

## LES CLASSES D'OBJETS

GASTON GROSS

## RÉSUMÉ

*La sémantique a été pendant longtemps tenue en suspicion dans la tradition harrissienne, sous prétexte qu'elle ne permet pas une description rigoureuse des faits linguistiques, comme le fait la syntaxe. À vrai dire, cet ostracisme concernait une sémantique autonome, détachée des structures syntaxiques. La situation change dès lors qu'on ne postule pas des niveaux différents, mais qu'on prend comme unité minimale d'analyse non pas le mot, mais la phrase. Le but de cet article est de montrer que les contraintes syntaxiques permettent de mettre au point des classes sémantiques, appelées classes d'objets, qui constituent des ensembles disjoints, ayant un caractère aussi reproductible que les descriptions syntaxiques. Ces classes permettent de séparer de façon précise les différents emplois des prédicats, de proposer des synonymes et des antonymes en contexte, de détecter de façon automatique les métaphores et les métonymies. Les classes d'objets rendent compte aussi du comportement des prédicats nominaux et, en particulier, de leur conjugaison. Leur intérêt le plus général est de rendre compte des différents emplois des prédicats, en déterminant leurs schémas d'arguments et en rattachant à ceux-ci l'ensemble de toutes les propriétés qui les caractérisent et seulement celles-là.*

## ABSTRACT

*For a long time semantics was regarded with suspicion from a Harrisian perspective under the pretext that it could not give rise to a rigorous description of linguistic facts as it was assumed to be the case for syntax. As a matter of fact, this rejection was based on a form of autonomous semantics, oblivious to syntactic structures. The situation is however very different once one no longer postulates different levels and once one takes the minimal unit of analysis to be the sentence and not the word. The goal of this article is to show that syntactic constraints can be used to construct semantic classes, called « object classes », that constitute disjoint sets which are just as reproducible as syntactic descriptions. These classes are a means to separate in a very precise manner the different uses of predicates, to construct synonyms and antonyms in context, and to automatically detect metaphors and metonyms. Object classes also yield a way to account for the behavior of predicative nouns and in particular for their conjugation. Their most general use is to be found in accounting for the different uses of predicates in terms of their argument schemas and in associating with these the set of all properties that characterize them.*

## 1. FONDEMENT MÉTHODOLOGIQUE DES DICTIONNAIRES

Une pratique scolaire fréquente consiste à demander à un élève de définir le sens d'un mot. Cet exercice semble banal, mais il pose des questions de fond de nature théorique. Imaginons qu'il s'agisse de définir le « sens » du verbe *abattre*. À une telle question, il est impossible de répondre de façon satisfaisante. Admettons que l'élève réponde par *tuer*. On voit qu'il aura sélectionné l'un des sens du verbe *abattre*, celui sans doute qui lui venait le plus naturellement à l'esprit et qu'il identifie à l'aide d'un synonyme. La réponse est arbitraire et il ne serait pas étonnant qu'un autre élève propose *démoraliser*, un autre encore *couper*. Chacun d'eux aura pointé, parmi les différentes significations du verbe *abattre*, celle qui s'impose à lui au moment de la question. Le maître ne saurait évidemment se contenter des équivalences proposées. Il lui reviendrait de faire découvrir les conditions dans lesquelles le verbe *abattre* prend la ou les signification(s) proposée(s). Il apparaîtrait alors clairement qu'en dehors de tout contexte, il est impossible de décider auxquels des huit ou dix « sens » du verbe *abattre* on a affaire.

C'est donc le contexte linguistique, autrement dit son environnement, qui détermine le sens d'un terme.

