Version de travail (non définitive)
A paraître dans : *Language and Complexity* (F. La Mantia dir.), Springer, 2016

Continuity in the interactions between linguistic units

Gilles COL, Rossana DE ANGELIS, Thierry POIBEAU

Laboratoire LATTICE (UMR 8094)
PSL Research University
CNRS/ENS/Université Sorbonne nouvelle
1 rue Maurice Arnoux, 92120 Montrouge, France

Introduction

Linguistic tradition has produced descriptions that can be assimilated (at least for the most part) to discrete models: discrete models of categories, discrete grammar rules and discrete lists of word senses, among others. This view is now widely discussed and more and more linguists agree that this model is at best partial and, more importantly, cannot account for the whole complexity of human languages. For example, Manning (2003) clearly shows that a discrete model of categories is both too poor and too restrictive:

> «Categorical linguistic theories claim too much. They place a hard categorical boundary of grammaticality where really there is a fuzzy edge, determined by many conflicting constraints and issues of conventionality versus human creativity.
> Categorical linguistic theories explain too little. They say nothing at all about the soft constraints that explain how people choose to say things (or how they choose to understand them).» (Manning 2003 : 297)

Similarly, dictionaries are interesting tools for providing references and definitions for humans but they are very unsatisfactory outside prototypical cases. It has been shown by a wide range of different authors that definitions overlap, and that different word senses (i.e. different dictionary entries) can correspond to a same occurrence in a corpus. The notion of graded word sense has been proposed (Erk and McCarthy, 2009), which is at the same time interesting and intriguing since it is not clear how it could afford the traditional model of dictionaries.
New linguistic models have thus been developed that address (at least to some extent) some of the limitations of traditional linguistic descriptions. In our opinion, two key features have contributed to the renewal of the domain: *i)* the availability of very large corpora that make it possible to access massive data and sets of examples and *ii*) new computational tools that make it possible to observe linguistic data as a complex system, i.e. as a system in which the key feature is the interaction of units between themselves. These interactions are complex, multi-faceted and multi-layered so as to give birth to dynamical and moving landscapes of linguistic behaviour. This has been stressed by different authors, see for example Girault and Victorri (2009):

> “Il faut donc changer de cadre de modélisation si l’on veut utiliser pleinement le nouvel observatoire que nous offrent les nouvelles technologies : les linguistiques de corpus se doivent d’être des théories s’appuyant sur les mathématiques du continu” (Girault, Victorri 2009: 153)

In computational linguistics, continuous models were marginally used compared to approaches based on dictionaries and rule based approaches that were dominating the period 1950-1980. With the rise of machine learning methods (especially distributional learning and “deep learning”, see Socher et al. 2013 and the last section of the paper for a quick overview), continuous models tend to now be dominating. It seems that it is not the case in linguistics yet, despite various attempts since one or two decades.
Continuous models are highly challenging for linguistics. But if interaction is at the heart of the linguistic process, it seems necessary to explain what is the source of the interaction, how it can be defined and what can be its dynamics. There is at least two apparently opposite views which could account for linguistic interactions and more particularly for their dynamics: a compositional one (i.e. the meaning of a complex unit depends on its parts) and a gestaltic one (i.e. the meaning of a complex unit is more than the

meaning of its parts). In fact, these views are not contradictory but should be combined. In what follows, we will see how we can reconcile them thanks to the notion of «instruction» and more specifically to the very content of these instructions.
The notion of instruction is basically considered here as a means to assess the notion of continuity since the content of an instruction is in fact a dual process of convocation and evocation. All linguistic units are considered as giving an instruction, thus taking part to the global meaning of the utterance and at the same enriching its own local meaning.
Some of these models, combining topology and dynamics, are presented in Section 1. Section 2 is devoted to the notion of instruction in a specific model (*gestalt compositionality*) and the last section enlarges upon continuity in language modelling.

I. Topological and dynamic models in linguistics : from « autopoesis » to the notion of « instruction »

1.1. A **short review of some models**

Autopoeisis

Maturana and Varela (1980) propose a topological and dynamical model of living systems inspiring a topological and dynamical model of cultural systems. First distinguishing between *self-referred systems* – only referred to themselves, as living systems – and *allo-referred systems* – referred to a context – they propose an epistemological model in which cultural systems are conceived as texts (Maturana & Varela 1980: xiii). Speaking about «systems defined as unities through the basic circularity of their production of their components» (Maturana & Varela 1980: xiv), Maturana defines *autopoiesis* as «what takes place in the dynamics of the autonomy proper to living systems» (Maturana & Varela 1980: xvii). Developing this concept, «notions of purpose, function and goal are unnecessary and misleading» (Maturana & Varela 1980: xix). Then being an *autopoietic system* means 1) to be autonomous, to present itself as 2) an individuality and 3) a unity, 4) and do not have inputs or outputs.
The notion of *autopoiesis* is based on the fundamental notions of *unity*, *organisation* and *structure*. *Unity* is «an entity distinct from a background, [which] characterizes both unity and background with the properties with which the operation endows them, and specify their separability. A unity thus specified is a simple unity that defines through its properties the space in which it exists and the phenomenal domain which it may generate in its interactions with other unities» (Maturana & Varela 1980: xix). Once identified a system as a unity, we can repeat this operation of distinction – the first perceptual and conceptual operation made by the observer – in a composite unity, going on with the identification of the components constituting the unity as a whole. «The relation between components that define a composite unity (system) as a composite unity of a particular kind, constitute its organisation. In this definition of organization the components are viewed only in relation to their participation in the constitution of the unity (whole) that they integrate» (Maturana & Varela 1980: xix). This specific definition of *organisation* calls back the hermeneutic conception of *text* considered as a whole. So, adopting the autopoietic model in linguistic analysis, we have to consider this hermeneutical principle. Finally, the notions of *unity* and *organisation* are in relationship with the notion of *structure*, identifying the actual relations which hold among the components (Maturana & Varela 1980: p. 77).
Focusing on the notion of *structure* we can explain different topological and dynamical models stem from different epistemological paradigms. For instance, the *gestaltic structuralism* (De Angelis 2014) carries out from the encounter between *Structuralism* and *Gestalttheorie*.

