} the *agent* relation (*agt*) which links an action to its typical agent and *patient* which links an action to its typical patient which is influenced by it. *agt*(‹*to eat*›)= ‹*cat*›; *pt*(‹*cat*›)= ‹*food*›.

D. Using of Lexical Functions

1) For Applications:

Machine translation: Igor Mel'čuk introduced lexical functions in MT because he noticed that some terms are associated to others whereas their direct equivalents are not used to mark a similar idea. Thus, we speak of "*grosse fièvre*" in French but not of *"*big fever*" in English, where "*high fever*" will be used instead. These phenomena were thus model by lexical functions. They can be applied to any language in the same manner and are considered universal. In MT, LF can be used as an *interlingua* i.e. as an intermediate language.

Information Retrieval: can be divided into two phases. The first one, *documents indexing* consists of building a computational representation for each document. The second one, the *search phase*, consists of transforming the request in similar representation and to extract the closest documents according to the given criteria. Lexical function can be useful to find synonymy of values. For example, we can imagine that the text representation does not directly refer to text segments like "*a high fear*" or "*crushing majority*" but rather to $Magn$(‹*fear*›) and $Magn$(‹*majority*›). Then, documents with "*a high fear*" or "*a strong fear*" and "*crushing majority*" or "*landslide majority*" would be more easily found than with simple distributional systems like SMART [19] or LSA [5].

2) For solving semantic analysis Problems: LFs can provide some clues which can help in the various tasks discribed:

Lexical Disambiguation : The two lexical function types can help us: (1) *LFLK* for identifying the syntagmatic relations between two words or at least to estimate its existence can help to identify the possible meanings for the corresponding lexical item. Thus, in "*For his recent election to the senate, Mr Smith obtained a crushing majority.*" ‹*majority*› can be partly disambiguated thanks to the LF $Magn$. Indeed, we can consider that ‹*majority*› can have as possible meanings the proportion which is related to the **age**, the **vote** or the one which is related to **assembly** but only $Magn$(*majority*/**vote**) = ‹*crushing*› and $Magn$(*majority*/**assembly**) = ‹*crushing*› exist. In the same way, synonyms or generics can indirectly contribute to the clarification via identity relation; (2) *LFWK* because they formalise world relations which can exist between the terms. Thus information such as "*Renault has connection with cars*" or "*Napoleon was an emperor*" can contribute to lexical disambiguation. Clarification can be done again here indirectly by identifying the identity relations thanks to hypernymy or instantiation.

Identity Relations Identification : These relations are partly supported by equivalent terms in context. They can be synonyms but also hypernyms. Knowing or identifying these relations can thus be a determining element for the meaning reconstitution.

III. LF instanciation:Lexical-Thematic Info

A. Thematic Information : Conceptual Vectors

We represent thematic aspects of textual segments (documents, paragraph, phrases, etc) by conceptual vectors. Vectors have long been used in information retrieval [19] and for meaning representation in the LSI model [5] from latent semantic analysis (LSA) studies in psycholinguistics. In computational linguistics, [3] proposed a formalism for the projection of the linguistic notion of semantic field in a vectorial space, from which our model is inspired. From a set of elementary concepts, it is possible to build vectors (conceptual vectors) and to associate them to any linguistic object. This vector approach is based on known mathematical properties. It is thus possible to apply well founded formal manipulations associated to reasonable linguistic interpretations. Concepts are defined from a thesaurus (in our prototype applied to French, we used Larousse thesaurus [13] where 873 concepts are identified)). Let $\mathcal{C}$ be a finite set of n concepts, a conceptual vector V is a linear combinaison of elements c_i of $\mathcal{C}$. For a meaning A, a vector $V(A)$ is the description (in extension) of activations of all concepts of $\mathcal{C}$. For example, the different meanings of ‹*door*› could be projected on the following concepts (the *concept*⌈*intensity*⌋ are ordered by decreasing values): V(‹*door*›) = (*opening*⌈*0.8*⌋, *barrier*⌈*0.7*⌋, *limit*⌈*0.65*⌋, ...

