Verbal Polysemy in Automatic Annotation

Maryvonne Abraham
GET-ENST-Bretagne ,
CS 83818, F29238 Brest cedex 3 France
LaLICC , Paris-Sorbonne University,
Maison de la Recherche, 2 3 rue Serpente, 750105 Paris, France
Maryvonne.Abraham@enst-bretagne.fr

Abstract

The linguistic theory of Applicative and Cognitive Grammar analyses the language in three levels as follows: the linguistic level, the predicative level, and the semantico-cognitive level.
The meaning of the words is described at the semantico-cognitive level. Here we give the description of the verbal semantics in order to build a lemmatized lexicon of polysemic words.
On the one hand, the contextual exploration of texts is based on a linguistic context made up of muted segments of sentences without grammatical analysis. In this paper, we attempt to justify this way of processing by describing the characteristics of the arguments of the parameters of the verb.

The problem: how do we read a text ?

When we read a text, how do we gain access to the meaning? What are the processes which allow us to read a text and understand its meaning?
In order to describe the meaning of a text, we break it up into smaller elements. Text has to be distinguished from the words of the text. The text is available as a syntactic structure that we can analyse from a (formal[1], categorial) grammar. The words are often polysemic. That is the reason why we try to break them up into smaller elements that we call primitives. We want to ensure a cognitive foundation to these primitives.
It is interesting to describe the polysemy of a word. To do this, we try to organize the different significations of a word into a net, then we search for an invariant, using an abductive method [Figure 1]. This invariant does not belong to the language, but to our mental organization, and is seen as a mental representation of the word[2].

[1] [Chomsky, 1979, 1981], [Jackendoff, R., 1978, 1983,1987], [Jacobson, R.].
[2] For [Desclés, 1990], [Abraham, 1995], [this invariant is a « cognitive archetype ». For Picoche, 1986], it is a « signifié de puissance ».

However, how do we build the meaning of a text from its components? Contextual exploration processes using textual indices. It does not use a lemmatized lexicon, but seeks flexed and conjugated words at the observable level of a text, voices of verbs, expressions dedicated to a domain, etc... The reading processing does not work by decomposing and recomposing the text and the words, but it builds the meaning of a text from the conjugated words seen as compiled knowledge: the transformations applied to the words indicate a part of the role of the words in a sentence.
Decomposing and recomposing the meaning are not symmetrical processes.

The problem of polysemy

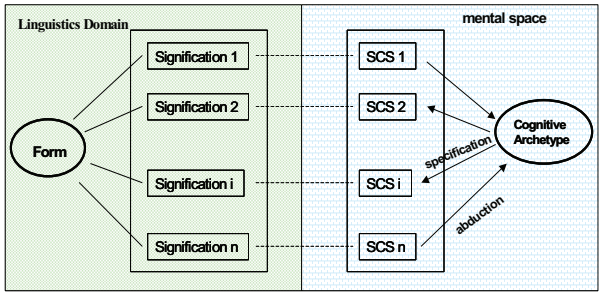

Figure 1 : From the form to the formal description of polysemy

We focus on the description of the verbal lexicon[3], at a semantico-cognitive level. Many significations can correspond to a single entry form of the lexicon. We describe them using a semantico-cognitive scheme

[3] The method to describe the lexicon is given in [Abraham 1995], [Descles, 187,1990], [Djioua, 2000], [Bogacki, 1983]. This method can be compared to [Pustejovky, 1991,1995].

language of representation[1] (SCSLR). Let us recall the main cognitive domains which describe the world: they are static and evolutive domains. From these notions, we can describe the significance of verbs: in order to do that, we define structuring primitives based on Static, Kinetic and Dynamic domains. Empirical primitives will allow us to complete descriptions of the world. We can then represent the significance of a verb by a combination of semantic primitives. The static part of the verb describes permanent relations of the SCS, the type of arguments; the kinetic part describes the modification of *y*, as in the example *La branche casse*[2] ; a dynamic part of the SCS can be applied to the kinematic scheme introducing an agent which can receive different degrees of agentivity for *x* , for example :
Le vent casse la branche. (the wind breaks the branch)
Jean casse la branche. (John breaks the branch)
To resume the problem, the verb is an open structure, combining static, kinematic and dynamic relations, which expects one or two places of arguments[3]. In French, it is necessary to introduce an additional place through a preposition.
A same verbal item can express static, kinematic or dynamic situations:

- The static part describes relations which are permanent during the spatio-temporal transformation
- The kinematic part describes the transformation which can act on the entity, its localization, its state, its properties, its attributes, its semantic type
- The dynamic part can indicate the degree of agentivity of an agent or of an instrument responsible for the transformation.

