Semantic visualization and meaning computation

Fabienne Venant

Semantic visualization and meaning computation

Fabienne Venant
Loria
Campus Scientifique
BP 239
54506 Vandoeuvre-les-Nancy
fabienne.venant@loria.fr

Abstract

We present here VISUSYN, a prototype we developed in order to study meaning construction. This software implements the model of dynamic construction of meaning proposed by Victorri and Fuchs (1996). It allows a semantic visualization that can be used to compute the meaning of a lexical unit in a given context

1 Introduction

Meaning construction – how words take meaning in a sentence or in a text –i s an important issue for computational linguistics. The major problem we have to deal with is that each word can have a lot of different meanings, depending on context. This phenomenon – called polysemy – is constitutive of language, and is the basis of its richness. However it is quite difficult to formalize. In most models of language, polysemy is considered as a kind of artefact. In these models, polysemy amounts to very little: a choice in a list of pre-existing meanings. However the omnipresence of polysemy always leads this kind of computation to combinatorial explosions. To avoid this problem, we want to give a central place in meaning construction to polysemy: that is why we define our model within the framework of continuous mathematics. This model was first proposed by Victorri and Fuchs (1996). It is deeply rooted in *Gestalttheorie* (Guillaume, 1979). Each linguistic unit is associated with a semantic space, where its different meanings are organized according to semantic proximity. The other units of the utterance define a potential function, which allows us to determinate the region of the semantic space corresponding to the meaning of the unit studied within the utterance.

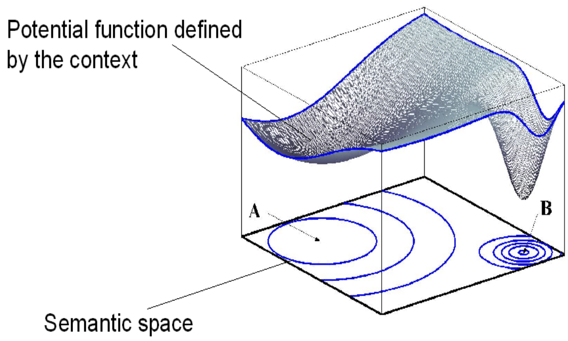

Figure 1: Model

We present here VISUSYN, the software we developed in order to implement this model. VISUSYN is still a work in progress. The working version was developed in Matlab. A Java version will be developed as soon as possible.

2 Presentation of VISUSYN

VISUSYN makes two kinds of operations, corresponding to the two componants of the model: it constructs semantics spaces and then computes meaning functions defined on these semantic spaces.

The algorithm for building semantic spaces relies on the analysis of a graph of synonyms. The aim of the exploration is to reveal the structure of the lexicon modelled by the graph so an automatic system can reach the information it contains. VISUSYN can construct local spaces, representing the semantic of a given unit, as well as global spaces, representing a lexical paradigm in its whole (for French adjectives, or adverbs for instance). Local spaces are used to compute meaning of the unit under study when accompanied by a given word in a sentence. To compute the corresponding potential function, VISUSYN uses cooccurence data from large scale corpora.

3 Small-world graphs

It has been discovered recently that most lexical graphs belong to the class of small-world graphs. This term denotes graphs where most nodes are neighbors of one another, but every node can be reached from other by a small number of hops or steps. Small-world networks were defined by Watts and Strogatz (1998). They noted that graphs could be classified according to their clustering coefficient (C) and their characteristic path length (L). Additionally, a third property can be associated with small-world networks even though it is not required for that classification. Specifically, if a small-world network has a degree-distribution which can be fitted with a power law distribution, it is taken as a sign that the network is small-world. These networks are known as scale-free networks. Ravasz and Barabási (2003) showed that a high clustering coefficient with scale free topology determines an original combination of modularity and hierarchical organization. It is not a simply pyramidal organization. The structure is made of groups of nodes, with small clusters at the bottom and very large groups at the top. Moreover, groups of nodes may overlap at any level. This self-similar nesting of different groups or module into each other forces a strict fine structure on real networks (Ravasz and Barabási, 2003).

4 Building semantic spaces

The algorithm will be here illustrated on the French adjective lexicon. The graph under study, called Synadj, is a graph of synonymy with 3,699 vertices and 22,568 links. We verified the small-word structure of this graph (C, L and degree distribution). This small-world structure led us to use the cliques[1] of the graph as a tool for building the semantic space. A clique in a graph is a maximal set of pairwise adjacent vertices, or – in other words – an induced subgraph which is a maximal complete graph. In the present case, a clique is made of adjectives which are all synonyms in a one to one relationship. By virtue of the definition, small-world networks will inevitably have high representation of cliques, and subgraphs that are a few edges shy of being cliques, i.e. small-world networks will have sub-networks that are characterized by the presence of connections between almost any two nodes within them. This follows from the requirement of a high cluster coefficient. We can consider as a first approximation that the cliques define very precise meanings that can be considered as the intersection of the meanings of all the units belonging to the clique. We thus define the semantic space as the euclidian space generated by the vertices of the graph (the adjectives, here). Each clique of the graph is associated with a point of this space, which coordinates depend on wich vertices belong to the clique. VISUSYN uses the chi-square distance to compute the distances between the cliques. Then a principal component analysis is applied to reduce the dimensionality of the space. In order to build a local semantic space (for example associated with a given word), we select a sub-graph, made only by the word under study and all its synonyms. The local space only contains the cliques of this sub-graph[2]. Figure 2 shows a visualization of the semantic space associated with the French adjective *sec* (*dry, severe, brusque...*). It accounts for the six main meanings we can find in a dictionary.

