ANT ALGORITHMS, CONCEPTUAL VECTORS AND FUZZY UNL GRAPHS

Mathieu Lafourcade[1]

***Abstract** – In the context on the UNL project, we focus on the automatization of enconversion process, that is the building of UNL graphs from sentences. We present an extension of the UNL graph structure aiming at handling lexical and relational ambiguities. On this intermediate structure, we can apply ant algorithm propagation of conceptual vectors and other constraints. Graph nodes and relations have a level of excitement and when this level remains too low for too long they are deleted. This way, both acception and attachment selections can be performed.*

***Index Terms** – Fuzzy UNL graphs, enconversion, ant algorithms, conceptual vectors, lexical and PP attachment ambiguities.*

INTRODUCTION

In itself, a text constitutes a complex system, but the computational problem is that the meanings are not strictly speaking active elements. In order to ensure the dynamicity of such a system, an active framework made of "meaning transporters" must be supplied to the text. These "transporters" are intended to allow the interactions between text elements and they have to be both light (because of their possible large number) and independent (word meanings are intrinsic values). Moreover, when some meanings stemmed from different words are compatible (*engaged* with *job* for instance), the system has to keep a trace of this fact. These considerations led us to consider ant algorithms. Ant algorithms or variants of them have been classically used for optimisation problems like traveling salesman problem [*Dorigo* et al. 1997] among many others, but they were never used in Natural Language Processing (most probably because the NLP community contrary to the psycholinguistics one, considered semantic aspects not very often as an optimization problem, nor explicitely modeled then as a dynamic complex system, [*Kawamoto* 1993] being a notable exception). However, [*Hofstadter* 1995] with the COPYCAT project, presented an approach where the environment by itself contributed to solution computation and is modified by an agent population where roles and motivations vary. Some properties of these models seem to be adequate for the task of semantic analysis, where word senses can be seen as more or less cooperating. We retain here some aspects that we consider as being crucial: (1) mutual information or semantic proximity is one key factor for lexical activation, (2) the syntactic structure of the text can be used to guide information propagation through possibly ambiguous relations. Finally, as pointed by [*Hofstadter* 1995], biased randomization (which doesn't mean chaos) plays a major role in this kind of model.

In the context on the UNL project, we focus on the automatization of enconversion process, that is the building of UNL graphs from sentences. We present an extension of the UNL graph structure, dubbed *fuzzy UNL graph*, aiming at handling lexical and relational ambiguities. On this intermediate structure, we can apply ant algorithm for propagating conceptual vectors and other constraints. Graph nodes and relations have a level of excitement and are deleted when this level remains too low for too long. This way, both acception and attachment selections can be performed. We construct fuzzy graphs on the basic of morpho-syntactic analysis trees which enumerate PP (prepositional phrase) attachments or are duplicated depending on the nature of syntactical ambiguities. Lexical ambiguities are represented as alternative nodes at leaf level.

The conceptual vector model is a *recall focused* approach which aims at representing thematic activations for chunks of text, lexical entries, locutions, up to whole documents. Roughly speaking, vectors are supposed to encode *ideas* associated to words or expressions. The main applications of the model are thematic text analysis and lexical disambiguation [*Lafourcade* 2001] and can find interesting approaches for vector refinement through the lexical implementation of taxonomies like the UNL knowledge base. Practically, we have built a system, with automated learning capabilities, based on conceptual vectors and exploiting monolingual dictionaries for iteratively building and refining them. For French, the system learned so far 130000 lexical entries corresponding to roughly 470000 vectors (the average meaning number being 5). We are conducting the same experiment for English.

In this paper, we first expose the main principles and assumptions about the treatment of ambiguities during the enconversion. Then, we present the conceptual vectors model and the fuzzy graph extension. The conceptual vector propagation through ant algorithm is then detailed with its consequences on weighting acception and relations. Some examples of fuzzy graphs are given, focusing mainly on simple acception selection and choice between *mod* (modifier) and *ins* (instrument) relations.