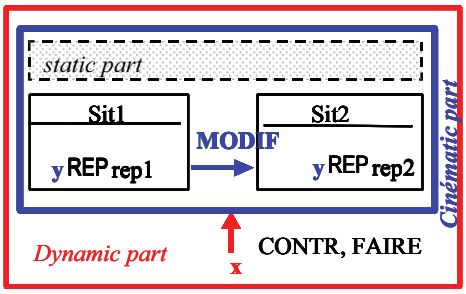

Figure 2 : structure of a SCS

Let us give a few kinematic examples which describe the modification of an agent:

- Its spatial movement : *les pommes tombent* (apples fall)
- Its state : les pommes mûrissent (apples ripen)
- Its properties : *couper une pomme* (cut an apple : the apple changes state from a whole state to a state in pieces)
- Its attributes : *peindre un mur* (paint a wall : the attribute of color is changed)
- Its semantic type : *commencer un livre* (to begin a book) : the book changes its semantic type from object to activity , which can be reading, writing, tearing, etc….

The following examples are dynamic schemes :

- *Le soleil mûrit les fruits* (the sun ripens fruit) the sun has an action without any control , which is denoted FAIRE applied to a change of state.
- *Jean peint le mur, coupe une pomme, cueille les pommes* (John paints the wall, cuts an apple, picks apples) : here, Jean has control over his action , which is noted CONTRoFAIRE .

But if *Jean marche* (John walks) is dynamic, *Jean tombe* (John falls) is only kinematic : Jean has no control over his movement when he falls. There would be control in the example *Jean se laisse tomber*, (John drops to the floor) or *Jean fait tomber x (John makes x fall)* . We see that the feature "animated" of the syntactic subject of a verb does not imply that the verbal scheme is dynamic; there are verbs which are only kinematic .
It is the kinematic part of the verb which expresses the modification (of its parameter) which is described by the verb. The kinematic part describes a modification which can affect the location, one or several properties of the attributes, the state and the type. The transformation depends on the domain, (spatial or not), on types, properties and attributes of the parameters. The transformation can be spatial or not spatial. It is not a reduction to spatial, but it is clear that the spatial can be represented more directly. It is more visual; it is easier to be perceived.

Contextual exploration

We described above how to represent the different significations of a same polysemic verb. The formalism of SCS allows us to build a lemmatized lexicon in which many significances can match a same entry form (being described by SCSs). We shall now see how we find the appropriate significance in the context.
When we read a text, we do not process following morphosyntactic, grammatical and semantic analysis as proposed in most conventional automatic natural language processing systems. We directly seek the flexed forms and their places in the sentence, surface marks resulting from grammatical operations. The system of contextual exploration copies this way of reading by using a context of

[1] SCS are formal structures which can describe the semantics of words indicating transformation.
[2] We are describing the French language, so our examples are given in French. The translation into English is : *the branch is broken*.
[3] [Abraham,M.Y., 1995], [Desclés, J.P., 1987,1990]

lists of conjugated words, without processing a morpho-syntactic analysis nor using a lemmatized lexicon.
From the lexicon point of view, this lexicon contains polysemic lemmatized words and entry forms for their morpho-syntactic, semantic and other descriptions. We describe the intrinsic polysemy of the verbal lexicon with SCSs, combinations of semantic primitives, which allow us to receive different places of actant parameters which are specified by their attributes, types and properties. In context, the concordance of attributes, types and properties of parameters with possible significations of a same verb should allow us to select the pertinent signification. But more information is needed to access this signification. The attributes, types and properties of the actants expected by a verb can constrain the aspectual forms in order to guarantee the compatibility of the resulting predicate with a larger context. In the sentence, aspecto-temporal and enonciative operations apply morpho-syntactic transformation to the instantiated verbal predicate and place the result into a context with respect to a grammaticalized semantics which is compatible with the lexical semantics of its parameters. So, a good detailed context allow us to choose one of the values of a polysemic verb.
We propose to illustrate this mechanism between lexicon and context with a few examples.