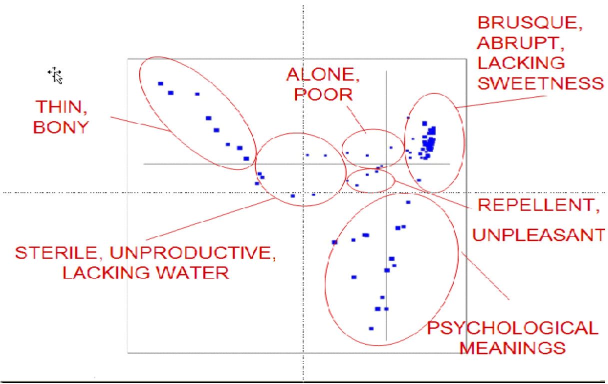

Figure 2: Local space associated with the French adjectives *sec*

The global semantic space is generated by all the vertices of Synadj. It contains all the 11,900 cliques of Synadj. Figure 3 shows a part of the global space of French adjectives. The

[1]Following the idea first proposed by Ploux and Victorri (1998)

[2]Ploux and Victorri (1998) first proposed the algorithm and built the lexical data. Local semantic spaces can be consulted at http://elsap1.unicaen.fr/dicosyn.html or http://dico.isc.cnrs.fr/fr/index.html.

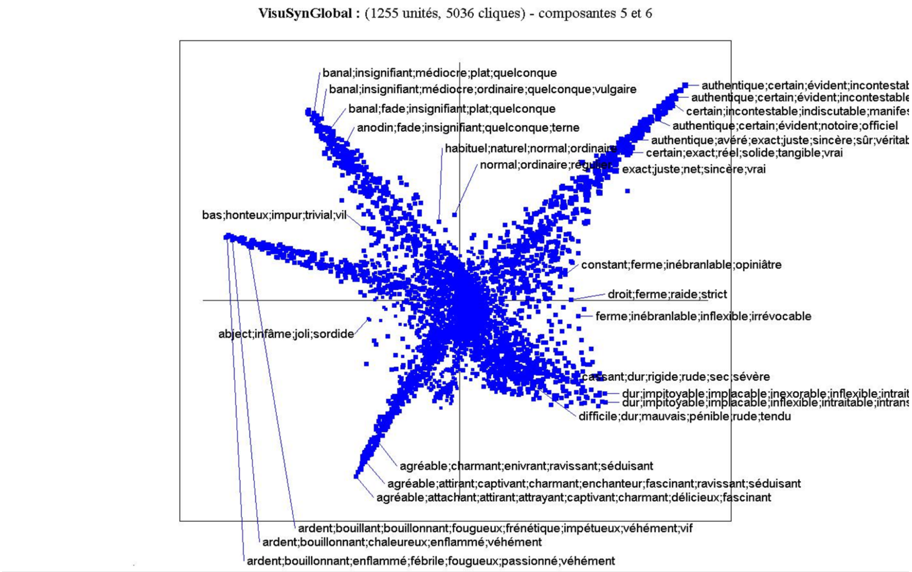

Figure 3: Global semantic space: French adjectives (2D)

very centre of the space only contains intensive meanings like *authentique; certain; evident; incontestable (authentic; certain; evident; incontestable)* or *agréable; charmant; enivrant; ravissant; séduisant (agreeable; delightful; exciting; attractive).* These meanings are very general meanings and can apply to any nouns.

There are now many semantic branches more or less long growing out from the central core. These branches are very dense near the centre and then go in all the directions becoming sparser and sparser. They are homogenous from a semantic viewpoint. Each branch only contains one sort of adjectival meaning: relational, qualificative, or intensive. VISUSYN had also been used to explore other global lexicons, like French verbs (Figure 4 – (Gaume et al., 2006)) or French adverbs ending in -ment (more or less corresponding to english adverbs ending in -ly).

5 Meaning Computation

Let's go on with French adjectives. An attributive adjective is always linked to a noun. It is this noun which mostly constrains the meaning of this adjective, even if other units like the article can play a role.

