Meaning representation: from continuity to discreteness

Fabienne Venant

Meaning representation: from continuity to discreteness

Fabienne Venant

Loria/Université Nancy2
Nancy France
fabienne.venant@loria.fr

Abstract

This paper presents a geometric approach to meaning representation within the framework of continuous mathematics. Meaning representation is a central issue in Natural Language Processing, in particular for tasks like word sense disambiguation or information extraction. We want here to discuss the relevance of using continuous models in semantics. We don't want to argue the continuous or discrete nature of lexical meaning. We use continuity as a tool to access and manipulate lexical meaning. Following Victorri (1994), we assume that continuity or discreteness are not properties of phenomena but characterizations of theories upon phenomena. We briefly describe our theoretical framework, the dynamical construction of meaning (Victorri and Fuchs, 1996), then present the way we automatically build continuous semantic spaces from a graph of synonymy and discuss their relevance and utility. We also think that discreteness and continuity can collaborate. We show here how we can complete our geometric representations with informations from discrete descriptions of meaning.

1.. Linguistic issues

Why should we use continuous semantic representations? As soon as we deal with semantic description, we are confronted with the question of categorization. A main lexical meaning, as it can be found in a dictionary for instance, can be viewed as an equivalence class of basic semantic elements. Depending on the theoretical framework these elements can be precise semantic nuances, semantic features, contextual-usage meanings of words, logical representations... A persistent question is to define criteria in order to decide if one given element belongs to one given class. As one single criterion is generally not sufficient to characterize a class, we use sets of criteria. This lead us to consider question of graduality :

> expressions satisfying the whole set of criteria can be said typical elements of the corresponding class, whereas other expressions can be viewed as more peripheral, further from the center of the class as they satisfy a smaller number of criteria? (Victorri, 1994)

One way to evade this duality, is the use of topological representations, associating each lexical item with small graphic configuration that outlines, in a continuous way, the kernel of their semantic value instead of splitting them into discrete classes. (Culioli, 1990; Lakoff, 1987; Langacker, 1987)
Using this kind of representation enables us to give to polysemy a central place in meaning construction. 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 preexisting meaning classes. However the omnipresence of polysemy always leads this kind of computation to combinatorial explosions. That is why we define our model within the framework of continuous mathematics. This model, called dynamic construction of meaning, was first proposed by Victorri and Fuchs (1996).

2.. Theoretical framework

As can be seen in Figure 1, 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. Thus, the precise meaning of a polysemous unit in a given sentence is modeled by a dynamical system on its semantic space. The dynamic is parametrized by the other units present in the sentence. Each attractor of the dynamics correspond to a possible semantic value.
This model can account for various semantic phenomena, like ambiguity or indetermination (depending on the number of attractors of the dynamic, and the form of their basins). In this paper, we will focus on the central element of this model: the semantic space upon which semantic phenomena are modeled. This space as to be continue and to account for the semantic topology of a lexical item. By semantic topology, we mean the different meanings it can take, depending

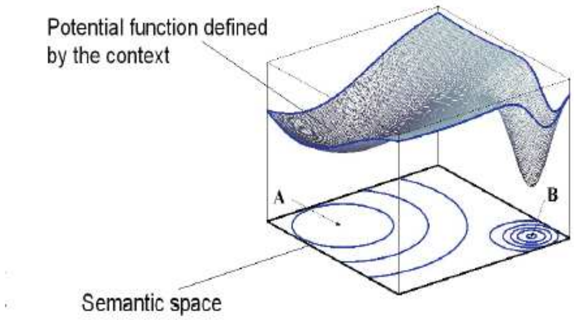

FIGURE 1. Model

on the sentence it is embedded in, the way this different meanings are organized within a lexicon, and their semantic similarities.

3.. Building semantic spaces

Ploux and Victorri (1998) developed an algorithm to automatically build semantic spaces. It relies on the analysis of a graph of synonyms. This algorithm constitutes a very useful way to explore such a graph. It reveals the structure of the lexicon modeled by the graph, so an automatic system can reach the information it contains. This method enables us to 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 can be used to compute the meaning of the unit under study when accompanied by a given word in a sentence.

Let's illustrate the construction of semantic spaces on the French verb *abandonner* (*to abandon, to be eliminated, to desert, to drop, to drop out, to forsake, to give up, to leave, to leave behind, to relinquish, to renounce, to retire from, to run out on, to walk out on, to withdraw from*). The graph of synonymy is provided by the Dictionnaire Electronique des Synonymes (DES, `www.crisco.unicaen.fr`). Given a lexical item like *abandonner*, we need the subgraph formed by *abandonner* and all (and only) its synonyms. Figure 2 shows an excerpt of this graph. The underlying idea is that only one synonym is generally not sufficient to define a meaning of a given lexical item.