Comparison between conceptual vectors is done using angular distance. For two conceptual vectors A and B, $D_A(A,B) = arccos(Sim(A,B))$ where Sim is $Sim(X,Y) = \cos(\widehat{X,Y}) = \frac{X \cdot Y}{\|X\| \times \|Y\|}$. Intuitively, this function constitutes an evaluation of the *thematic proximity* and measures the angle between the two vectors. We would generally consider that, for a distance $D_A(A,B) \leq \frac{\pi}{4}$ (45°), A and B are thematically close and share many concepts. For $D_A(A,B) \geq \frac{\pi}{4}$, the thematic proximity between A and B would be considered as loose. Around $\frac{\pi}{2}$, they have no relation. D_A is a real distance function. It verifies the properties of reflexivity, symmetry and triangular inequality. We have, for example, the following angles(values are in radian and degrees). D_A(V(‹*tit*›), V(‹*tit*›)) = 0 (0°); D_A(V(‹*tit*›), V(‹*sparrow*›)) = 0.35 (20°); D_A(V(‹*tit*›), V(‹*bird*›)) = 0.55 (31°); D_A(V(‹*tit*›), V(‹*train*›)) = 1.28 (73°); D_A(V(‹*tit*›), V(‹*insect*›)) = 0.57 (32°)

The first one has a straightforward interpretation, as a ‹*tit*› cannot be closer to anything else than to itself. The second and the third are not very surprising either since a ‹*tit*› is a kind of ‹*sparrow*› which is a kind of ‹*bird*›. A ‹*tit*› has not much in common with a ‹*train*›, which explains the large angle between them. One may wonder why ‹*tit*› and ‹*insect*›, are rather close with only 32ř between them. If we scrutinise the definition

of ‹*tit*› from which its vector is computed (*Insectivourous passerine bird with colorful feather.*) perhaps the interpretation of these values would seem clearer. Indeed, the thematic distance is by no way an ontological distance.

B. Limitation of Conceptual Vectors For LF Detection

As shown in [2], distances computed on vectors are influenced by shared components and/or distinct components. Angular distance is a good tool for our aims because of its mathematical characteristics, its simplicity to understand and to linguistically interpret and ultimately allow it efficient implementation. Whatever chosen distance, used on this kind of vectors (represanting ideas and not term occurences), the smaller the distance, the bigger the number of lexical objects in the same semantic field (Rastier call it isotopy).

In the framework of SA as outlined here, we use angular distance to take advantage of mutual information carried by conceptual vectors in order to make disambiguate words pertaining to the same or closely related semantic fields. Thus, "*Zidane scored a goal.*" can be disambiguated thanks to common ideas concerning sport, while "*The lawyer pleads at the court.*" can be disambiguated thanks to those of justice. Furthermore, vectors allow to attach properly prepositions due to knowledge about vision. For example, the prepositional phrase "*with a telescope*" would be attached to the verb "*saw*" in the sentence "*He saw the girl with the telescope.*".

On the contrary, conceptual vectors cannot be used to disambiguate terms pertaining to different semantic fields. Actually, an analysis solely based on them might lead to misinterpretation. For example, the French noun ‹*avocat*› has two meanings. It is the equivalent of ‹*lawyer*› and the equivalent of the fruit‹*avocado*›. In the French sentence "*L'avocat a mangé un fruit.*", "*The lawyer has eaten a fruit*", ‹*to eat*› and ‹*fruit*› convey the idea of ‹*food*›, hence the interpretation computed by conceptual vectors for ‹*avocat*› will be ‹*avocado*›. It would have been good to realize that "*a lawyer is a human*" and "*a human eats*", yet this is not possible by using only conceptual vectors. They are simply not sufficient to exploit the instanciation of lexical functions in texts, however, a lexical network can help to overcome these shortcomings. These kind of limitations have been shown in experiments for the SA using ant algorithms in [12].

C. Lexical Information : Lexical Networks

1) Principles: Natural language processing has used lexical networks for more than fourty years, with Ross Quillian's work going back to the end of the sixtie's [**?**]. Authors differ concerning the network type and the way to use them. Some authors use directly graph microstructures (cliques, hubs) while others use them indirectly through similarity operations and/or activation of nodes (neural networks, pagerank).