Gestaltic structuralism

Gestaltic structuralism is based on the assumption that any perception has to be considered referred to a biological organism and a perceptive organisation, what explains the common agreement of *Structuralism* and *Gestalttheorie* on the concept of *structure* (Petitot-Cocorda 1990: 29-33). Perceptions are considered as *structures* (*Gestalten*), organic unities individualized in the spatial and temporal field of perception (Guillaume 1979: 23-27). Structures are organised and complex, resulting from a process of generation in which the constitutive elements can't be considered independently from the whole in which they are included.

One of the basic principles of *Gestaltpsychologie* is that things we observe in the environment are perceived regularly as stable (Köhler 2000: 191 sgg.). So the object of *gestaltic structuralism* is the way structures maintain stable or change, and how we can identify them in spite of their evolutions.

Morphogenesis

The *theory of catastrophes* proposed by Réné Thom (1980, 1985) makes a synthesis between two concepts, *morfogenesis* and *structure*, and two different structuralisms, *biological**[1]*** and *linguistic**[2]***. Thom shows the way structures ensue one another, first identifying this phenomenon as *morphogenesis*, then identifying the processes able to create or destroy structures (Thom 1985: 3). In fact, structures can occupy a certain space and last a certain time (Thom 1980: 3). The possibility to recognize a structure as being always the same depends on the possibility to recognize a plurality of aspects identifying the structure itself as a same object of analysis. Finally, the concept of *morphogenesis* supposes to identify a certain discontinuity interpreted as a change of structures itself.

The morphogenetic model applied to structures in general is then developed by Petitot-Cocorda in a more specific morphogenetic model applied to linguistic structures. As Thom himself stresses introducing his work (Petitot-Cocorda 1990: 8), the approach to linguistic analysis proposed by Petitot-Cocorda has an important ontological consequence: we consider structures as a reification of connections, so we recognize to the terms connected a *positional value* and we reject *a priori* all phenomenological aspects of the terms composing the structure as a whole (Petitot-Cocorda 1990: 15). This model adopted in linguistic analysis is based on a *dynamical topology* (Petitot-Cocorda 1990: 8), concerning places and connections, based in turn on a *differential topology* (Thom 1985: 4), concerning specifically the differential connections of the terms. So, supposing that the *Gestaltic structuralism* is based on a *topological* model, despite of a logical one, Petitot-Cocorda proposes a *morphogenesis of sens*. This perspective is also based on an hermeneutical principal: as Thom says, *topology* is a mathematical discipline which lets to consider *the passage from global to local* (Thom 1980: 169). Adopting this point of view in semantics, we can identify *semantic structures* as stable, complex, locally determined, and connected to other semantic structures, all together constituting a whole, a global structure which is also stable, complex and coherent: the *text*.

According to the *morphogenesis of sense*, each structure is stable and (auto)governed by the connections defining the *positional values* of the terms involved: because it is topologically determined, the value of a semantic entity is defined only by the place it occupies in the structure. So the *positional values*, their stability and change, can explain the organisation, the stability and the transformation of the semantic structures. Finally, the notion of *positional value* supposes the notion of *differential value* proposed by Saussure[3]: we can identify linguistic entities only by the differential connections they entertain in language.

1.2. The articulation between *global* and *local* dimensions

The relation between *global* and *local* dimensions in linguistic analysis is essentially drawn by the *gestaltic structuralism*. However, this relation is stressed also in general linguistic. As Victorri and Fuchs (1996) remind us, formerly we can read about the relation between global and local in F. de Saussure's *Cours de linguistique générale*:

> « Il ne suffit pas de considérer le rapport qui unit les diverses parties d'un syntagme entre elles (par exemple *court* et *tous* dans *contre tous*, *contre* et *maître* dans *contremaître*); il faut tenir compte aussi de celui qui relie le tout et ses parties (par exemple *contre tous* opposé d'une part à *contre*, de l'autre à *tous*, ou *contremaître* opposé à *contre* et à *maître*). » (Saussure [1916] 1972: 172).

Also E. Benveniste speaks about this relation in his work *Problèmes de linguistique générale*:

> « Une phrase constitue un tout, qui ne se réduit pas à la somme de ses parties; le sens inhérent à ce tout est réparti sur l'ensemble des contituants. Le mot est un constituant de la phrase, il en effectue la signification; mais il n'apparaît pas nécessairement dans la phrase avec le sens qu'il a comme unité autonome. » (Benveniste 1966: 123-124).

Nevertheless, the relation between global and local dimensions becomes crucial to identify *semantic*

forms in the approach named *interpretative semantics* as developed by François Rastier. As we can read in his work *Sense et textualité*, understanding a linguistic sequence is an activity consisting in identifying *semantic forms* which are already understood or produced during the analysis of text (Rastier 1989: 9). The *interpretative semantics* has to be conceived as belonging to the rethorical-hermeneutical tradition (Rastier [1987] 2009, 1989). In fact, one of the fundamental principles in hermeneutics concerns the relation between global and local dimensions of texts: particular entities are determined by their relation to the text conceived as a whole which is in turn composed by them. Nevertheless, we can recognize the same principle in the *gestaltic structuralism*.
The notion of *semantic form* evokes the one of *perceptive form*: in fact, Rastier considers language in general, expression and meaning in particular, as perceptive objects, so he proposes the notion of *semantic perception* (Rastier 2009: XIV). What he calls *semantic perception* depends on a hierarchy of three kinds of data: the *forms*, the *grounds*, the *backgrounds* from which grounds and forms emerge, that is to say the paradigms of other grounds and forms concurrent connecting the actual semantic perception to the corpus of previous linguistic experiences.
The relation between *semantic forms*, *grounds* and *backgrounds*, evokes the way in which the *Gestaltpsychologie* explains perceptive phenomena, in particular the ones of visual perception (Köhler 2000: 188 sgg.). When a visual *form* is perceived, the one emerging by subtraction is absorbed by the *background* and its own *form* becomes invisible. However, when the last one emerges from the *background*, the first one disappears. The *form* emerging on the top and the *form* staying in the *background* can't be perceived at the same time. So the existence of a visual form depends on the visual unite emerging from the background. In parallel, in semantics a *form* implies the existence of a semantic unite which acquires its *form* after being isolated from the *background*. The *form* stays on a semantic *ground* on which differently extended semantic *unites* stand on.
In the *interpretative semantics* the relation between *forms* and *background* depends on the differentiation of *semes* (the smallest units of meaning). The *semes* are distinguished as *generic* and *specific*: the first ones represent the *ground* of a lexical class – or an *isotopy* – from which the second ones emerge identifying the *semems* (the content of a morpheme) constituting this class as *semic molecules* (a group of *semes*, not necessarily lexicalized) which go through the text. So we can realise in semantics what happens for the *perceptive forms*: during the interpretation of a text, *semantic forms* emerge from the background, distinguishing themselves from the other ones dispersed in the text.
So the sense is re-produced tracing a network of *semantic forms* laying on a network of *expressive forms* showing the *interpretative paths* of the text. The network of *semantic* and *expressive forms* is a global network constituted in turn by different local networks. As Rastier (2006) explains, concerning the dimension of meaning, networks are made by *semic molecules* and *semes* constitute their hubs; concerning the dimension of expression, networks are made by *phemic molecules* and their hubs are constituted by *phemes* (the smallest units of expression). The possibility to cross these two textual dimensions, meaning and expression, depends on the *interpretative paths* which take shape in the text and explain the bustle between a local and a global dimension according to an hermeneutical approach[4].