We can see in Figure 2 that *laisser (to leave)* is at the same time synonym of *quitter (to quit)* and of *confier (to entrust)*, which correspond to two different meanings of *abandonner*. We thus characterize a meaning by a set of synonyms. To be more precise, we use cliques of the graph. 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.

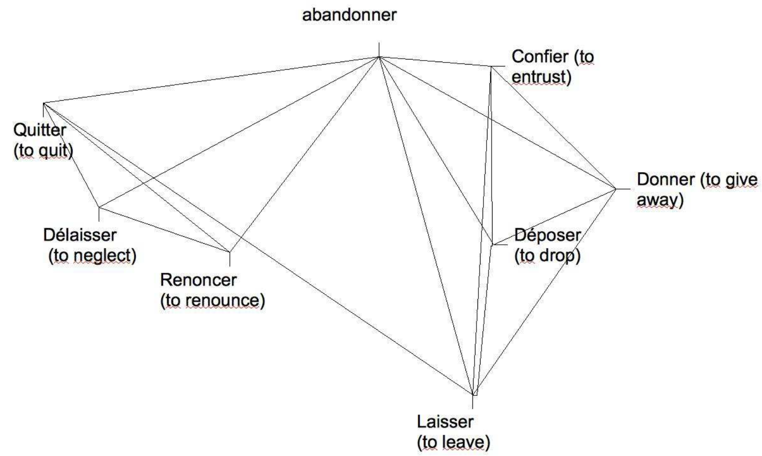

FIGURE 2. Subgraph of synonymy for *abandonner* – an excerpt

In the graph in Figure 2, there are three cliques

$$< abandonner, d\acute{e}laisser, quitter, renoncer >$$

$$< abandonner, laisser, quitter >$$

$$< abandonner, confier, d\acute{e}poser, donner, laisser >$$

We can consider as a first approximation that a clique corresponds to a precise meaning of the word studied. We define the semantic space as the space generated by the synonyms of the lexical unit studied. Each clique of the subgraph corresponds to a point whose coordinates depends on the synonyms it contains. Ploux and Victorri showed that the canonical Euclidean distance does not work in this space. This distance does not account for real semantic proximity because it gives the same weight to all the cliques and all the synonyms. Thus, they proposed to use the chi-square distance. This works better because each synonym is balanced according to the number of cliques it belongs to, and each clique according to the number of synonyms it contains. The more a synonym belongs to different cliques, the less it is specific, the less the role it plays in meaning discrimination is important. The more a clique contains non specific synonyms, the nearer its corresponding point to the origin of the space. Figure 3 shows the semantic space built for *abandonner*. We obtain such a 2D representation via a component projection. Its main interest is to account for a continuum of meaning. Figure 3 shows four main meaning poles, and the way we can, in a continuous way, go from one to another.

We can compare this representation with the article consecrated to *abandonner* in the French dictionary TLFI (Trésor de la Langue Française informatisé).