Seeking good indexes for contextual exploration

In contextual exploration, the verb is found as conjugated (tense, person) with a specific voice, in a precise structure of the sentence. This structure gives the roles of actance (subject, complement) with the position in the sentence; in addition, these actancial roles verify a compatibility of typing, control properties of the parameter. They also verify a compatibility of typing, properties and attributes. At possibly, a locative, or cultural precision is added, as in the example :
Les menhirs avancent vers la mer à Noël
(menhirs move toward the sea at Christmas)
where, ontologically, the menhirs do not usually have properties of movement ; in the Breton culture, it is well known that they receive this property at certain periods of the year.
To return to the verbal organization, we can now ask what is the organization around an invariant; how can we use the dictionary , the attributes and properties of the parameters of a verb to choose the "correct" signification? How do we read and how do we understand the signification of a verb?
To answer this question, let us see the criteria which structure the polysemy of the verb. Contextual criteria, but which context? The verb, an open structure, expects parameters (actants) ; their position in the sentence indicates the interpretative roles. The semantic typing, properties and attributes of the verb are taken into account in the modification that the verb describes. The kinematic parameter – the one which is modified – must have attributes or properties interested in modification in time. The entity which is moving must have properties of moving, with or without control. We note that a "pure" location has ontologically no property of movement.
We can also notice an analogy between activity and location in the uses of the verb. A spatial location is continuous; a location of activity can be considered following successive steps in time (ingredience).
The following example only describes a property of the argument of the French verb "circuler" (to circulate) :

- *Le sang circule* (blood circulates)
- *Les voitures circulent* (cars run)
- *Les planètes circulent* (planets revolve)
- *Les informations circulent* (information goes round)

If we build a context for these examples , we must consider the location of the transformation indicated by the verb "*circuler*". This location must be compatible with the properties of the argument.

- *Le sang circule dans le corps humain* (if we say : « on the road », we must invent a context of a terrible accident to accept the sentence , i.e., change an ordinary context)
- *Les voitures circulent sur la route, dans les chemins, dans la montagne,* (cars run on the road, in lanes, in the mountains)
- *Les planètes circulent autour du soleil* (planets revolve round the sun)
- *Les informations circulent dans les journaux* (information is spread in newspapers)

It is possible that the parameter of the verb does not have the properties which are normally expected by the verb. In this case, the language uses the inheritance of the property of an attribute. We denote MVT the property of movement, MVTh if this property is inherited, MVTa if the property is auto-controlled, MVTs if the movement is externally controlled.

exemple	Propriété
Le Pape s'envole The Pope flies off	MVTh
Le moulin tourne The mill turns	MVTh
La pendule avance The clock is fast	MVTh
Le ministre roule (en voiture) The minister travels (by car)	MVTh (inheritance of a property of a containing entity)
La voiture décide de tourner The car decides to turn	CONTRh
Le tonneau fuit, coule, déborde The barrel leaks, leaks, runs over	MVTh

La maison surgit derrière le tournant The house appears after the bend	Inheritance of the perception of the entity responsible of statement

In French, there are many rules of inheritance, called "metonymy" , but we prefer to detail them in our theory:

•rule I-1 of inheritance for the relation of ingredience
if the property P1 belongs to an ingredient of y, y can inherit the property of this ingredient in the language ; it is lexicalized by the same verb, but with a derived signification which catches the relation of ingredience in its signification.
For example, we have :
la roue roule, (the wheel rolls)
the wheel is an ingredient of "car"
then *la voiture roule*, (the car is running) : in French the same verb is used, but it describes a movement which differs from that of the wheel.
Let us take another example :
les aiguilles (de la pendule) avancent,
the hands of the clock are fast
the hands are ingredients of the clock
then *la pendule avance*. the clock is fast.

So :
if x is localized in the interior of an object y, if z is an ingredient of y, then z is an ingredient of x.

rule I-2 of inheritance for the relation of attribution
if the property P1 belongs to an attribute of y , y can, in the language inherit the property of its attribute, which is expressed by the same lexeme : the verb which is used is the same, but with, for entity y, a derived signification, which interpretes the relation of attribution as a specification. For example, *Marie rougit* (Mary goes red).

rule I-3 of inheritance for the relation of localization
In the language, if y is localized from a moving entity, then y can inherit the property of movement[1], of its localizer. For example *Madame Dupont roule* because we have a relation of localization of Madame Dupont inside the location "*voiture*" (she is inside a car) and *voiture* (car) has the property of movement.

rule I-4 of inheritance for the propery of movement
In the language, if y contains an entity which is moving, then y can inherit the property of movement of its contents. For instance, in *le tonneau fuit, coule, déborde, ...* (the barrel flows) the language gives the barrel the property of its fluid contents.

rule I-5 expressing perception of the entity responsible for the statement
In the language, a verb can apply to an entity having a semantic type which is "impossible" in an ordinary context, that is, a type which does not match an expected type. In this case, the verb expresses a particular perception of the entity responsible for a statement. For example, *la maison surgit derrière le tournant* expresses the perception of the entity responsible for the statement.