We show here how Visynsyn can be used to automatically find which synonyms match better the meaning of an adjective (here the French adjective *sec*) when used with a given noun. In order to do this, VISUSYN associates a characteristic func-

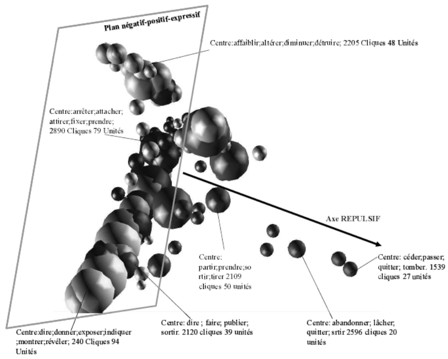

Figure 4: Global semantic space: French verbs (3D)

tion with each synonym. To compute the value of the function in a given point, it evaluates wether the synonym belongs to the corresponding clique or not. The basins of this function represent the meaning zones of the semantic space in which the synonymy between the word and the given synonym is relevant.

To compute the meaning of *sec* when used with a given noun, VISUSYN associates a potential function with the noun. The value of this function in each point depends on the frequencies of cooccurrence of the noun with the adjectives of the corresponding clique. The basins of the function determine the zone of the semantic space corresponding to the meaning of *sec* when used with the associated noun. Figure 6 shows the potential function associated with *fleur* (*flower*). In this case the noun *fleur* forces *sec* to take a precise meaning in the zone 'lack of water'.

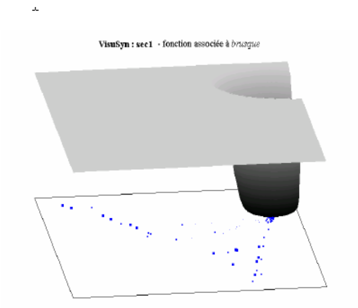

Figure 5: Characteristic function associated with the adjective *brusque*

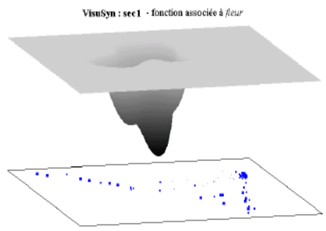

Figure 6: Potential function associated with *fleur (flower)*

The method of disambiguation consists in comparing the function of a synonym to the function of the noun under focus. The more the functions overlap, the more you can replace *sec* with its synonym without changing the meaning of the syntagm. This method has been used to compute the overlap rate for 20 nouns among the most frequently used with *sec* in a large corpus (Frantext, http://atilf.atilf.fr/frantext.htm). The results automatically processed were compared with the human answers on the same task. The rate of success is 79% (Venant, 2006).

This works for a fined grained disambiguation. VISUSYN can also deal with a more macrocopic partition of the semanctic space. Figure 2 shows a division of the semantic space of *sec* in 6 zones. VISUSYN can decide automatically which of these six zones correspond to the meaning of *sec* , when used with a given noun. The method is similar to the previous: It now associates each zone (and not only each synonym) with a characteristic function. This function is computed according to the cliques belonging to the zone. For each noun, VISUSYN compares its potential function with that of each zone. It thus computes an affinity rate between each noun and each zone. This method gives better results than the previous

6 Conclusion

Although VISUSYN is still a prototype, it seems very promising. Our basic result is that the kind of visualization presented here displays the structure of lexical graphs. It also constitutes an original tools to explore the structure of small-world graphs. As the algorithm is independent of the nature of the relation modelized by the graph, it could be used to explore other small-world graphs

It offers interesting possibilities of using semantic informations in computational linguistics, especially in the domain of automatic disambiguation. (We also evaluated its results on verbs and noun). From a more theoretical point of view, it accounts for the validity of the underlying model. Of course this model has some limits and the work is still in progress. We have to develop the system. However, this work shows how continuous mathematics can be relevant for semantic modelization and encourages us in the challenge of using continuity for corpus linguistics.

References

Gaume, Bruno, Fabienne Venant, and Bernard Victorri. 2006. Hierarchy in lexical organisation of natural languages. *Hierarchy in Natural and social Sciences Denise Pumain (ed)*, pages 121–142.

Guillaume, Gustave. 1979. *La psychologie de la forme*. Flammarion.

Ploux, Sabine and Victorri Bernard. 1998. Construction déspaces sémantiques á láide de dictionnaires de synonymes. *Traitement automatique des langues*, 39:161–182.

Ravasz, Erzsébet and Albert László Barabási. 2003. Hierarchical organization in complex networks. *Physical Review*, E 67:026112.

Venant, Fabienne. 2006. A geometric approach to graph exploration. In *Fourth Workshop in Information and Human Language Technology (TIL'2006)*.

Victorri, Bernard and Catherine Fuchs. 1996. *La polysémie, construction dynamique du sens*. Hermès.

Watts, Duncan .J. and Steven H. Strogatz. 1998. Collective dynamics of small-world networks. *Nature*, 393:440–442.