> «Le sens consiste pour l'essentiel en un réseau de relations entre signifiés au sein du texte – et dans cette perspective, les signifiants peuvent être considerés comme des interprétants qui permettent de construire certaines de ces relations. Elles demeurent de type perceptif: estimation de similarité, reconnaissance de forme, catégorisation» (Rastier 2001: 189-190).

Considered according to the *Gestaltic structuralism* (Rastier 2009), the notion of *semantic perception* (Rastier 1991: §VIII) can explain the identification and the management of the *semantic forms*. This notion represents also a synthesis of Rastier's approach to semantic analysis: to the cognitive perspective supposing the meaning subordinate to the concept and its compositionality, he opposes a structural and hermeneutic one, first kept close to the *Gestaltpsychologie* (Rastier 2010: 206). This approach reveals to be both *gestaltic* and *hermeneutic*, implying that the local dimension is determined by the global one, and supposing that the semantic analysis seems more like a process of identification of forms than like a process of computation (Rastier 2010: 207).

1.3. *Instruction*: a dynamical and topological notion

The notion of *instruction* in linguistic analysis comes from different epistemological paradigms. It is used by many linguists, from Harris (1954) to Fauconnier (1997) and more recently Tyler and Evans (2003) or Harder (2009). French linguists like Ducrot (1980), Ducrot and Anscombre (1983), or Nemo (2001) use it as well. We encounter it in the field of cognitive psychology (Barsalou 1999, 2008, 2010) as well as in the pragmatic tradition (Nølke 1994, Moeshler 2005).

According to Weinrich (1976) who proposes a «*C-I-T [Communication – Instruction – Text] linguistics*», *texts* are «communicative units» and *communication* is essentially a praxis whose pragmatic value is condensed in the notion of *instruction* (Weinrich 1976, chap. IV) considered as an order or a suggestion produced by a subject to induce his interlocutor to react anyway, even only understanding the text. The text considered as *texture* – made by a plurality of connections in which its components are involved – suggests to the interpreter the different steps of his interpretative action, offering to him some instructions to follow and go through. Furthermore, it has a continuous communicative relation with a *context*, so its analysis begins identifying big units composing the «text-in-context» delimited from some evident interruptions of communication allowing to consider the text itself as an autonomous unit (Weinrich 1976: 15-16).

Considered as a *texture*, the text is an *actual system* of connections which realises partially the possibilities offered by language considered as the *virtual system*. This process of actualisation includes the selection of different options and their presentation as a *structure*: «A structure treated by processing as a single unit can be termed a *chunk*, and may range from a *local* (small-scale) *micro-structure*, over to a *global* (large-scale) *macro-structure*. A chunked micro-structure constitutes a *micro-state* of the actual system, while a chunked macro-structure constitutes a *macro-state*» (Beaugrande 1980, on line). Linguistic elements becomes *instructions* for the actualisation: speakers need to know some *procedures* to have access to the resources offered by the linguistic system, as well as in its syntactic and semantic dimensions, supposing two kinds of knowledge: assertive and procedural. So each linguistic element becomes an *instruction* for the interpretative work made by the speaker, «e.g., a word being a pattern of sounds, a piece of a phrase, an instruction to "activate" a meaning, and so on» (De Beaugrande 1996: 11), as an intermediate step for a more general *instruction of sense* (Cf. Segre in Weinrich 1976, trad. it. 1988, p. 23).

Weinrich adopts a dynamical-pragmatic concept of linguistic sign: it's an instruction to act, supposing an intransitive concept of meaning. So the *semantic value* of a linguistic sign consists in showing to the interpreter what he has to do for interpreting and understanding. The model of communication is dynamical**[5]**: every linguistic sign is understood as an *instruction* for the interpreter to act in a particular context (Weinrich, 1976, trad. it. 1988, p. 15).

In the text conceived as a whole, linguistic signs become strategic components of its texture. In fact, they produce an expectation of sense which can be confirmed or denied by the text itself and by which the listener or reader develops his interpretative work (Weinrich 1976, trad. it. 1988, p. 87). Studying grammar in relation with the process of learning of foreign languages, Oller (1979) describes a phenomenon called *Expectancy Grammar*, consisting in the ability of the interpreter of a foreign language to make plausible hypothesis of what will be said or read in a specific context. The *Expectancy Grammar* depends on three parameters: 1) the awareness of situations and contexts, which makes anticipate the uses and the goals attended in communication; 2) the redundancy, that is to say supplementary information given by contexts, co-texts and paratexts; 3) the encyclopaedic knowledge, by which the interpreter creates some hypothesis on what will be said or written and solve the ambiguity of sense.

Semantic and syntactic information delivered by linguistic signs in relation with other linguistic signs belonging to a same linguistic system produce some expectations of sense in relation with the other textual components. Then Weinrich identifies three kinds of *preliminary informations* given respectively by 1) language; 2) text; 3) context. These ones represent the firsts steps to advance in global understanding of text. For instance, according to Weinrich, we can consider the title as a *macrolinguistic instruction to the expectation* (Weinrich, 1976, trad. it. 1988, p. 23, corsivo mio). So text becomes a *global structure* generated by the connection of different *local structures* whose connections make the *textuality* of the text (Weinrich 1976, trad. it. 1988, p. 24). Finally, all linguistic signs are for the reader *local instructions* whose connections give a *global instruction* of sense of the text considered as a whole (Weinrich 1976, trad. it. 1988, p. 24).