[1]LIRMM, Université Montpellier II, 161, rue Ada, 34392 Montpellier cedex 5, Mathieu.lafourcade@lirmm.fr, http://www.lirmm.fr/~lafourcade

HOLISTIC ALGORITHMS FOR DISAMBIGUATION

Thematic representation and mutual information sharing The constraints present in the UNL knowledge base is instrumental for an automated enconversion process but is by far too scarce to be considered as a thematic (or semantic) representation. We use conceptual vectors to convey a rich meaning representation both for acception and for each entry of the knowledge base.

Analysis viewed as a Non-Ending Iterated Process Very often, the semantic analysis is viewed as processing sequentially more or less like an expert system. In our views, this process should be done incrementally by adding little pieces of informations (dubbed as *clues*) at a time, and letting some induction process structuring the result. The process may converge (it is the case most of the time), but for very ambiguous results some oscillations could occur. Furthermore, all kinds of semantic ambiguities are holistically processed, that is at the same time, with all representation clues being solicited.

Explicitly Managing Uncertainty More than often, uncertainty about domain or about data interpretation are considered as problems to be absolutely solved, and in case of irreconcilable constrains, some heuristics are called or experts questionned. We think that uncertainty should be explicitly represented and managed, as it can never be completely eliminated. This is why, we advocate that each relation in the graph to be associated with a *confidence value* or (depending on the view adopted) *excitement level*. This value may be increased (or lowered) according to the clues discovered or the induction undertaken. Distributional aspects of free texts are an excellent source for managing uncertainty on the basis of existing items and relations found in dictionaries.

Mixing Meanings and Vocables Lexical and syntactical ambiguities are the issues at stake. More than often in texts, word senses may not be clearly separated. Morevover, it is now well accepted in psycho-linguistics that language is processed at the same time at vocable (terms, compounds, etc.) and meaning (thematically and associatively) levels.

CONCEPTUAL VECTORS AND FUZZY UNL GRAPHS

Conceptual Vectors

The Model of Conceptual Vector has already been presented the context of UNL in [*Lafourcade* et al. 2002] and what follows is a short description (towards the unfamiliar reader) of the main principles. Thematic aspects (or ideas) of textual segments (documents, paragraphs, syntagms, etc.) are represented thanks to vectors of interdependent concepts. Lexicalized vectors have been used in information retrieval for long [*Salton* et al. 1983] and for meaning representation by the LSI (Latent Semantic Indexing) model from latent semantic analysis (LSA) studies in psycholinguistics [*Deerwester* et al. 90]. In computational linguistics, [*Chauché* 90] proposed a formalism for the projection of the linguistic notion of semantic field in a vectorial space, from which our model is inspired [*Lafourcade* 2001]. From a set of elementary notions, dubbed as *concepts*, it is possible to build vectors (conceptual vectors) and to associate them to lexical items. The hypothesis that considers a set of concepts as a generator to language has been long described in [*Roget* 1852] (*thesaurus hypothesis*). Polysemous words combine the different vectors corresponding to the different meanings considering several criteria as weights: semantic context, usage frequency, language level, etc. Concepts are defined from a thesaurus (in our prototype applied to French, we have chosen [*Larousse* 1992] where 873 concepts are identified to compare with the thousand defined in [*Roget* 1852]). To be consistent with the thesaurus hypothesis, we consider that this set constitutes a generator space for the words and their meanings. This space is probably not free (no proper vectorial base) and as such, any word would project its meaning(s) on this space.

Thematic Projection Principle and Angular Distance. Let be $\mathcal{C}$ a finite set of n concepts, a conceptual vector V is a linear combination 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 $\mathcal{C}$.