The types of networks depends on entities chosen for nodes (lexical items, meanings, concepts) and on lexical relations chosen for edges. We can consider two families of lexical networks : (1) *semantic lexical networks* such as Quillan's [4], or, more recently, [20], WordNet [7] where nodes correspond to lexical items, concepts or meanings and, usually, there are several kind of edges to qualify a relation (synonymy, antonymy, hypernymy, ...); (2) *distributional lexical networks* such as [21] where two terms are linked with an edge provided they cooccur in a corpus. In this kind of network there is only one type of edge. For SA, lexical networks are used only for lexical disambiguation. On the other hand, Jean Véronis, for example, showed that distributional networks are *small worlds* and used this property to find every possible meaning for a word [21]. He made partitions on graphs to extract the different components organised around a hub, a central node to which are linked terms used in a same context. For a SA, these components are exploited while searching for the partition containing the words in the co-text of the target term. With regard to the indirect use of the structure of the graph, it is done step by step by mutual activations and excitation of the nodes to cause compatible solution to emerge. [20], for example, use a technique inspired by *"neural networks"* on a graph made from dictionaries definitions while [16] built a network with words of a sentence and their possible meanings and edges weighted according to a similarity between definitions. Excitation of nodes is done with a *pagerank* algorithm.

Very few authors use edge labels in their experiments. We have found only the *Leacock and Chodorow measure* [14] based on WordNet *is-a* relations.

2) Limits of Lexical Networks: All these methods help to solve only one of the problems mentionned, i.e. lexical ambiguity. They provide a way to make a preference concerning the meaning of each word of a text taken individually. This last feature makes it impossible to even obtain the compatible paths of interpretation. By their very nature, it is hard to imagine how to extend the above mentioned methods in order to solve at least one of the other problems. Indeed, they all consider that the important information to be found in the networks lie only in the node, whereas in reality they **also** lie in the edges. However, as mentionned in part II-D2, to find the relations between items in a statement can contribute to the resolution of other types of ambiguity (e.g. lexical ambiguity).

Of course, this last comment has to be considered with respect to the specifically used networks. In the previous examples, none present both paradigmatic and syntagmatic information as the network we manage to build. Nevertheless, some research converges towards this idea. Syntagmatic information is crucially lacking in a network like WordNet. This phenomenon is known as the *tennis problem*. The lexical item ‹*racket*› is in one area while ‹*court*› and ‹*player*› are in others. Of course this is true, no matter what field chosen. Syntagmatic and paradigmatic relations are essential for natural and flexible access to the words and their meaning. Michael Zock and Olivier Ferret have made a very interesting proposal in this respect [8].

D. Hybrid Representation of Meaning : Mixing Conceptual Vectors and Lexical Network

While lexical networks offer unquestionable precision, their recall is poor. It is difficult to represent all possible relations between all terms. Indeed, how can we represent the fact that two terms are in the same semantic field? They may be absent from the network because they are not connected by

"traditional" arcs. Introducing arcs of the type "semantic field" is also problematic for us because of two reasons implicated by the fuzzy and flexible of this relation: (1) the first one is related to the database creator's understanding on this relation: when are two synsets considered to be in the same semantic field? In an unfavourable case there would be very few arcs, while in the extreme opposite case we could have a combinative explosion in the number of arcs; (2) the second and more fundamental problem is related to the representation itself. How could a fuzzy relation, the essence of which is a continuous field, be represented with discrete elements?

Thus, the continuous domain offered by conceptual vectors gives flexibilities that the discrete domain offered by the networks cannot. They are able to bring closer words which share ideas, including less common ones. A network, on the other hand, cannot do so, however common the ideas are. The conceptual vectors and the operation of thematic distance can correct the weak recall inherent of the lexical networks. This, then, is why conceptual vectors and lexical networks are complementary tools to each other: the defects of one are mitigated by qualities of the other.

IV. Ant Algorithms and SA

It has been demonstrated that cooperation inside an ant colony is self-organisated. It results of simple interactions between individuals which allow the colony to solve complicated problems. This phenomenon is called swarm intelligence and is more and more used in computer science where centralised control systems are often successfully replaced by other types of control based on interactions between simple elements [6].