Referring to *textual pragmatics,* supposing a concept of text defined as a *communicative unit,* considering the relation between the linguistic and the extra-linguistic context, the speaker and the reader, and also

between the supposed encyclopaedic knowledge and their will to cooperate in communication (Bar-Hillel 1968: 270-271), the influence of textual pragmatics in Eco's theory of text is condensed in the notion of *instruction*, as we can observe first in the analysis of terms considered as «instructions oriented to the text» (Eco 1979: 15). So he refers to an *Instruktionssemantik* textually oriented[6].

The specifity of this notion in compositionnal gestalt semantics is that it implements the two step process of convocation-evocation and that it accounts for the assembly of the units of an utterance. Furthermore, an instruction is unique and has various effects. Eventually, an instruction is underspecified but the meaning of the unit enriches in the course of discourse.

II. Continuity, instruction and the online processing of meaning

2.1. Two fundamental hypotheses concerning meaning

Following the theoretical background exposed in the previous section, our analysis is based on two main hypotheses:

- Individual words do not correspond to a list of word senses. Instead, we think that meaning emerges from the interaction with the context. We thus adopt an 'emergentist' approach that radically contrast with traditional approaches defining the different meanings of a word a priori, generally through a structured list of entries stored in the dictionary.
- Another related hypothesis is that there are not clear-cut sense boundaries. Instead, we consider meaning as a continuous phenomenon that emerges dynamically from all the interactions of the different linguistic units within a sentence (and, more generally, within the context as it may also include non linguistic elements).

Both hypotheses need a thorough analysis, involving different kind of utterances so as to explore their consequence. For example, the first hypothesis considers meaning as an emergentist phenomenon but does this entail that linguistic units do not have any meaning per se? If this is the case, how could one give the definition of a word out of context? Or alternatively, if each word has a core meaning, how can we define this core meaning? As for the second hypothesis, what is the nature of continuity in linguistics? Is it just a metaphor? Does it involve a multidimensional analysis? In other words, is meaning directly related to continuity or does this continuity correspond to the composition of binary functions for example?
In the next section, we address some of these issues by considering practical examples in English, especially particle verbs and the case of "over", since these items are known to be hard to categorize in English.

2.2. The case of English complementation

Particle verbs are especially challenging in English. There has been numerous discussions to try to determine whether particle verbs can be decomposed or not (i.e. if he meaning of the particle verb can be derived from the meaning of the base verb combined with the meaning of the particle). Of course, it all depends on the particle verb under consideration: some particle verbs seem more compositional than others (if it makes sense to consider degrees of compositionality). If fact, the problem is not straightforward since the notion of particle verb itself is a difficult and not so well defined one. Base verbs can be complemented with words that are sometimes considered as particles (when they are closely linked to the verb), sometimes as prepositions (when the word seems more closely related to the complement noun phrase).

Things are probably clearer if we take an example. Let's have a look at the following phrases:

- *Run into the room*: *into* introduces a location (where?), cf. "enter the room";
- *Run into financial difficulties*: *into* introduces the topic (what?), cf. "encounter difficulties
- *Run into a telephone pole*: *into* introduces the topic (*what?* but not *where?*), cf. "collide with a pole"
- *Run into John*: *into* introduces a person (*who*? but not *where*?), cf. "*encounter John*".

From this example, one can see that the meaning largely depends on the noun that is involved: its nature (location, person or object) has a major influence on the meaning of the verb. The so called literal (or prototypical) meaning of *to run* (*to walk fast*) just reflects, in our opinion, a tendency to favour concrete

meaning over more abstract ones. Most cognitive linguists also defend the fact that literal meaning come first mostly based on evidence from complexity and learning theory. However, if one believes in usage and statistical significance, the supremacy of literal meaning does not hold and is even not justified any more. Instead, we observe that abstract meaning is prominent (in "*run into an obstacle*", *run* does not correspond to *walk fast* and the noun is not a physical object). Depending on the nature of the noun and the verb, into can be categorized as a preposition or a particle.

2.3. The specific case of 'over'

Let's take an example and let's now examine the case of "*over*". Over is know to be plurifunctional and highly polysemous and hence has been extensively studied by linguists, especially in the cognitive linguistics tradition, see Taylor 1988, Deane 1993, Dewell 1994, for the last century, and Deane 2005, Talmy 2005, Tyler and Evans 2001, 2003, Van der Gucht *et al* 2007 or Zlatev 2003 more recently. We have proposed our own study of over in (Col and Poibeau, 2013) where we have shown that patterns of usage can be derived from corpora.

The following table provides an overview of the main patterns that can be derived from a corpus analysis:

$N_{procès}$	over $_{prep\ adjoint}$	$N_{temporel}$
$N_{cognitif}$	over $_{prep\ argument}$	$N_{topique}$
$N_{contrôle}$	over $_{prep\ argument}$	$N_{topique}$
$V_{procès}$	over $_{prep\ adjoint}$	$N_{temporel}$
	over $_{adv}$	$Q_{quantifieur}$
$V_{procès}$	over $_{adv}$ and over $_{adv}$ over $_{adv}$ again	
$N_{processuel}$	over $_{adjectif}$	

Some examples seems quite easy and simple to analyse:

- "they are often doing speeds **over** 50mph.": *over* is here an adverb expressing the notion of 'excess';
- "he could see **over** the tops of the trees": preposition with a spatial meaning;
- "the idea of a black sheep, has changed **over** the years": preposition with a temporal meaning;
- "controversy **over** bluefin tuna": preposition expressing the topic;
- "in November 1989 he took **over** as CDU leader": particle expressing control.

These examples seem straightforward because of the high frequency of some patterns or constructions in corpora:

- N1 over N2 is a structure that favours the meaning of topic for N2 and thus entails the categorization of over as a preposition, esp. with N1;
- over $N_{quantifieur}$: favours the excess meaning of over;
- over $N_{Temporal}$ (over the years): favours the categorization f over as a preposition with a temporal meaning introducing a modifier[7];
- only a few verbs are regularly associated with over as a particle (be, go, take, come, turn) and are lexicalized as such in lexical databases like Wordnet.

One can also observe more complex cases. For example in "then they have to go through a grueling contest over a period of time", it is not clear whether "period of time" refers to a topic (in which case the complement is an argument) or just to a period of time (in which case the complement is a modifier). There is here a tension between "over a period of time" that is most of the time a modifier and contest over, which is generally followed by a noun referring to the topic of the contest.