Let us define $Sim(A, B)$ as one of the *similarity* measures between two vectors A et B, often used in information retrieval as their normed scalar product. We suppose here that vector components are positive or null. We, then, define an *angular distance* D_A between two vectors A and B as their angle.

$$\begin{gathered} Sim(A, B) = \cos(\widehat{A, B}) = \frac{A \cdot B}{\|A\| \times \|B\|} \\ D_A(A, B) = \arccos(Sim(A, B)) \end{gathered} \tag{1}$$

Intuitively, this function constitutes an evaluation of the *thematic proximity* and is the measure of the angle between the two vectors. We would generally consider that, for a distance $D_A(A, B) \leq \frac{\pi}{4}$, (i.e. less than 45 degrees) 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 can have, for example, the following angles (values are in degrees; examples are extracted from `http://www.lirmm.fr/~lafourcade`):

D_A(‹*tit*›, ‹*tit*›) = 0°
D_A(‹*tit*›, ‹*animal*›) = 32°
D_A(‹*tit*›, ‹*passerine*›) = 10°
D_A(‹*tit*›, ‹*joy*›) = 42°
D_A(‹*tit*›, ‹*bird*›) = 19°
D_A(‹*tit*›, ‹*sadness*›) = 65°

A ‹*tit*› is thematically closer to a ‹*passerine*› than a ‹*bird*› than an ‹*animal*›. Here the thematic proximity follows some kind of ontologic relation. However, ‹*cell*› nonewithstanding the polysemy begins to be poorly related. The term ‹*sadness*› has almost no thematic sharing with ‹*tit*›.

Meaning Selection. From a given thematic context under the form of a conceptual vector, it is possible to select (or weight) the meanings of a vocable. For a vocable w with k meanings $w_1 \ldots w_k$ and a context C, the weights α of the meanings are non-linearly related to the amount of mutual information between the context and a given meaning:

$$\begin{aligned}\alpha_i &= \cot(D_A(V(w_i), C) \\ &= \frac{\cos(D_A(V(w_i), C)}{\sin(D_A(V(w_i), C)}\end{aligned} \quad (2)$$

We recall that *cot* refers to the *cotangent* function, with $\cot(0) = +\infty$ and $\cot(\pi/2) = 0$. The rational is the following. The *similarity* between two objects A and B is the cosine of the angle between these two objects. Inversely the *dissimilarity* is the sine. The weight of selection of B towards A if the ratio between what is common (the similarity) on what is different (the dissimilarity).

For example, take the vocable ‹*frégate*› (Eng. frigate) with ambiguity between the boat and the bird. Let C be the vector related to ‹*plume*› (feather) which is itself ambiguous, we have the following values:

$D_A(V($‹*frégate(icl>boat)*›$), V($‹*plume*›$)) = 1.1$
$\alpha_i = cot(1.1) = 0.5$
$D_A(V($‹*frégate(icl>bird)*›$), V($‹*plume*›$)) = 0.5$
$\alpha_i = cot(0.5) = 2.18$

Thanks to the thematic context, the most activated meaning of ‹*frégate*› in the context of ‹*plume*› is the bird, as it has much more weight than the other interpretation. Although useful, this process may no be sufficient as more than often words and meanings are related while not being in the same semantic field. This is why, the construction and the exploitation of lexical and semantic network is necessary. The construction of such a network is done through templates but also by filtering through thematic proximity.

Fuzzy UNL Graphs

We extend UNL graph by adding to new types of nodes: lexical and relation nodes. These nodes are only instrumental in the process of choosing which acceptions or relations have to be selected (see Fig. 1 and Fig. 2) . To link these nodes to standard nodes we use two new types of arc: *acc* for linking acceptions to their corresponding lexical node and, *rel* for linking relation nodes to lexical nodes.

ANT ALGORITHM ON FUZZY UNL GRAPHS

Each acception node behaves like an *ant nest* producing ants that propagate on the graph the conceptual vector associated to the acception. However, at each cycle of the simulation, the probability for a nest to create an ant is a function of its activation level $\mathrm{E}(N) \in [-\infty, +\infty]$. There is a cost ϵ (we set ϵ empirically to 0.1) for producing an ant, which is deducted from the nest energy. Each time, a nest produces an ant, its probability to generate another one at the next cycle is lowered. The probability of producing an ant, is related to a sigmoid function (see Figure 3) applied to the energy level of the nest. The definition of this function ensures that a nest has always the possibility to produce a new ant although the odds are low when the node is inhibited (energy below zero). A nest can still borrow energy and thus a word meaning has still a chance to express itself even if the environment is very unfriendly. For a given lexical node, at each cycle at least one ant should be produced among the various acceptions.

