

## **« LE COUPLE SEMANTIQUE/PRAGMATIQUE ET LE CALCUL DU SENS »**

M. Caterina Manes Gallo & Jacques Rouault  
Equipe Cristal-Gresec, Université Stendhal Grenoble

La lexie comprendra tantôt peu de mots, tantôt quelques phrases;  
ce sera affaire de commodité : il suffira qu'elle soit le meilleur espace possible  
où l'on puisse observer les sens : sa dimension, déterminée empiriquement,  
au juger, dépendra de la densité des connotations, qui est variable  
selon les moments du texte : on veut simplement qu'à chaque lexie  
il n'y ait au plus que trois ou quatre sens à énumérer.  
(R. Barthes, *S/Z*, 1970, pg. 18)

### **I - Introduction**

L'objet de notre communication est une modélisation de sémantique linguistique sur la structure profonde des faits linguistiques de type prédicatif. Les faits linguistiques analysés sont les constructions prédicatives, formées à partir soit d'un verbe avec ses actants régis, soit de l'association entre la copule *être* et un adjectif ou un participe passé. D'un point de vue méthodologique, nous nous plaçons dans le cadre du Traitement Automatique de la Langue Ecrite (T.A.L.E.) et visons des applications où la langue joue un rôle central: recherche d'information, communication homme-machine en langue naturelle écrite essentiellement.

D'après la perspective adoptée, nous considérons que la fonction du traitement sémantique est de relier le résultat du traitement syntaxique à la représentation pragmatique relative à l'évolution des propriétés des objets du discours. Le modèle présenté est fondé sur une combinatoire des caractéristiques lexicales des constructions prédicatives. Il implique une sémantique *panachée* qui s'appuie à la fois sur une vision continuiste et une vision compositionnelle du sens véhiculé par les séquences linguistiques.

#### **I-1 - Contraintes méthodologiques**

Un modèle pour le (T.A.L.E.), orienté vers la communication Homme/Machine doit satisfaire deux contraintes, en apparence opposées. Primo, il s'agit de définir un modèle à large couverture linguistique, et donc transportable. Ce qui implique la prise en compte de la variabilité des phénomènes prédicatifs de surface, indépendamment du domaine d'application de l'interface et sachant que ces phénomènes ne sont pas normés. Secundo, le modèle défini doit être opérationnel c'est à dire formalisable en un algorithme. Ce qui signifie essayer d'unifier le plus possible la variabilité des phénomènes prédicatifs, en essayant de les représenter sous la forme d'un ensemble restreint de règles combinatoires. D'un côté, on veut rendre compte du plus grand nombre de modalités selon lesquelles un locuteur peut exploiter le système de la langue naturelle, afin de s'exprimer selon un langage adapté à ses intentions communicationnelles. De l'autre côté, il est nécessaire de limiter les connaissances linguistiques et les structures abstraites à partir desquelles reconduire et/ou générer les séquences de surface.

Il ne s'agit donc pas de définir un modèle pour valider une théorie linguistique spécifique, mais de combiner les suggestions en provenance de plusieurs théories afin de définir un modèle opérationnel (Rouault, 1989). L'adoption d'un méta-modèle tel que le *schéma de lexis*, défini par Culioli (1990) se justifie par la nécessité de disposer d'une méta-structure linguistique homogène pour représenter les informations nécessaires au calcul du sens des séquences de surface, indépendamment du niveau de complexité de leur forme.

## **I-2 - Contraintes théoriques**

La formulation du modèle de sémantique linguistique implique deux grandes étapes distinctes. La première étape, doit aboutir à l'identification du type d'énoncé noyau sous-jacent à la séquence de surface. Un type d'énoncé est défini à partir des mécanismes prédicatifs réguliers hors incidents d'énonciation. Nous avons analysé ces mécanismes dans les termes de l'opposition action vs. propriété, donc dans le cadre de leur fonction par rapport à la progression du discours (cf. § III). En revanche, la deuxième étape doit intégrer les transformations induites sur l'énoncé noyau par les opérations énonciatives (temps, aspect, modalité, déterminants,...), afin de repérer le type d'énonciation à laquelle renvoie la séquence de surface. Cette deuxième typologie implique une analyse des régularités des contraintes exercées par les mécanismes prédicatifs réguliers sur les opérations d'énonciation (cf. § IV).

## **II - Cadre théorique**

Nous considérons que, à la fois en reconnaissance et/ou en génération, le traitement sémantique constitue un *pont* qui permet de relier la sortie du traitement syntaxique avec le traitement pragmatique. Il s'agit donc de rendre compte des mécanismes linguistiques qui contribuent à cette *fiction conceptuelle inédite et provisoire* qui correspond à la construction du sens par le discours (Berrendonner, à paraître).