Ce petit exercice scolaire permet d'envisager de façon nouvelle la notion de polysémie. Comme on le sait, la plupart des dictionnaires signalent dans leur préface l'étendue de leur nomenclature et précisent le nombre déterminé de significations auxquelles correspondent ces entrées. Cette différence justifie la notion de polysémie. Le nombre de significations est parfois impressionnant, comme dans le cas du verbe *prendre*. Or, il se pourrait que le concept même de polysémie soit un artefact de la lexicographie. En effet, dans un texte et de façon générale dans tout discours, les mots ne sont que très rarement ambigus. L'ambiguïté peut apparaître en cas de télescopage de constructions syntaxiques, comme le faisaient remarquer avec insistance les premiers travaux de la grammaire générative (cf. N. Chomsky, 1956: *Les Américains jugent les Vietnamiens coupables*). Quant à la polysémie proprement dite, on observera qu'il n'existe aucun prédicat qui ait plusieurs significations dans un environnement constant. La polysémie n'est donc pas un problème textuel, à la différence de la synonymie, qui correspond à une vraie réalité linguistique.

Les conséquences théoriques qu'on doit tirer de ces observations sont nombreuses et bouleversent les découpages habituels de l'analyse linguistique. Si un mot ne peut pas être défini en lui-même, c'est-à-dire hors contexte, mais seulement dans un environnement syntaxique donné, alors le lexique ne peut pas être séparé de la syntaxe, c'est-à-dire de la combinatoire des mots. La sémantique n'est pas autonome non plus : elle est le résultat des éléments lexicaux organisés d'une façon déterminée (distribution). Qu'il en soit ainsi est confirmé par les auteurs de dictionnaires eux-mêmes qui, timidement et sans aucune méthode, notent pour un prédicat donné, le ou les arguments qui permettent de séparer un emploi d'un autre. Pas de sémantique sans syntaxe donc, c'est-à-dire sans contexte.

## 2. CONTEXTES ALÉATOIRES ET CONTEXTES APPROPRIÉS

Nous allons examiner, dans ce qui suit, ce que nous entendons par le mot « contexte » et qui ne saurait être assimilé à ce que l'on appelle quelquefois, de façon superficielle, les environnements de gauche et de droite d'un mot.

### 2.1. Contextes aléatoires

On peut être tenté de dire que toutes les informations nécessaires à la reconnaissance du sens d'un mot sont définies par ses environnements stricts, susceptibles d'être dégagés à l'aide de logiciels d'extraction opérant sur de vastes corpus. Il suffirait d'éliminer les doublons pour déterminer ainsi comment ces environnements permettent de sélectionner le sens d'un mot en contexte. Or, cette recherche des co-occurents ne peut pas être mécanique. Une requête consistant à sélectionner les cinquante caractères qui précèdent ou qui suivent immédiatement un verbe, par exemple, ne sauraient constituer les conditions nécessaires à cette expérimentation. En effet, les éléments contextuels qui définissent un mot ne lui sont pas nécessairement contigus, comme on le postule par définition pour les éléments de la phrase simple, où les arguments encadrent strictement le prédicat. En effet, dans les textes, on observe souvent entre un prédicat et ses arguments des insertions de diverses natures, qui n'appartiennent pas en propre au schéma d'arguments et qu'il faut banaliser si l'on veut retrouver la distribution qui rend compte du sens du prédicat. Ces insertions peuvent être :

- des incisives :

Il faut tailler, comme on le sait, les arbres au début du printemps.

Louis XIV, à ce qu'on dit, a réduit le pouvoir des nobles.

Il se serait servi, à ce que j'ai cru comprendre, d'un exemple très ambigu.

L'État, nous le savons tous, tire ses ressources de l'impôt indirect.

- des négations :

Les adolescents ne jouent plus aux billes.

Il n'a répondu, sur aucun point, à mes questions.

- des adverbes :

Pierre a participé grandement à cette victoire.

Paul a changé radicalement de conduite.

- des circonstancielles :

Pierre a oublié, avant de partir, de fermer le gaz.

- des déterminants nominaux :

Paul a lu un tas de livres.

Il y avait à l'arrière du front un monceau de cadavres.

Il a versé dans son thé un nuage de lait.