Nests should count on ants of other nests to improve their energy level. In effect, in their wandering other ants may arrive at a given acception node (not their own) and give an amount of energy δ equal to :

$$\delta = \mathrm{DS}_A(N, A) = 1 - \frac{2D_A(V(A), V(N))}{\pi} \quad (3)$$

where $V(A)$ is the vector of the ant A and $V(N)$ the vector of the node N (N should not be the nest of A). W call this value DS_A (as Distance Similarity) as it is the distance D_A mapped from $[0, \frac{\pi}{2}]$ to $[1, 0]$. We see here that if A bears a vector that resembles very much the node it encounters, then a large amount of energy will be given. To induce some population control, each ant has a life span L of a finite number of cycles after which it dies (we found experimentally that $L = 30$ is a good trade-off between convergence of the simulation and resources).

Each time an ant traverses an arc, it increase the excitement level of this arc (this is metaphorically a small amount of pheromone that give its name to ant algorithms). This excitement slowly decays over time, and if this arc is not visited for a long time it may reach a null excitement and be deleted. Only *rel* and *acc* links can be deleted. At the beginning of the simulation, each arc excitement is equal to 1. Each time an ant enters a node that is not an acception, it modifies slightly the node vector:

$$V'(N) = V(N) \oplus \alpha V(A) \qquad \text{with} \quad \alpha = 0.01$$

This way, each ant propagates the vector on the graph. The ant displacement behavior is directly related to node vectors.

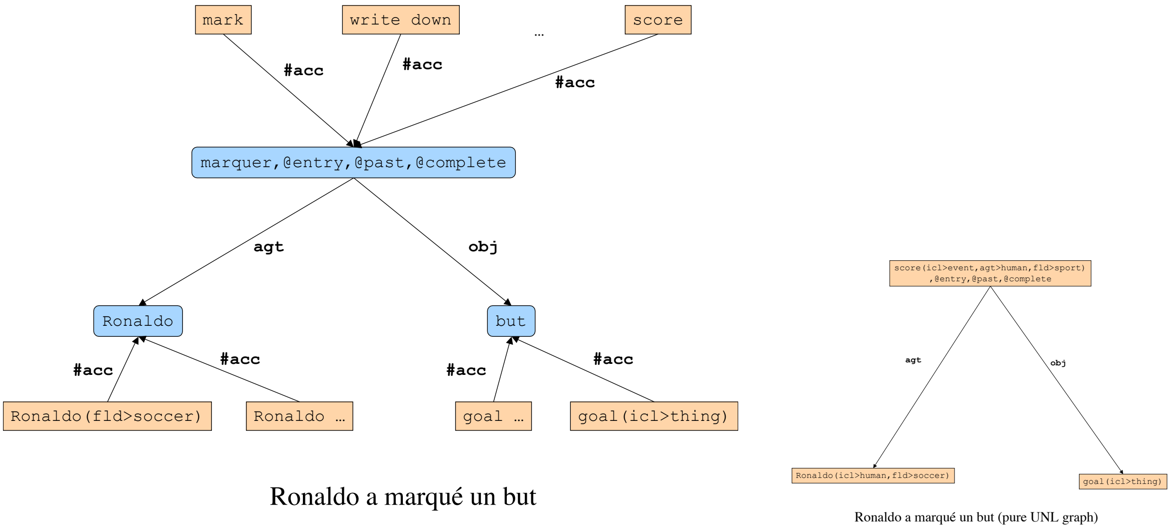


FIGURE. 1. Example of the French sentence *Ronaldo a marqué un but.* (Lit. Eng. Ronaldo scored a goal). One the right, possible UNL graph. On the left, the fuzzy graph where each content word is represented through one lexical node which is linked to each corresponding acception. An example with *rel* relation is given with Figure 2.