[1] A pure location has no movment property.

Some signification of a polysemic verb *avancer*

The verb *avancer* (to move forward, to advance) indicates a movement, but which entity is moving? In *le train avance*, (the train is moving forward) there is no doubt, the train is moving (forward) , but, in *la montre avance*, (the watch is fast) , the watch does not move, only its hands move (forward). SCSs allow us to separate the significance of *avancer* between *le train avance* (the verb is kinematic, written : cin) and *Jean avance* (the verb is dynamic, written : dyn). The train and Jean have the semantic type J, which indicates an individual entity (countable) , the train has a property of movement MCTs without control, and Jean has a property of movement MVTa , of self-control : the verb *avancer* can be applied to Jean and to the train, for these entities can move.
The compatibility with a cultural context can take place too, as in the example *Les menhirs avancent dans la mer à Noël* .

example	properties
Le train avance	cin, J,no control MVTs
Jean avance	dyn, J, self-control MVTa
Le travail avance Work is progressing	cin, activity, no control MVTa
Les aiguilles de la montre avancent	cin, J,no control MVTs
La montre avance The watch is fast	cin, J, no control MVTh , inheritance of an ingredient
Les menhirs avancent dans la mer à Noël	cin, J,no control MVT

There are many other uses of the verb *avancer*. In certain cases, many interpretations can exist at one and the same time: a*vancer la réunion* (to bring forward the meeting) can express *avancer l'heure de la reunion* (to bring forward the time of the meeting) where *réunion* inherits the property of movement of its attributes *heure* (time), or *avancer le travail qui se fait dans la réunion* (to make the work (undertaken in the meeting(progress) , where this time the type of *réunion* is activity, as in *la réunion avance* (the meeting progresses).

Change of type of an entity

Let be the entity *livre* (book), which for us has the type individal entity. Certain words in the language can be applied to this entity to change its semantic type.

Operator	*example*	Type of *livre*
The verb	*Commencer , finir un livre* Begin ,finish a book	activity
a demonstrative	*Ils ont lu ce livre*	information

	They have read this book	
a preposition	*Dans, sur le livre* In, on the book	location
a determinant	*Les, des, plusieurs, leurs, livres* The, some, several, their books	PL
A classificator	*Une tonne, un paquet, une pile de livres* A ton, a pile	massive

Actancial roles

The same verb can receive many actancial constructions, which are partially indicated from its syntactic construction, and also from the knowledge of the properties of the actants.

Constructions of the verb	
X avance X goes forward	Without or with or CONTR
X s'avance X goes forward	ACONTR
X avance Y X makes y go forward	ECONTR or FAIRE

Knowledge of the world linking up X,Y are :

- type and properties of X (movement, self-controlled or with external control) :
- attributes or parts of X
- type and properties of Y (movement, inside, …) ,
- attributes or parts of Y
- semantic binding between X and Y

How do we understand a text?

Applicative and cognitive grammar analyses the language following three levels : an observable linguistic level (the one which is explored by the contextual exploration) ; a predicative level, where grammatical operations are formally expressed; a semantico-cognitive level, which describe the significance of the verbs. On the one hand, we have described verbal semantics in order to build a referential lemmatized lexicon, in which we try to explain polysemy through a cognitive archetype. On the other hand, we have described the process of contextual exploration which does not use the lemmatized lexicon, but seeks the use of the conjugated verb contextualized with its actants.

In this case, we question whether it is useful to describe the verbal semantics and the verbal lexicon. We have seen here how the kinematic part of the verb expresses the transformation that the verb applies to its parameter. In the contextual exploration, it is the result of this transformation, as compiled knowledge, which is directly taken into account. The marks from the grammatical operations give the interpretative roles of the flexed parameters which have their semantic type, properties and attributes which point to the correct meaning. The description of the semantics of the verbal lexicon which expects parameters with type and properties expected by the verb explain why the contextual exploration can process : it is based on rebuilding the verbal segment processed on the second level of contextual exploration in Applicative and Cognitive Grammar.