Nuances of meaning can also been found in examples like "*Craig ran over the neighbor's cat*" *vs*. "*She ran over the fields, past the barn, and to the white house*". The nature of the complement of over entails a largely different interpretation of the scene described by each of these sentences.

By these examples, we hope to have made it clear that the meaning of a word like over cannot be reduced to one dimension and can hardly be defined in itself. Instead, it is the interaction with the context that makes it possible to justify:

- The continuity of meaning
  - The different categories over refers to;
  - The different function of the phrase introduced by over (esp. argument or modifier is we adopt a syntactic point of view)

It is probable that these very regular and highly frequent patterns are registered as such in memory. Bolinger has shown as soon as in the 1970s that the language is made of co-occurring words with different kinds of constraints. Bolinger (1968) proposes to call prefab local co-occurring units to encompass words, collocations and idioms. For Bolinger (1976) these prefab should play a major role in semantics, beyond just simple and isolated words. Construction grammars (Goldberg, 1995) can be seen as going a step further since they propose to generalize the analysis at different levels, taking into consideration the different kinds of interactions at any level in the analysis (words, idioms, phrases, sentences).
This tradition has also been acknowledge in the domain of corpus grammar. Collocations and prefab are especially important for Sinclair (1991) and there are attested traces of the interest of Bolinger for corpus linguistics (Fortis, 2014).

The previous paragraphs make it clear that the core meaning of a linguistic unit like "over" is to be understood and described in the complexity of its uses. So as to account for the various uses associated with syntactic patterns and categories, and the extremely large (or maybe unlimited) number of possible interactions of the unit with its co-text, Col et al (2012) suggests a model based on a gestalt compositionnal approach. We consider that the meaning of an utterance is constructed in the course of a double, dynamic process leading to a temporary, tentative semantic balance between the different interactions. This balance is also largely self-organised as it is the result of two opposed movements, a bottom-up one and a top-down one. Each unit interacts with its co-text, thus the meaning of each unit depends on the meaning of the other units with which it interacts, and the global meaning of the utterance retro-acts on the meanings of the units present in the utterance and the co-text. This approach is then both compositionnal and gestalt as it focuses on the dynamicity of the interactions in the construction of meaning.
From this viewpoint, the meaning of a unit does not have to encapsulate all its attested meanings. It does not have to express a very general, abstract meaning either. It is rather its dynamicity which should be its core meaning and more generally of any linguistic units, and this dynamicity should be expressed according to what Victorri (1996) calls a convocation-evocation principle. The double movement we described above actually corresponds to two steps of the construction of meaning. The first step determines what needs to be present in the co-text and more generally in the intersubjective field so that the unit can play its role in the construction. These elements that are necessary but not supplied by the linguistic unit itself are said to be "convoked". The second step refers to what the unit brings to the construction when interacting with other linguistic units. This interaction has an effect on the verbal scene under construction: this is what the unit "evokes". This capacity of convocation-evocation is what defines a linguistic unit; it is not its "meaning" per se but what we call its "instruction". As an example, the instruction supplied by "over" will be the following one (Col 2012 and Col et al 2014):

> Over convokes a bounded domain and evokes a movement of covering of the domain, including its bounds.

As can be noticed in the formulation of this instruction, no particular meaning is priviledged and no proto-meaning either. The bounded domain is neither spatial or temporal. It does not correspond to the use of "over" as a preposition or as an adverb but it is compatible with these various categories. In an example such as: "*Go over the bridge and turn right immediately onto a track leading into the trees*", it is clear that the "bounded domain" is a spatial one: it construes the space under the bridge (a river or a road) that is minimally delimited by both ends of the bridge itself. In "*As he released his anger, he felt his love for his ex-wife, and wept over their divorce*." The domain is not spatial but still "over" convokes a bounded domain, the divorce, whose boundedness comes from the specificity of the divorce, i.e. *their* divorce.It also comes from the imperfectivity construed both by the tense of the verb "weep" (*wept*) and by the morpheme <ex-> in *ex-wife*. This last example is particularly interesting as we clearly notice that it is in the interaction with *weep* and *divorce* that *over* succeeds in giving the meaning of "topic" to the sentence. Still, *over* alone does not make the Topic meaning emerge; it needs the other lexical units to construe it, and particularly an 'object to weep over'. The interaction between *over* and the lexical units of the uterance (as

well as its recurrence in regular pattern like N_1 $over_{prep}$ N_2) facilitates the emergence of the Topic meaning. Finally, the dynamic aspect of the formulation aims at anticipating the assembly of the unit with the neighbouring ones. We enlarge on this last point now.

2.4. An instruction at work: the processing of the instructions of an utterance

The processing of a complete utterance follows the same intructionnal principles: all the units are processed according to their instructions in the unfolding of discourse and when they are perceived. More precisely, Col et al (2012: 161) suggests four main principles controlling the processing of the utterance, and above all, controlling the order of processing: 1) the units are taken into account one after the other as soon as the sentence is perceived; 2) once they are taken into account, and if it is possible for them to be processed, they are processed at once in conformity with their instruction. Thus, each unit plays its evocation role as soon as the elements it convokes are available; 3) if they cannot be processed, they are left pending until the elements they convoke are introduced on the scene that is being built; 4) once evoked, the elements of the scene continue to be determined or even transformed by later convocations by other units.
Let's take a simple example so as to illustrate the whole process:

Britain ***1*** *consum* ***2*** *ed* ***3*** *over* ***7*** *9* ***5*** *million* ***4*** *tons of paper* ***6*** *.*