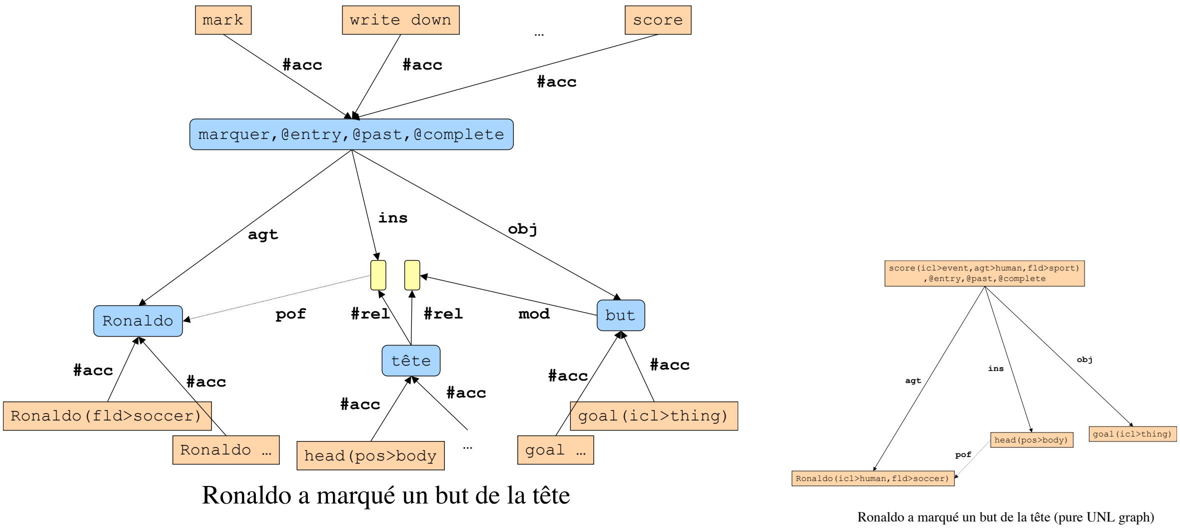


FIGURE. 2. Relation nodes are used (for example) when attachment are ambiguous. In the sentence *Ronaldo a marqué un but de le tête.* (Lit. Eng. Ronaldo scored a goal of the head), the GN *de la tête* may be a *mod* of *goal* or an *inst* of *score* (proper interpretation).

Before moving, an ant examines each nodes linking to its current position. The probability $P(N_k, A)$ for an ant A to choose a particular node N_k is computed as follows:

$$P(N_x, A) = \mathrm{DS}_A(N_k, A) / \sum_{1 \le i \le p} \mathrm{DS}_A(N_i, A) \tag{4}$$

At the beginning, only acception nodes have a conceptual vector. A node without vector is considered having the null vector. Over time, non acception nodes have vectors that correspond to the ant population distribution passing by them. From an acception, its vector slowly propagates outward, and ants may eventually find some *friendly* nests. The algorithm is purely altruistic as a nest will receive energy only by stranger ants. To be successful, which means being able to maintain a high level of energy and a large ant population, a nest should find some support in other nests.

After some cycles (round 300 for the examples given in this paper), the activation and vectors of the graphe have converged. That is they are not much modified by ant activity. A cleaning stage is then performed to obtain a standard UNL