In these algorithms, the environment is usually represented by a graph. Virtual ants exploit pheromone deposited by others and pseudo-randomly explore the graph. Pheromone quantity plays the role of heuristic. These algorithms are a good alternative for the graph modelised problems resolution. They allow fast and efficient walkthrough close to other resolution methods. Their main interest is their important ability to adapt themselves to changing environment.

We think that phenomena to be addressed for a proper SA should be globally considered for at least two reasons. (1) They are dependent on each other. We exemplified it with Lexical Functions in II-D2 and this demonstration can be easily extended to other phenomena. (2) It is problematic to combine expertises with a supervisor. Criteria are often contradictory and their possible weighting are function of the others (again because they are related). Finally, the bottleneck is not only the expert agent conception but the precise definition of an aggregate function for the returned values. Ants algorithms constitute an easy and efficient way to handle SA issues in a hollistic manner. Each ant caste is associated to heuristics intended to solve a particular problem (in the presented model, to instanciate a LF type) and thus has its own behavior partly influenced by other castes. The idea is to constitute a beam of clues which causes one (or several) compatible solutions to emerge. Thus, when elements needed for an ambiguity resolution are present, solving one problem is able to help in the resolution of another. In this way, somewhat like domino theory, resolution is done progressively.

V. The MCSE Model

The Multi-Caste and Sharing Environment Model is not in the scope of this article we just present its characteristics. Mathematical heuristics which can be found in Didier Schwab's PhD. dissertation.

A. Principle

1) Bootstrapping: On the morpho-syntactic tree of the sentence (cf. figure 1(b)) we put (1) an ants nest for each ACCEPTION of the lexical item; (2) on each node a quantity of energy which corresponds to the reward of the ants. At each discovery of a lexical network node by an ant, we will also place there the same quantity of energy; (3) a conceptual vector with all its coordinate are equal, an odor, on each node of the tree. At step cycle of the experiment, we consider this odor as a representation of the thema of the subgraph.

2) Simulation: Simulation consists in a potentially infinite iteration of cycles. The simulation can be stopped and the current state observed. During a cycle, the following tasks are done: (1) eliminate the oldest ants (a number of fixed cycles); (2) for each nest, request the production of an ant (an ant can be or not be born, in a probabilistic way); (3) For each edge, decrease the rate of pheromone (evaporation of the traces); (4) for each ant:determine its mode (search for food, return to the nest) and make it move and create an interpretative bridge if necessary; (5) compute move consequences of the ants (on the activation of the edges and the energy of the nodes);

In an abstract way, one can summarize the move of an ant as follows. A newborn ant (ie. just produced by its nest) looks for food. It is attracted by the nodes which carry much energy (food). It collects as much energy as it can carry, it transports more and more food and the probability of wishing to return to the mother nest increases. When it wants to return, it moves following (statistically) the ways which contain the mother nest pheromone and then try goes back to deposit its loading there. When an ant is on a nest, it has the choice between an edge of the morpho-syntactic tree or an edge of the lexical network. If it chooses to follow a network edge, we recopy the arrival node and we place there a quantity of energy equal to the one which was placed on each node of the morpho-syntactic tree during the bootstrapping phase. Just like a nest, this node corresponds to an ACCEPTION, thus its vector cannot be modified. The ant will then seek to explore other nodes gradually and will be likely to build a bridge toward its nest mother if it finds a node thematically close.