Once perceived, "Britain" can be processed rather rapidly. It evokes an entity identified as a proper name referring to a country. Nothing about its status as an argument or another syntactic function is supplied yet; it is just an entity directly perceived on the verbal scene. The same could be more or less said about "consumed". Concerning this unit, we should make a distinction between the lexeme and the morpheme <–ed>. "Consume" immediately evokes a process – there is no noun corresponding to this lexeme in the English lexicon. But the <–ed> morpheme may evoke at least two different elements: either a process located in the past time (or located on a different level from the actual one), hence the interpretation of <-ed> as a preterite, or an accomplished process, hence the alternative interpretation of <-ed> as the past participle. But, since an entity has been introduced on the scene (Britain) and as this entity has no real function yet, the <–ed> morpheme is processed as a marker of the preterit and at the same time, Britain acquires a syntactic function: subject of "consumed". Both elements are determined simultaneously in the course of their interaction and of their perception as a "good form", in a gestalt way.
The rest of the processing is more complex. Actually, when "over" is perceived, it cannot be processed because at this stage of the processing, it is not recognisable as a preposition or an adverb (nothing is associated with it except a verb and a subject), nor as a verb particle as "consume over" is not a phrasal verb. "Over" is then left pending until other units are processed. The following unit, "9" cannot be processed either: "Britain consumed over 9" is barely understandable and needs the expression of a quantity to gain some meaning. "Million" is the next unit to be processed as "9" is left pending and together, they evoke the quantity required by the process "Britain consumed". We may say that 9 *attracts* this quantity as soon as it is perceived. As for the object of the quantity ("tons of paper "[8]), it is rapidly processed insofar as a quantity plus a process are present on the scene and they can assemble.
At the stage of the processing, "over" is finally processed. This stage is a double one in fact: "over" is processed, i.e. a semantic trajectory is given, as it were, and at the same time, this unit is categorised, namely as an adverb thanks to the instruction supplied both by the verb and the nominal complex evoking a quantified entity. "Over" is the last unit to be processed and it evokes something which goes beyond the movement of covering supplied by its instruction. In the interaction with the rest of the utterance, its contribution is larger than identifying the meaning of a unit: it finalizes, at least tentatively, the semantic gestalt construed by the whole utterance.

III. Continuity in language modelling

Since the availability of numeric methods and very large corpora, lot of progress has been made in the automatic exploration of lexical semantics. Most approaches rely on distributional methods, based on a

hypothesis first formulated by Harris in 1954. This hypothesis states that words appearing in the same context tend to have a similar meaning (the same tradition was popular in linguistics, cf. Firth (1957): "you shall know a word by the company it keeps"). The number of papers published on distributional methods is nowadays so large that it is hard to give a comprehensive picture of the field. However, most approaches so far were considering isolated words to calculate word similarity (for information retrieval or information extraction applications for example). A large effort has been made recently to go beyond the word barrier and take into account larger phrases and more generally the syntactic context so that distributional methods can cope with the word limitation problem.

The hypothesis that makes it possible to take into consideration sequences more complex that isolated word is the notion of compositionality, often attributed to Frege (1892). Compositionality means that the semantics of a sequence can be calculated from the semantics of its parts. In other words, it would be possible to combine the semantics of the parts to compute the semantics of a complex sequence. This is supposed to explain why humans can understand sentences they have never seen before: by combining the meaning of isolated words, we are able to compute the meaning of larger sequences even if we have never seen these sequences before.

These methods are powerful and seem well fitted for language modelling. We will not detail here the mathematical basis for these models, which is can be quite complex. We think it is more relevant to focus on different features that make these models interesting from a linguistic point of view.

1. Thanks to distributional methods, it is not any more necessary to define in advance what is the meaning of a word. Distributional methods take literally the principle that words are defined by the contexts in which they appear, which lead to representations that are quite different from those of existing human dictionaries. Dictionary induction is a task examining to what extent the meaning of a word can be induced only by looking the context: results obtained with this approach are often interesting, sometimes debatable and at least question existing lexical descriptions.
2. The description obtained by such methods is consistent with the notion of "continuity" (Fuchs and Victorri 1994). One does not need to a priori define clear cut word senses but different descriptions, with a different granularity, can be obtained depending on how precise the context taken into account is. It is thus possible to define more or less fine grained word senses for a same word, and propose semantic maps showing that certain senses are closer than others (see Figure 1).

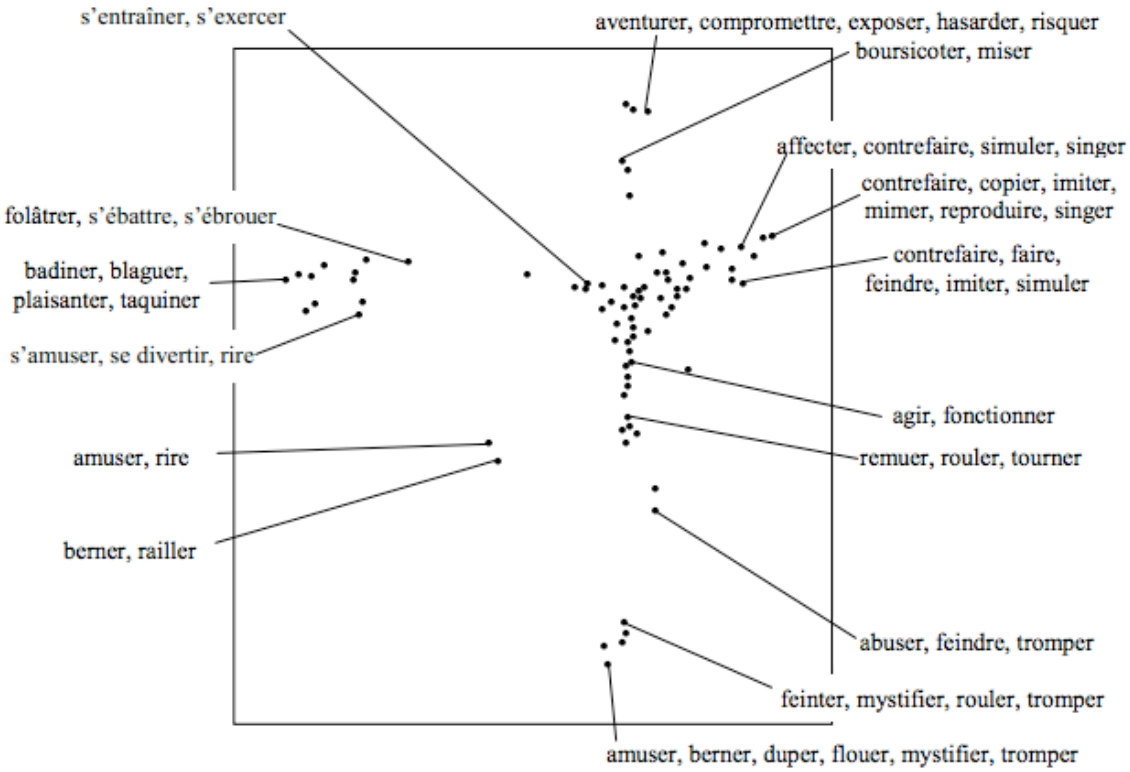

Figure 1. The semantic space associated with the French verb jouer, from Gaume et al. 2005

The techniques used are also able to compare different words so as to automatically identify synonyms or other related words, identify regular patterns of meaning changes (metonymy, metaphors, etc.), cf. Shutova *et al.*, 2012.