The description of the lexicon which is presented in Applicative and cognitive grammar bring to the fore elements used for the contextual exploration. If a same polysemic verb expresses different transformations, what has to be annotated in order to choose one among the significations?

- The interpretative role is given by the syntax;
- The place as a subject position has or receives agentive properties ;
- The kinematic parameter has the properties expected by the verb; they come from the ontology, or from inheritance or from a type change;
- The type gives properties
- An extra linguistic context (or an ontology) describes the locations which are compatible with the verb
- These locations can be spatial, location of activity, or notional following the type and properties of the parameter located in relation to this location.
- We can see that we are able to give the signification of words , and to use them in the correct context. This means that we have both knowledge of the words and of the context , and of the way to put these words into a sentence to build meaning.

References

Abraham, M.Y., 1995 *Analyse sémantico-cognitive des verbes de mouvement et d'activité; contribution méthodologique à la constitution d'un dictionnaire informatique des verbes.,* Ph.D., Ecole des Hautes Etudes en Sciences Sociales, Paris.

Bogacki, K., Bogustawska, G., Kleisberg, A., Lewicka, H., Tozinska, M., Thieme, M., Wojcik, L., and Zielinska, T. 1983, *Dictionnaire sémantique et syntaxique des verbes français*, Panslwowe Wydownictwo Naukowe.

Chomsky, N., *Structures syntaxiques.* Point n°98, Paris, 1979.

Chomsky, N., *Réflexions sur le langage*, Champs Flammarion, Paris, 1981.

Desclés, J.-P., « Réseaux sémantiques : la nature logique et linguistique des relateurs », *Langages* , n° 87, 1987, 57 -78.

Desclés, J.-P., 1990, Langages applicatifs, langues naturelles et cognition, Paris : Hermès.

Fauconnier, G., 1984, *Les espaces mentaux*, Editions du Seuil, Paris.

Djioua, B., 2000, *Modélisation Informatique d'une base de connaissances lexicales (DiSSC): Réseaux polysémiques et Schèmes Sémantico-Cognitifs*, Ph.D., sous la direction de Jean-Pierre DESCLES Paris-4-Sorbonne.

Givon, and Talmy , 1979, *On understanding Grammar*, Academic Press.
Jackendoff, R., 1978, "Grammar as an evidence for Conceptual Structure" in Morris Halle, Joan Bresnan, George Miller, (eds), *Linguistic Theory and Reality*, Cambridge, Mass., MIT Press, 201-228
Jackendoff, R., 1983, *Semantics and Cognition*, Cambridge, Mass : MIT Press.
Jackendoff, R., 1987, *Consciousness and the Computational Mind*, Cambridge, Mass. : Bradford/MIT Press.
Jacobson, R., 1971, "Boas"s View of Grammatical Meaning" *Selected Writings of Roman Jacobson*, Vol. 2, 489-496.
Lakoff, G., 1987, Women, Fire, and Dangerous Things : What Categories Reveal about the Mind, Chicago/ London : University of Chicago Press
Langacker, R., 1987, *Foundations of Cognitive Grammar*, Vol 1: *Theoritical Perequisites*, Stanford, Calf. : Stanford University Press.
Langacker, R., 1991, *Foundations of Cognitive Grammar*, Vol 2 : Descriptive Application, Stanford, Calf.: Stanford University Press.
Picoche J., 1986, *Structures sémantiques du lexique français*, Nathan, France.
Pottier, B., 1974, *Linguistique générale*. *Théorie et description,* Paris.
Pottier, B., 1987, *Théorie et analyse en linguistique*, Paris.
Pottier, B., 1993, *Sémantique générale*, Presses Universitaires de France.
Pustejovsky J., 1991, *The generative lexicon*. Computational Linguistics, vol. 17, n° 4, 1991.
Pustejovsky, J., 1995, *The generative lexicon.* MIT Press, Cambridge, Ma.
Talmy, L., 1983, "How Language Structures Space". *Spatial Orientation / Theory, Research, and Applications*, ed. by Herbert Pick and Linda Acredolo, 225-282, New York : Plemnum Press.
Vandeloise, C., 1986, *L'espace en français*, Edition du Seuil, Paris.