3) Creation, Suppression and Bridge Type: As soon as an ant is on a node corresponding to a ACCEPTION , (i.e. a nest or a recopied node) of the lexical network, it can build a bridge. A bridge can be created when an ant reaches a potentially friendly nest. In this case, the ant evaluates the node corresponding to its mother nest like the nodes structurally related to this nest. If this node is selected, there is creation of a bridge between the two nests. This bridge is then considered as standard by the ants, i.e. the nodes which

it links are regarded as neighbors. A bridge can be seen like a compatibility between two nests, a possible interpretative way. This bridge is covered at the same time by the pheromone of passage deposited by each ant which borrows it and by the pheromone specific to each class. If all the pheromone of the bridge evaporates, the bridge is removed. Indeed, not only ants make it possible to know the various possible interpretative ways but they also make it sometimes possible to qualify these ways. Thus, if a bridge between two nests is often borrowed by ants of a caste ***c*** , one will be able to deduct from it a certain number of more or less direct information according to the caste. For example, a bridge strongly borrowed by ants of the caste ***seek_magn*** will probably represent this relation. It will be the same with the ants ***seek_predicat*** or the ants ***seek_patient*** . On the other hand, some are less easily interpretable like the synonymy or the hyperonymy which can contribute to discovered relations of identity if the nests join the same terms morphology.

4) Energy: At the beginning of simulation, the system has a certain energy which is distributed equitably on each node. The nests use energy they have to manufacture ants. These last move in the environment and bring back energy to the nests which will use it to produce other ants. When an ant dies on a node, the energy which was carrying and the energy which was necessary to produce ant is deposed on the node. There is thus neither a loss nor contribution of energy at any time. The system is completely closed. The quantity of energy is a fundamental element of the convergence of the system toward a solution. Indeed, since total energy is limited, the nests are in competition and only alliances may permit emergence of solutions. If we didn't choose to limit energy, all nests would receive energy and all would be strongly activated and none would be inhibited.

B. Example of Semantic Analysis in the MCSE Model

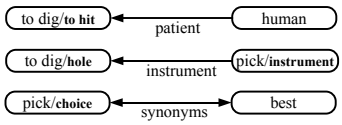
Let us take an ultra-simplified example to understand how is held an analysis in hybrid model. Let us consider the sentence "*He digs with the pick.* " and the mini lexical network presented in figure 1(a). The most important thing to understand here is the overall dynamics of the system. From some relatively simple heuristics presented in the preceding section, we have, by simple emergence, at the resolution of the various problems of analysis raised by the text. In our example, the only difficulties are at the level of lexical ambiguity: is ‹*pick*› the **instrument** or the **choice** and does ‹*to dig*› mean "*to hit*" or "*to make a hole*"? It is thus probable to understand how the bridges (E), (F) and (G) of the figure 1(b) will be formed and how the system chooses this interpretation rather than the others, can contribute to the comprehension of this dynamic. In this simple example, the *to dig*/**to hit** and *pick*/**choice** nests cannot be reasonably combined in order to emerge an interpretative way. Indeed, the lexical network given does not connect them and the topics given by each one are relatively distant. This fact has a significant consequence on ant moves on the morpho-syntactic tree. In this environment, it can only be chaotic at the beginning of the experiment and only influenced by the network. Let us consider each nest and the behavior of the ants which of it result.

Ants from *to dig*/to hit (2) and *pick*/choice (4) nests explore the lexical network or the tree and get lost since they can find nothing sufficiently tangible to come back toward their mother nest. Thus, they often die, seldom build bridges which, if they happen to exist, are seldom crossed and quickly disappear.

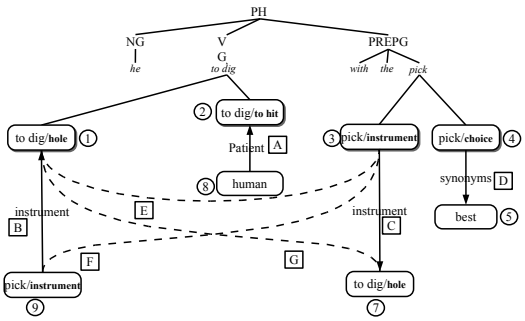
Ants from *pick*/instrument (3), in particular those of the caste ***look_for_instrument*** cross the edge (C) to arrive on the ACCEPTION *to dig*/**hole** (7). The ants which are in the morpho-syntactic tree go down again toward the leaf and specifically reach the nest *to dig*/**hole** (1). Statistically, a stable bridge (E) cannot be directly considered now because the arrival of the ants is not very probable since it is only possible from the tree. These ants thus start to go in great majority on the lexical network by the edge (B), on nodes most probably already copied by the ants from the nest *to dig*/**hole** (1). Arrived on *pick*/**instrument** (9), they probably create a bridge toward their mother nest (3) since the odor criterion will be then maximum.