3. When coupled with a way to deal with compositionality, distributional methods can be extended so as to calculate the meaning of phrases or even sentences (however note that most models are limited due to the complexity entailed by pure distributional models. An open research question consists in extending these methods using tractable models). The advantage of applying these methods beyond the word limit is to represent using a uniform mechanism notion such as idioms, compounds or multi-word expressions. An idiom can be seen as a regular phrase with limited possibilities of variations, which can be to a large extent characterised and calculated automatically using statistical and / or distributional methods.

The complexity of nowadays techniques is high since distributional methods require to calculate the co-occurrences of words and phrases through very large corpora (sometimes gigabytes of data, several billions words). Since the current trend is to go beyond the word limit, there is potentially an infinite number of sequences to take into account, which quickly lead to combinatorial explosion and other complexity problems.

Different proposals have recently been made to implement this kind of models. A first approach consists in generalizing the use of matrices (that can represent the co-occurrences of a series of words) through tensors (Van de Cruys, 2013). The key idea is that compositionality is modelled as a multi-way interaction between latent factors, which are automatically constructed from corpus data. The composition component can itself be based on a formal description of interaction between linguistic items (Grefenstette et al. 2014). Another very popular way of dealing with the problem is to use multi layered neural network, a method known as "deep learning", widely used initially in speech recognition and nowadays more generally but also in natural language processing (Sarakiya et al., 2014; Socher et al. 2013). It seems that there is thus a real convergence between automatic methods and linguistic description here.

IV. Conclusion

In this paper, we have considered the notion of continuity and interaction in linguistic models. There is now a quite rich tradition of work based on the hypothesis that natural language is not a discrete model. Instead, continuous models consider word sense, grammar rules and categories as continuous notions: some words are hard to categorize and some rules do not fully apply in certain contexts. Word meaning in context often correspond to different meanings according to the dictionary, etc.
New models have been proposed that help to provide a more accurate view of these phenomena. Recent advances in natural language processing also support this idea by providing very rich models based on a multi-level representation of the context of use of linguistic items. These models have been applied to different problems (from lexical categories to grammar rules, but also to more complex tasks such as information extraction or machine translation). From this point of view, we can see an interesting convergence between linguistic and computational models.
There is anyway still a long way to go in this domain. The models considered here are just in their infancy, especially for linguistic descriptions.

References

Bar-Hillel Y., «Communication and argumentation in pragmatic languages [1968]», in Aa. Vv., *Linguaggi nella società e nella tecnica*, Milano, Comunità, 1970.
Barsalou WL (1999). « Perceptual symbol systems ». *Behavioral and brain sciences* 22, 577-660
Barsalou WL (2008). « Grounded cognition ». *Annual Review of Psychology*, 59:617-45
Barsalou WL (2010) « Grounded Cognition : Past, Present, and Future ». *Topics in Cognitive Science*, 2:716–724

Benveniste E., *Problèmes de linguistique générale*, Paris, Gallimard 1966.
Col G., Aptekman J., Girault S., Victorri B., « Compositionnalité gestaltiste et construction du sens par instructions dynamiques », *CogniTextes* [En ligne], Volume 5 | 2010, mis en ligne le 11 juillet 2010, Consulté le 10 mai 2014. URL : http://cognitextes.revues.org/372
Col, G. et Poibeau, T. (2014). « An instruction-based analysis of *over* », accepted in *Language and Cognition*.
Col, G., Aptekman J., Girault S. et Poibeau, T. (2012). « Gestalt Compositionality and Instruction-Based Meaning Construction », *Cognitive Processing*, volume 13, numéro 2, pages 151-170.

Col, Gilles (2011), « Modèle instructionnel du rôle des unités linguistiques dans la construction dynamique du sens », in Le Langage et ses niveaux d'analyses. Jean Chuquet (dir). Rennes, Presses Universitaires, 45- 60.
De Beaugrande R. A., « Text, discourse and Process. Toward a Multidisciplinary Science of Texts », *Advances in discourse processes*, vol. XI, Norwood (New Jersey), Ablex Publishing Corporation, 1980.
De Beaugrande R. A., 1996, « New foundations for a science of text and discourse. Cognition, communication and the freedom of access to knowledge and society », *Advances in discourse processes*, vol. LXI, Norwood (New Jersey), Ablex Publishing Corporation, 1996.
Ducrot O. (1984) *Le Dire et le dit.* Editions de Minuit, Paris
Ducrot, O. (1972). *Dire et ne pas dire. Principe de sémantique linguistique*. Paris, Herman.
Fauconnier G (1997) *Mappings in Thought and Language*. Cambridge University Press, Cambridge
Firth, John Rupert. 1957. « A synopsis of linguistic theory 1930-55 ». In *Selected papers of J.R. Firth (1952-59)*. London: Longmans. 168-205.

Frege Gottlob. 1892. Uber Sinn und Bedeutung. ¨Zeitschrift fur Philosophie und philosophische Kritik,100:25–50.
Fuchs C., and Victorri B. (eds). 1994. Continuity in linguistic semantics. Amsterdam ; Philadelphia : J. Benjamins, c1994.
Gaume Bruno, Venant Fabienne, Victorri Bernard. 2005. Hierarchy in lexical organisation of natural languages. Dans : Hierarchy in Natural and social Sciences. Denise Pumain (Eds.), Springer, p. 121-143,

Vol. vol. 3, Methodes series.
Girault, S. et Victorri, B. 2009. "Linguistiques de corpus et mathématiques du continu", *Histoire, Epistémologie, Langage*, tome XXXI, fascicule 1, 147-170
Grefenstette Edward, Sadrzadeh Mehrnoosh, Clark Stephen, Coecke Bob and Pulman Stephen. 2014. Concrete Sentence Spaces for Compositional Distributional Models of Meaning. In *Computing Meaning, Volume 4*, pp.71-86, H. Bunt, J. Bos and S. Pulman (eds), Springer, 2014
Guillaume, P., *La Psychologie de la forme*, Paris, Flammarion, 1979.
Harder, P. (2009). « Meaning as input : The instructional perspective », in V. Evans et S. Pourcel (dir), *New Directions in Cognitive Linguistics*, 15-26