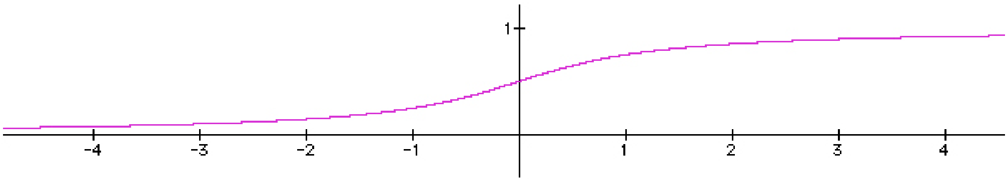


FIGURE. 3. Sigmoid function: $\mathrm{Sig}(x) = \frac{1}{\pi}\arctan(x) + 0.5$. Some values are: $\mathrm{Sig}(0) = 1/2, \mathrm{Sig}(1) = 0.75, \mathrm{Sig}(2) = 0.852, \mathrm{Sig}(-1) = 0.25, \mathrm{Sig}(-2) = 0.147$.

graph. On remaining *acc* and *rel* links related to a lexical node, only the most activated one is kept, others are deleted. Then, inaccessible nodes are suppressed. Finally, each lexical node are replaced by the unique acception left.

Example with only lexical nodes

In the sentence presented in Fig. 1 we have only a lexical ambiguity with *marquer*, *but* and possibly *Ronaldo*. Each acception, are producing ants that are slowly spreading their conceptual vectors. Notice that each produced ant decrease the energy level of its nest, thus the ant production, after an initial burst, tends to rapidly decrease. However, if we focused on the node *score*, even early in the simulation, most of the ants attaining it come from acception sharing much information (namely *goal(fld>soccer)* and *Ronaldo(icl>human,fld>soccer)*). Other acceptions are not able to maintain their population and the graph is swarmed by ant from activated acceptions. Figure 4 illustrates an intermediary steps where everything seems to be already settled.

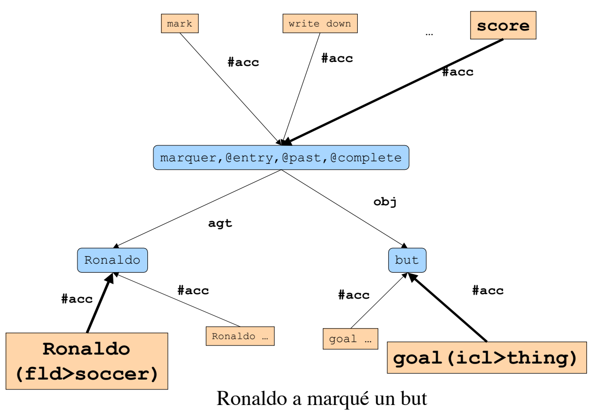


FIGURE. 4. By mutual information sharing with conceptual vectors, the ant circulation quickly converges between some selected acceptions. After some time, poorly activated nodes are not able to maintain any population level and related links disappears.

Example with lexical and relation nodes

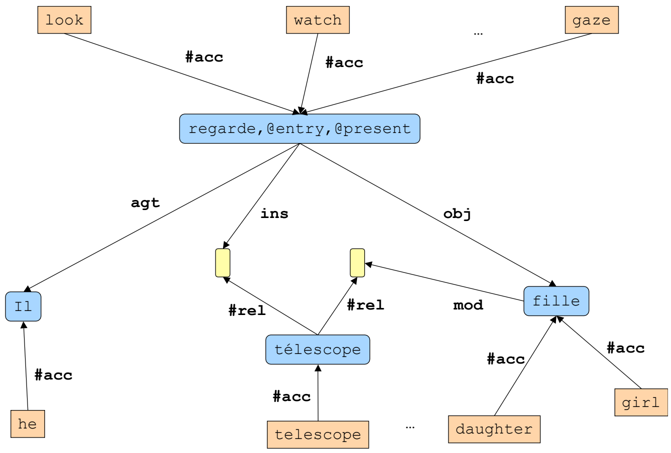
In the French sentence *Il regarde la fille avec un telescope*, we focuse our attention on relations and attachments (Fig 5). The acceptions *watch* and *telescope* support mutually more than any others. Furthermore, the whole path between both acceptions is shorter through the *ins* relation which induces less information dissipation. Eventually, the *rel* link related to the *mod* relation disappears. We should note here, that for *fille* the thematic context doesn't help, other insformation like acception distribution should be used.