Ants from *to dig*/hole (1) act in a symmetrical way to those of *pick*/**instrument** (3). Some ants choose to cross edge (B) to *pick*/**instrument** (9). In parallel, those which choose to go in the tree go down again towards the leafs and in particular towards the nest *pick*/**instrument** (3). Statistically, at this time, the bridge (E) can be created but its conservation is not very probable considering the relatively weak flow of ants of (1) newcomer in (3). The majority of these ants then will explore the lexical network by the edge (C) towards *to dig*/**hole** (7). Arrived at this node, they have a rather strong probability to create a bridge (G) towards their mother nest (1).

The most significant point in this example relates to the **collaborative behavior of *pick*/instrument and *to dig*/hole**. The ants of (1) created the bridge (G) and ants of *pick*/**instrument** (3) can thus cross it and find themselves on the nest (1). From there, they can manufacture a bridge (E) which this time will have statistically more chance to be preserved since it is compatible with available information of circuit CEG. In the same way, this bridge will be reinforced by the ants of *dig*/**hole** (1) which, they, will use EFB.



(a) Mini lexical network for example



(b) Environnement modifications by ants

acceptions	global	nouns	adjectives	verbs	adverbs
recall	0,79 (+4%)	0,79 (+5%)	0,79 (+3%)	0,74 (+6%)	0,82 (+0%)
precision	0,78 (+3%)	0,78 (+2%)	0,82 (+4%)	0,76 (+4%)	0,8 (+2%)

edges	global	interpretation	prep. attachments	references	FL instantiation
recall	0,82 (+80%)	0,81 (+14%)	0,83 (+2%)	0,78 (+50%)	0,83 (+∞%)
precision	0,85 (+77%)	0,82 (+12%)	0,89 (+6%)	0,81(+63%)	0,87 (+∞%)

FL	syn	hyper	Magn	Ver	Bon	mero	holo	instr	agent	patient	dest
recall	0,85	0,77	0,7	0,72	0,73	0,62	0,7	0,72	0,72	0,6	0,61
precision	0,86	0,78	0,72	0,74	0,75	0,66	0,69	0,7	0,74	0,62	0,62

VI. Experiment and Results

A forty short texts corpus was constituted. These texts were selected for their representativeness of the semantic phenomena which we seek to solve (cf. I). In this corpus, each sentence was manually annotated to describe, ideally, its complete SA. In practice, for each sentence, one describes each possible interpretation i.e. (1)ACCEPTION *used for each word* , (2) *references* (3) *prepositional attachments* and (4) *lexical functions instantiations*. The evaluation then consists in comparing nests and edges created by ants. At the end of five minutes of analysis of each text, computation is stopped (in all our tests, we did not find convergence exceeding two minutes). Only the nests whose activation level is higher than 0 are preserved. In other words, the inhibited meanings are ignored as well as the possible edges they would be linked to (what is very little probable). Usually, one compares results according to the traditional method of recall-precision. The experiment presented was undertaken on 11 FL and 22 castes. The table presents hybrid model results, percentages show rates augmentation comparing to a pure conceptual vectors model. First, we can notice that all semantic phenomenon are solved and, thus, validate the model. We also see that LF usage improves results. As an example, references are the best results as precision rate goes to 63%. Results of disambiguation also show an indirect qualitative profit of the instantiation of the FL on interpretation edges and terms disambiguation. The significant instantiation of the adjectival FLs explains in particular the good rate for adjectives and nouns. The same phenomenon is found for the verbs although the rate of instantiation of the verbal FL is less except for the *agent* relation.

VII. Conclusion and Further Work

In this paper, we presented some of the ambiguity problems that a semantic analysis should solve. The resolution of these linguistic problems can take advantage of both thematic information (conceptual vectors) and lexical information (lexical network) through instanciation of lexical functions. Handling these tasks can be difficult when they are considered separately and often contradictory when they are globally considered. Instead, they can be process by emergence through ant algorithms. We have presented a model with multi-caste (each looking for a particular lexical function) ant colony and shared environment(morpho-syntactic tree and lexical network). This is just a first study which paves the way toward many extensions. Some problems still remain such as passive form (as in "*L'avocat a été mangé*" (cf. I) where "avocat" is the lawyer with the presented model) or the system stop.