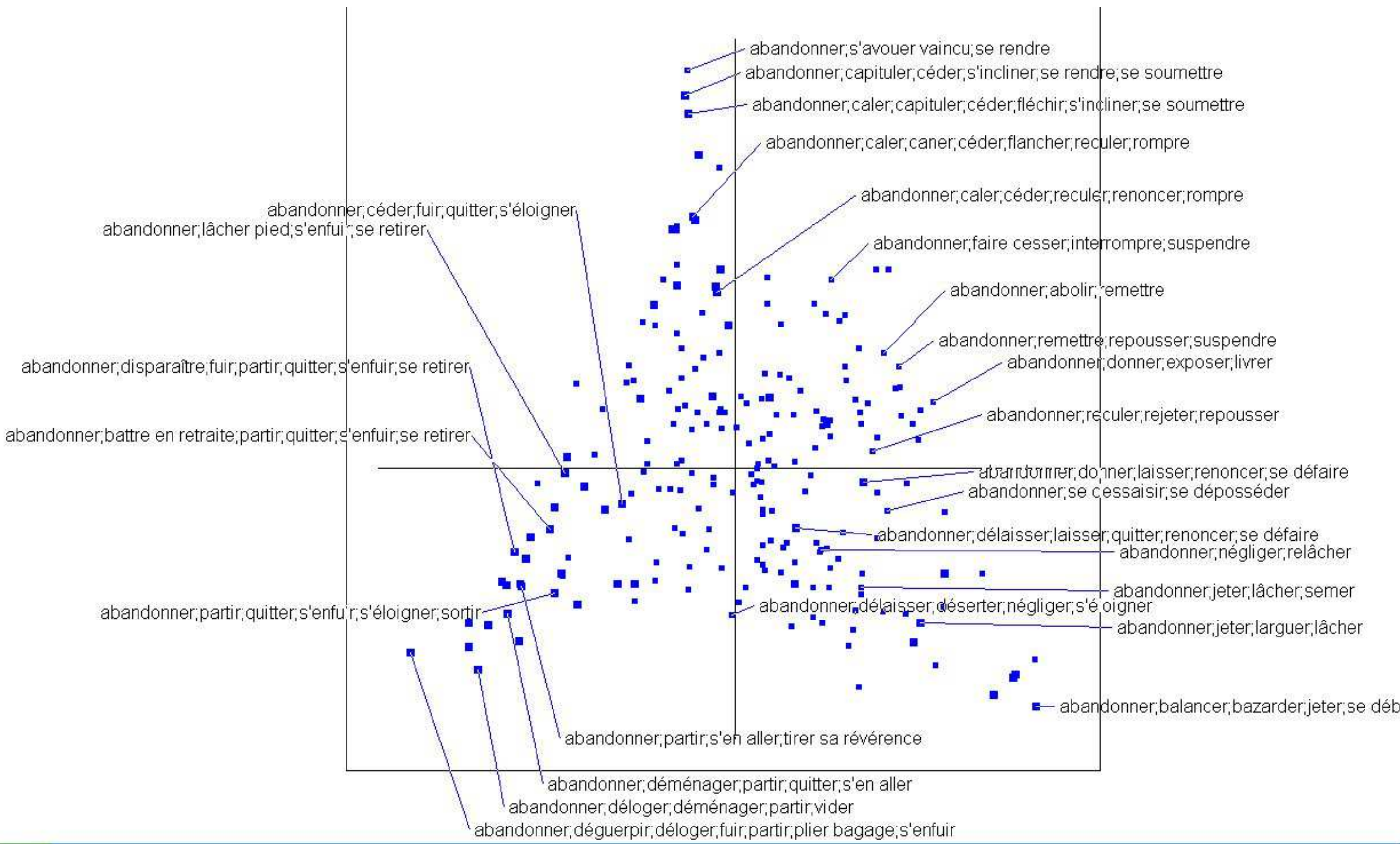

FIGURE 3. Semantic space of *abandonner*

Abandonner is very polysemic, but we can group its meanings in six main areas:

(1) to renounce, to abdicate (to give up a power, a right, owning something)
(2) to leave a place
(3) to give up a principle, a cause, common sense...
(4) to give up an action, an activity, a project...
(5) to leave someone
(6) to entrust someone or something to someone

What we see on Figure 3 is that the superior part of the semantic space contains synonymes like *(s'avouer vaincu (to admit defeat), céder (to give in), s'incliner (to be defeated) interrompre (to interrupt), renoncer (to give up)* which correspond to meaning 1, at the very top, and meaning 4 just below. This notion of dispossession leads us to synonyms like *jeter (to throw), lâcher (to drop), larguer (to drop, to dump, to chuck up)* wich correspond to meaning 5. The left inferior part of the space contains synonyms like *déguerpir (to clear of), fuir (to run away)* which correspond to meaning 2
The important points are the following:

- we obtained the semantic space through a totally automatic process
- it really accounts for main senses from a dictionary and enables to figure out the relationship between one meaning and each other
- it accounts for a continuum in the semantic of lexical items: we can see a continuous path to go from one sens to another. Frontier between main senses are hard to define.

4.. Graduality and granularity

As we said, our model is based upon a continuous representation of meaning, and more precisely on the use of cliques as the minimum semantic unit for meaning representation. Such a representations is very useful to accounts for subtle semantic phenomena or to deeply study the semantic of a given lexical item
We defined an automatic method, based on coocurences data from a big corpus, to define potential functions (as in Figure1) and be able to disambiguate an item giving the previous or following word (Venant, 2008; Jacquet, 2004) Such a task requires to precisely define what is a contextual meaning. We can't use a clique as a semantic tag, because semantic distinctions between cliques are too subtle. Thus we need to group cliques in order to obtain a more macroscopic tagset. For instance, we'd like to group cliques

in sets corresponding more or less to dictionary definitions. So far we manualy defined macroscopic meaning zones in semantic spaces. Several ways of automating this fundamental step are under study. They bring into play several methods. For example clustering enables scale changes in visualization. An other way is to obtain granularity changes is to interact with other semantic resources. For instance, we tried to match cliques of a given item with semantic descriptions coming from a discret model. We first exploited the TLFI definitions. The first step uses results described in Falk and *al.* (Falk et al., 2009) , a work where we used a similarity measure to pair all possible meanings of a given item with a relevant set of synonyms. This work was done for 27 French verbs.
The next step is to project the each pairs (synonym, definition) on the semantic space. In a first experimentation, we computed for each clique of the verb *abandonner* an affinity rate between this clique and a given definition from the TLFI:
Given c a clique and d a definition:
We call Sc the set of synonyms c contains and Sd the set of synonyms associated with d according Falk and *al.*. The affinity rate between Sc and Sd is called T(c,d) and defined as follow:

$$T(c, d) = |Sc \cap Sd|/|Sc|$$

Let's have a look on Figure 4 and 5. On Figure 4, we automatically marked with an asterisk the cliques which rate with the definition 1 'To renounce, to abdicate' is more than 60%. We can see that this cliques are relevant, containing synonyms like *renoncer (to renounce), plaquer (to plack in), laisser tomber ((to ditch)...*
On Figure 5, we marked with an asterisk the cliques which rate with the definition 5 'To leave someone' is more than 60%. Once again, this cliques are relevant, containing synonyms like *quitter (to leave), délaisser (to neglect), rompre(to break up), quitter (to leave)...*
We think that this affinity rate can constitute an efficient way to determine frontiers between macroscopic meaning. It can be a way to automatically structure our semantic spaces with information provided by other semantic resources. For instance, we could enrich our semantic representation with information like subcategorization, domain, construction, and use them in our disambiguation process.

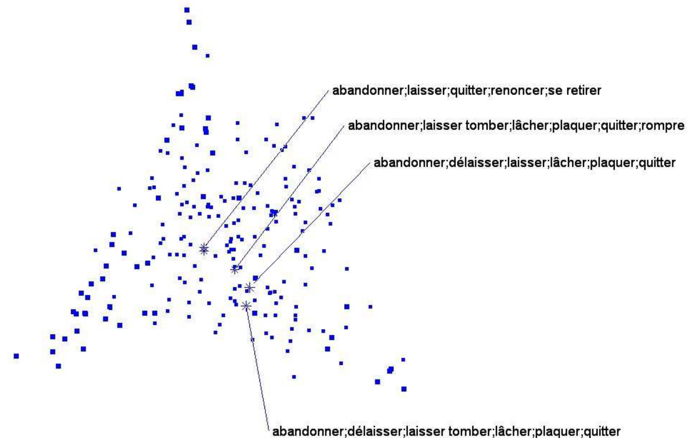

FIGURE 4. Semantic space of *abandonner*
$T(c, \text{"}To\_renounce, to\_abdicate\text{"}) > 60\%$

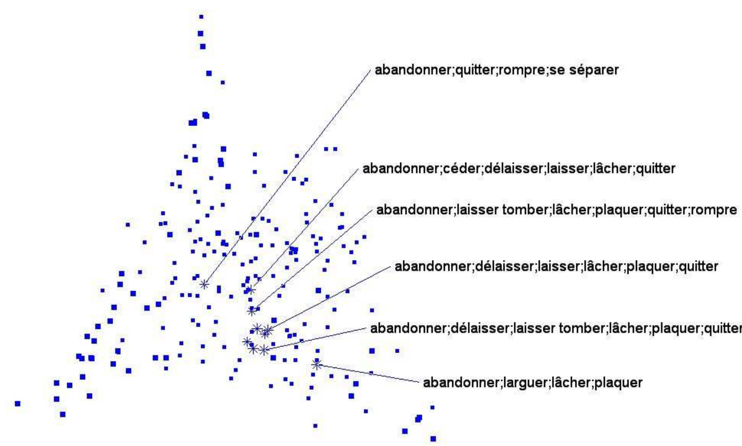

FIGURE 5. Semantic space of *abandonner*
$T(c, \text{"}To\_leave\text{"}) > 60\%$

5.. References

A. Culioli. 1990. *Pour une linguistique de l'nonciation*, volume 1. Ophrys.

I. Falk, C. Gardent, E. Jacquey, and F. Venant. 2009. Grouping Synonyms by Definitions. In Ruslan Mitkov, editor, *RANLP*, page 6, Borovets Bulgarie.

G. Jacquet. 2004. Using the construction grammar model to disambiguate polysemic verbs in french. In *procedings of ICCG3 (International Conference on Construction Grammar)*, Marseille.

G. Lakoff. 1987. *Women, fire and dangerous things : what categories reveal about the mind.* University of Chicago Press.

R.W. Langacker. 1987. *Foundations of cognitive grammar.* Standford University Press.

S. Ploux and B. Victorri. 1998. Construction d'espaces sémantiques à l'aide de dictionnaires de synonymes. *Traitement Automatique des Langues*, (39):161–182.

F. Venant. 2008. Semantic visualization and meaning computation. In *demo*. COLING 04.

B. Victorri and C. Fuchs. 1996. *La polysémie, construction dynamique du sens*. Hermès, Paris.

B. Victorri, 1994. *Continuity in linguistics semantics*, chapter The use of continuity in modeling semantic phenomena. C.Fuchs and B. Victorri– Benjamins.