Dans ces phrases, le système doit être capable de reconnaître que :

- l'environnement de droite de *tailler* est non pas *comme on le sait* mais le substantif *arbre* ;
- l'environnement de droite de *Louis XIV* est *réduire* et que *à ce qu'on dit* n'est pas pertinent pour la description de ce verbe ;
- la négation dans b) est une règle générale de la grammaire et qu'elle ne joue aucun rôle dans la définition syntaxique de *jouer*. Il arrive parfois que la négation fasse partie de la structure même de la phrase simple comme dans *n'en faire qu'à sa tête* ; ou *ne pas en rater une*. Ce type d'information doit alors figurer dans un dictionnaire ;
- les adverbes *grandement*, *radicalement* ne font pas partie du schéma d'arguments des prédicats *participer* ou *changer*, contrairement, par exemple, aux constructions suivantes où l'adverbe est obligatoire et constitue pour ainsi dire un « argument » : *se comporter Adv*, *présenter bien*, *bien / mal accueillir que* ;
- les propositions et compléments circonstanciels qui s'insèrent à l'intérieur d'un schéma d'arguments. *Avant de partir* doit être mis entre parenthèses pour pouvoir relier *oublier* et son complément *fermer le gaz*.

Un logiciel doit être en mesure de considérer *un tas de* et *un monceau de* comme des déterminants nominaux quantifieurs et non des compléments du verbe *lire*, par exemple.

On voit que si on définit les prédicats par la nature de leurs arguments, on est obligé de faire le tri, dans l'environnement immédiat, entre ce qui est ou n'est pas pertinent dans la définition des mots. La distribution d'un élément ne s'identifie donc pas à son

contexte matériel immédiat, mais repose sur une analyse qui détecte les éléments dont l'interdépendance constitue des ensembles significatifs.

## 2.2. Contextes contraints

On observe un grand nombre de cas où l'environnement immédiat, dépourvu d'éléments hétérogènes, n'est pas non plus éclairant pour la détermination de la construction et donc du sens d'un prédicat. C'est le cas des constructions figées.

### 2.2.1. Insertions dans les suites figées

M. Gross (1993) a signalé que les constructions verbales figées n'ont pas une structure interne différente des constructions régulières. Elles acceptent des insertions du type qu'on vient de voir, ce qui ne les distingue pas formellement des constructions libres, au regard de la contiguïté des éléments :

Pierre prend toujours les vessies pour des lanternes.

La moutarde lui a, semble-t-il, monté au nez.

Il faut donc chercher sur un autre plan la pertinence des environnements.

### 2.2.2. Un environnement contraint et non compositionnel

Indépendamment de ce fait, les constructions figées constituent, du point de vue sémantique, de « faux » environnements, puisque le sens du prédicat n'y est pas déterminé par la nature des arguments. C'est l'ensemble constitué par le verbe et la séquence des substantifs qui le suivent qui forme une unité sémantique. L'établissement du sens ne peut pas se faire sur la base d'une combinatoire, c'est-à-dire de la syntaxe régulière, mais de façon externe et globale. L'ensemble est doté d'une signification qui n'est pas fonction de celle de ses éléments composants. Il arrive que ces suites figées aient pour synonyme un lexème simple, mais ce n'est pas toujours le cas. De plus, un verbe figé, c'est-à-dire suivi d'un substantif qui n'a pas de réelle fonction syntaxique, peut en outre avoir des arguments réguliers, tout comme un verbe simple :

(tirer les vers du nez) à N hum

(donner du fil à retordre) à Nhum

(apporter de l'eau au moulin) de Nhum

### 2.2.3. Contextes significatifs

La détermination de sens d'un prédicat à l'aide du contexte n'est donc pas une opération triviale. Le contexte n'est une donnée immédiate que quand il s'agit de textes. Il est vrai qu'un discours n'est jamais une suite de phrases élémentaires, telles qu'elles sont définies dans les grammaires. Et pourtant la phrase simple est l'outil le plus important dans l'analyse des textes. Le premier travail de tout traitement automatique consiste donc à reconnaître toutes les phrases simples qui le composent. L'étude du « ciment » qui relie ces phrases entre elles constitue une seconde étape. C'est dans le cadre de la phrase simple que l'on peut mettre au point les environnements pertinents et, comme elle est définie par l'existence d'un prédicat, il faut donc être en mesure de détecter dans un texte l'ensemble des prédicats qui le composent. L'analyse consiste alors à reconnaître les structures phrastiques qui figurent dans un texte à partir de la description exhaustive des prédicats qui sont décrits dans le dictionnaire électronique.