In the French sentence *Ronaldo a marqué un but de la tête*, we have the situation of Figure 4 plus an attachment and relation difficulty similar to Figure 5. The lexical desambigation is reinforced with *tête* as an instrument of *score* and a part-of *Ronaldo*. Furthemore, the sharing between acceptions of *tête* and *but* is too low to compete and maintain.

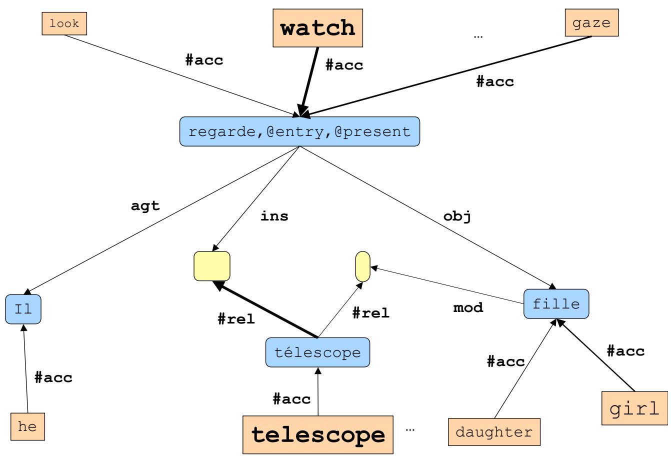
CONCLUSION

This paper has presented an approach extending UNL graph by including lexical and relation nodes and links, such a way to accommodate word senses and attachment ambiguities. This *fuzzy* UNL graph is created by some transformation on a morpho-syntactic tree. On this structure, we do propagate constraints to performs a disambiguation task. The propagation is directly inspired from *ant algorithm* and is formally identical to the Traveling Salesman Problem. The information exploited for the ant propagation are the topology of the graph and the mutual information between the conceptual vectors used for meaning representation.

We have defined some underlying principles to our approach. First, it is interesting to combine rich thematic representation like conceptual vectors and symbolic constraints as found in the UNL knowledge base. Then, uncertainty should be tackled explicitly and globally both under lexical and relation aspects. If we consider how vocables and knowledge are processed psycholinguistically, we have definitive advantages to mix vocable nodes and meaning nodes. This last aspect is very instrumental for the selection process.



Il regarde la fille avec un télescope



Il regarde la fille avec un télescope

FIGURE. 5. BECAUSE OF THE MUTUAL SUPPORT BETWEEN *watch* AND *telescope* THE *ins* RELATION EMERGES COMPARED TO THE *mod* RELATION.

Our strategies have been prototyped and tested on various French sentences and shorts texts. The obtained UNL graphs are very satisfactory, and all in all the approach seems very promising. For texts, sentense graphs were sequentially linked to each other by an abstract *text* node. It is also used for comforting conceptual vector calculation and detecting inconsistencies either in thematic association or in relations between vocables. Nevertheless, in some quite difficult cases, the graphs activation does not converge but oscillates between states. This is especially true of humorous sentence with double entendre. A desirable extension of our model is to enrich the representation with other types of constraints like lexical preferences, statistical co-occurences, to name a few.

REFERENCES

[*Chauché* 90] Jacques Chauché., *Détermination sémantique en analyse structurelle : une expérience basée sur une définition de distance.* TAL Information, 31/1, pp 17-24, 1990.

[*Deerwester* et al. 90] Deerwester S. et S. Dumais, T. Landauer, G. Furnas, R. Harshman., *Indexing by latent semantic anlysis.* In Journal of the American Society of Information science, 1990, 416(6), pp 391-407.

[*Dorigo* et al. 1997] Dorigo M., and L. Gambardella., *Ant colony system : A cooperative learning approach to the travelling saleman problem.*, *IEEE Transactions on Evolutionary Computation*, 1(1), p. 114-128, 1997.