Harris Zellig S.. 1954. Distributional structure. Word, 10(23):146–162.
Harris, Z. (1954). « Transfer grammar », *International Journal of American Linguistics* 20, num. 4, 259-270

Kanizsa, G., *La teoria della gestalt: distorisioni e fraintendimenti*, in Kanizsa, G., & Legrenzi, P., (a cura di), *Psicologia della gestalt e psicologia cognitivista*. Bologna: Il Mulino, 1978, pp. 39-61.
Köhler, W., *Psychologie de la forme. Introduction à des nouveaux concepts en psychologie*, Paris, Gallimard, 2000.
Manning, Christopher (2003). «Probabilistic Syntax», Rens Bod, Jennifer Hay & Stefanie Jannedy (éd.), *Probabilistic Linguistics*, MIT Press, 289-341.
Maturana H. R., Varela F. J., *Autopoiesis and Cognition. The Realization of the Living*, D. Reidel Publishing Company, Dordrecht: Holland / Boston : U.S.A. / London : England, 1980 [ed. or. *De Maquinas y seres vivos*, Editorial Universitaria S. A., Chile, 1972].
Nemo, F. (2001). « Pour une approche indexicale (et non procédurale) des instructions sémantiques. », *Revue de Sémantique et de Pragmatique* 9-10, 195-218.
Nølke, H. (1994). *Linguistique modulaire : de la forme au sens*. Bibilothèque de l'Information Grammaticale, Paris-Louvain, Peeters

Oller J. W., *Language tests at school: a pragmatic approach*, London-New York, Longman, 1979.
Petitot-Cocorda, J. >> *Morfogenesi del senso: per uno schematismo della struttura*, Milano, Bompiani, 1990.
Rastier, F. – Bouquet, S. (a cura di), *Une introduction aux sciences de la culture*, Paris, PUF, 2002.
Rastier, F., *Anthropologie linguistique et sémiotique des cultures*, in F. Rastier, S. Bouquet, *Une introduction aux sciences de la culture*, Paris, PUF, 2002, pp. 243-267.
Rastier, F., *Du signe aux plans du langage*, in "Cahiers Ferdinand de Saussure", 54/2001, pp. 177-200.
Rastier, F., *Formes sémantiques et textualité*, "Langages", 163/2006, pp. 99-114.
Rastier, F., *Sémantique et recherches cognitives*, 3ª ed., Paris, PUF, 2010.
Rastier, F., *Sémantique interprétative*, Paris, PUF, [1987] 2009.
Rastier, F., *Sens et textualité*, Paris, Hachette, 1989.
Sarikaya, R., Hinton, G. E. and Deoras, A. 2014. Application of Deep Belief Networks for Natural Language Understanding. IEEE Transactions on Audio, Speech and Language Processing.
Saussure, F. de, *Cours de linguistique générale*, C. Bally et A. Séchehaye éds., avec la collaboration de A. Riedlinger, Paris, Payot, [1916] 1922; tr. it., introduction et notes de T. De Mauro, *Corso di linguistica generale*, Roma-Bari, Laterza, [1967] 2003
Saussure, F. de, *Écrits de linguistique générale*, Paris, Gallimard, 2002.
Schmidt S., *Texttheorie*, München, Fink 1976; tr. it. par S. Muscas, *Teoria del testo*, Bologna, Il Mulino, 1982.
Shutova Ekaterina, Van de Cruys Tim and Korhonen Anna. 2012. *Unsupervised Metaphor Paraphrasing Using a Vector Space Model*, In Proceedings of COLING 2012, Mumbai, India.
Socher Richard, Manning Chris and Bengio Yoshua. 2013. Deep Learning for NLP (without Magic). Tutorial, Association of Computational Linguistics 2013 (http://www.socher.org/index.php/DeepLearningTutorial/DeepLearningTutorial).
Thom, R. 1980 *Stabilità strutturale e morfogenesi. Saggio di una teoria generale dei modelli*, Einaudi, Torino.
Thom, R. 1985 *Modelli matematici della morfogenesi*, Einaudi, Torino.
Van de Cruys, Tim, Poibeau, Thierry and Korhonen, Anna. 2013. A Tensor-based Factorization Model of Semantic Compositionality. Proceedings Human Language Technologies: Conference of the North

American Chapter of the Association of Computational Linguistics, Atlanta, Georgia, pages 1142-1151.
Victorri B., Fuchs C., *La polysémie: construction dynamique du sens*, Paris, Hermès, 1996.

[1] May I have explicit references??
[2] May I have explicit references??
[3] «Il n'y a dans la langue ni signes [dans le sens de "signifiants" – N.d.A.], ni significations , mais des *différences* de signes et des *différences* de significations : lesquelles 1° n'existent les unes absolument que par les autres, (dans les deux sens), et sont donc inséparables et solidaires ; mais 2° n'arrivent à se correspondre directement» (Saussure 2002: 70). Cf. also CLG, part II, chapp. III, IV.
[4] "La manière la plus simple d'éluder la question consiste à considérer le texte comme un signe. C'est la solution que choisissent Peirce, comme Greimas ou Eco (*cf*. Eco 1988, p. 32 : "le Message équivaut au Signe"). Cette esquive fait évidemment peu de cas de la différence de niveau de complexité entre le signe et le texte, mais surtout empêche de penser l'incidence du global sur le local, en l'occurrence du texte sur chacun des signes qui le composent" (Rastier 1997).
[5] Weinrich adopts the classical communicative model proposed by Shannon e Weaver (1952).
[6] As we can read in the italian translation of this essay, «un lessema si può concepire teoricamente come una *regola* (*in senso ampio*) o un'*istruzione* per la produzione di un dato "comportamento" verbale e/o non verbale… Il campo-contesto [il campo lessematico] assegna al lessema le sue possibilità generali di funzionamento nei testi» (Schmidt 1976: 56).
[7] In most studies, *over* has a temporal meaning in this kind of examples but it should be discussed if this is necessary, since the temporal meaning is mostly supported by the noun following *over*.
[8] We do not describe the full processing of this nonimal complex but various elements at different levels like the morpheme (-s) or preposotions (of) are to de distinguished.