Indeed, we are strongly convinced that our model, at least in its principle if not in its implementation, carries many interesting tracks of research. In particular, the genericity of the approach makes it possible to easily define new ant castes corresponding to new heuristics.

References

[1] K. R. Bangha. *La place des connaissances lexicales face aux connaissances du monde dans le processus d'interprétation des énoncés* . PhD thesis, Université de Montréal, Canada, 2003.

[2] Romaric Besançon. *Intégration de connaissances syntaxiques et sémantiques dans les représentations vectorielles de texte* . Thèse de doctorat, École Polytechnique Fédérale de Lausanne, 2001.

[3] Jacques Chauché. Détermination sémantique en analyse structurelle : une expérience basée sur une définition de distance . *TAL Information*, pages 17–24, 1990.

[4] Alan Collins and Ross Quillian. Retrivial time from semantic memory. *Verbal learning and verbal behaviour*, pages 240–247, 1969.

[5] Scott C. Deerwester, Susan T. Dumais, Thomas K. Landauer, George W. Furnas, and Richard A. Harshman. Indexing by Latent Semantic Analysis . *Journal of the American Society of Information Science*, pages 391–407, 1990.

[6] Dorigo and Stützle. *Ant Colony Optimization*. MIT-Press, 2004.

[7] Christiane Fellbaum, editor *WordNet: An Electronic Lexical Database*. MIT Press, 1988.

[8] Olivier Ferret and Michael Zock. Enhancing Electronic Dictionaries with an Index Based on Associations . *COLING'2006*, pages 281–288, Sydney, Australia, July 2006.

[9] Núria Gala Pavia. Une méthode non supervisée d'apprentissage sur le Web pour la résolution d'ambiguïtés structurelles liées au rattachement prépositionnel. *TALN'2003*, pages 353–358, France, juin 2003.

[10] J.Haiman.Dictionaries and encyclopedias.Lingua,pages 329–357, 1980.

[11] Nancy Ide and Jean Veronis. Word sense disambiguation: the state of the art . *Computational Linguistics*, pages 1–41, 1998.

[12] Mathieu Lafourcade and Frédéric Guinand. Ants for Natural Language Processing . *IJCI Research*, 2006. to appear.

[13] Larousse(ed.)*Thésaurus Larousse*, Larousse, 1992.

[14] C. Leacock and M. Chodorow. *WordNet: An electronic lexical database* , Combining local context and WordNet similarity for word sense identification. MIT Press, 1998.

[15] Igor Mel'čuk. *Lexical Functions in Lexicography and Natural Language Processing* , Lexical Functions: A Tool for the Description of Lexical Relations in the Lexicon, pages 37–102, 1996.

[16] Rada Mihalcea, Paul Tarau, and Elizabeth Figa. PageRank on Semantic Networks, with Application toWord Sense Disambiguation . *COLING'2004*, pages 1126–1132, Geneva, Switzerland, Août 2004.

[17] Peter Norvig. Multiple simultaneous interpretation of ambiguous sentences. *10th conference of the cognitive science society*, Août 1988.

[18] Alain Polguère. *Lexicologie et sémantique lexicale*. Les Presses de l'Université de Montréal, 2003.

[19] Gerard Salton and Michael McGill. *Introduction to Modern Information Retrieval*. McGrawHill, New York, 1983.

[20] Jean Véronis and Nancy Ide. Word Sense Disambiguation with Very Large Neural Networks. Extracted from Machine Readable Dictionaries. *COLING'1990*, volume 2, pages 389–394, 1990.

[21] Jean Véronis. Hyperlex : lexical cartography for information retrieval. *Computer, Speech and Language*, pages 223–252, 2004.

[22] Anna Wierzbicka. *Semantics: Primes and Universals*. Oxford University Press, 1996.