[*Hofstadter* 1995] Hofstadter, D. R., *Fluid Concepts and Creative Analogies: Computer Models of the Fundamental Mechanisms of Thought (together with the Fluid Analogies Research Group)*, NY: Basic Books, 1995.

[*Lafourcade* 2001] Lafourcade M., *Lexical sorting and lexical transfer by conceptual vectors.* Proc. of the First International Workshop on MultiMedia Annotation (Tokyo, Janvier 2001), 6 p.

[*Lafourcade* et al. 2002] Lafourcade M. , and Ch. Boitet., *UNL lexical*

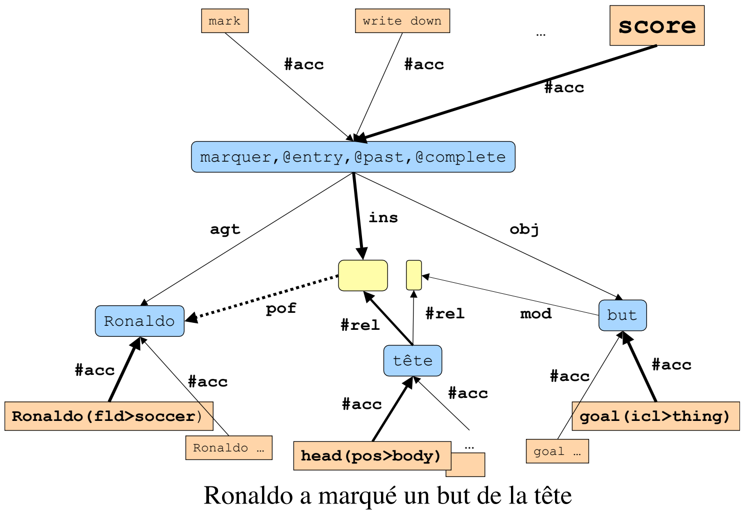


FIGURE. 6. Lexical selection induces relation selection (of *ins* opposed to *mod* in this example), which in turn reinforces acception activation.

Selection with Conceptual Vectors Proc. of LREC'2002, Las Palmas, Canary Island, Spain, May 27, 2002.

[*Larousse* 1992] Larousse., *Thésaurus Larousse - des idées aux mots, des mots aux idées.* Larousse, ISBN 2-03-320-148-1, 1992.

[*Lowe* 2000] Lowe, W., *Topographic Maps of Semantic Space*, *PhD Thesis*, institute for Adaptive and Neural Computation, Division of Informatics, Edinburgh University, 2000.

[*Kawamoto* 1993] Kawamoto, A. H., *Nonlinear dynamics in the resolution of lexical ambiguity: A parallel distributed processing account.* Journal of Memory and Language, 32, 474-516.

[*Miller* 1991] G.A Miller and C. Fellbaum., *Semantic Networks in English.* in Beth Levin and Steven Pinker (eds.) *Lexical and Conceptual Semantics* , 197–229, Elsevier, Amsterdam, 1991.

[*Pasteels* et al. 1987] J.M. Pasteels, J.L. Deneubourg, S. Goss, D. Fresneau, and J.P. Lachaud., Self-organization mechanisms in ant societies (ii): learning in foraging and division of labour. *Experientia Supplementa,* 54:177–196, 1987.

[*Prince* et al. 2003] Prince V. and Lafourcade M., *Mixing Semantic Networks and Conceptual Vectors: the Case of Hyperonymy.* Proc. of ICCI-2003 (2nd IEEE International Conference on Cognitive Informatics), South Bank University, London, UK, August 18 - 20, 2003.

[*Roget* 1852] Roget P., *Thesaurus of English Words and Phrases.* Longman, London, 1852.

[*Salton* et al. 1983] Salton G. and M.J. MacGill., *Introduction to modern Information Retrieval* McGraw-Hill, New-York, 1983.

[*Yarowsky* 1992] Yarowsky D., *Word-Sense Disambiguation Using Statistical Models of Roget's Categories Trained on Large Corpora, COLING'92*, Nantes, p. 454-460, 1992.