### **II-1 - Sémantique linguistique**

Nous considérons que les opérations de prédication contribuent à la progression du discours, à partir de la manipulation des propriétés des objets qui y sont posés explicitement ou par inférence. Le but d'une modélisation de sémantique linguistique est donc d'identifier les opérations prédicatives qui correspondent à l'opposition (intuitive) entre action et propriété. Le sens implicite véhiculé par le verbe dépend d'un référent abstrait qui est le type de procès (TP) auquel il renvoie, indépendamment de son référent extralinguistique (Fuchs, 1992). Par exemple, dans le cas de : *la tondeuse a coupé l'herbe, Jean a coupé le gâteau, un bus a coupé la route à la trottinette, le paysan a coupé le vin avec de l'eau* le sens implicite pertinent véhiculé par ces énoncés renvoie toujours à un procès accompli, qui induit une modification de l'actant de droite. Phénomène non réalisé par l'énoncé *devant mon objection il a coupé court* qui renvoie au contraire à un procès inaccompli. La référence à un (TP) est indépendante à la fois des modalités d'exécution de l'action « mondaine » (se rapportant au domaine d'application) figurée par le signifié concret du verbe et des effets de l'exécution de cette action, selon les caractéristiques mondaines des objets sur lesquels elle porte. Les signifiés concrets de *couper l'herbe* ou *couper la route* sous-entendent une série de sous-actions impliquant un déplacement dans l'espace de l'agent coupant. Eventualité plus improbable dans les cas de *couper un gâteau* et *couper le vin*. De plus, le résultat *l'herbe est coupée* et *le gâteau est coupé* implique une discrétisation de l'objet visé tandis que après l'accomplissement du même procès, *la route* et le *vin* demeurent deux objets compacts.

Les connaissances linguistiques qui permettent de déterminer le référent abstrait (ou TP) d'une construction prédicative sont représentées en termes de deux structures objet définies *schéma d'action* (SA) et *schéma d'état* (SE) (Manes Gallo, Rouault, 1998). Comme détaillé par la suite, le type de procès référent d'une construction prédicative est déterminé par une double combinatoire des compléments régis par le prédicat et les traits sémantiques des arguments occupant les différentes places syntaxiques (Manes Gallo, Rouault, 1997). Nous avons fait temporairement abstraction des prédictions portées par des nominalisation et par des compléments régis réalisés à travers des complétives.

### **II-2 - Pragmatique linguistique:**

Nous considérons la représentation pragmatique comme une *mémoire discursive* qui est mise à jour par les résultats du traitement sémantique au fur et à mesure que le discours avance. Le modèle pragmatique a pour tâche de représenter les objets, les propriétés et les actions que le discours met en oeuvre. Il a été défini comme un langage de représentation des connaissances à base nominale et orienté objet. Ce formalisme a la fonction d'un côté d'organiser l'information extraite du discours, et de l'autre d'apparier cette information avec l'information en provenance des connaissances sur le domaine d'application (Berrendonner, Rouault, 1991). Le FONCTIF est une sous-structure qui permet de décrire les connaissances linguistiques et extra-linguistiques associées à un constituant nominal (Berrendonner et al., 1992). D'où la nécessité de disposer de structures d'interface entre niveau sémantique et niveau

pragmatique. Les (SA) et (SE) sont les deux méta-structures sémantico-pragmatiques qui ont la fonction de représenter, de façon exploitable par la *mémoire discursive*, le sens implicite d'une énonciation.

Toutefois cette conception constructiviste du fonctionnement de la langue dans le discours bien que nécessaire n'est pas suffisante. Si on vise un système transportable on ne peut faire complètement abstraction des connaissances extra-linguistiques qui permettent d'associer aux prédicats leurs référents mondains, selon le domaine d'application de l'interface. La possibilité de reconnaître l'occurrence de synonymies lexicales dans le discours s'appuie en partie sur la prise en compte des *systèmes complexes de représentation intra-culturels des phénomènes* (Culioli, 1990). En reprenant l'exemple précédent, *couper l'herbe* peut être appareillé avec *hacher* et *couper le vin* avec *mélanger*.

Le (SA) est constitué d'une série de sous-objets qui contiennent des informations linguistiques et extra-linguistiques sur les constructions prédicatives qui renvoient à des actions. Les connaissances linguistiques du (SA) sont relatives aux PRE-CONDITIONS pour le déclenchement du procès, aux ROLES des différents actants, selon leur place syntaxique, et au RESULTAT-UISE qui peut être inféré après l'accomplissement du procès.

Les connaissances extra-linguistiques véhiculées par le signifié concret d'un verbe, et qui dépendent du domaine d'application, sont représentées dans la case PLAN du (SA). Ces connaissances peuvent être assimilées aux *occurrences phénoménales* qui permettent d'appréhender un certain domaine notionnel (Culioli, 1990). Le (SA) en tant que méta-structure sémantico-pragmatique sert à aiguiller le calcul du référent concret d'une construction prédicative, à partir en partie d'informations linguistiques sur la fonction de cette construction par rapport à la progression du discours.

Une telle représentation permet la mise en oeuvre d'un certain nombre de raisonnements de nature linguistique: raisonnement déductif permettant d'inférer un état résultant à partir d'un énoncé de type processus résultatif présenté à l'accompli; raisonnement de type abductif permettant de remonter, à partir d'un état, vers les processus résultatifs qui ont pu conduire à l'état présent de la mémoire discursive, etc.

En récapitulant, dans le cadre de la communication Homme/Machine en langue naturelle écrite, le passage du niveau sémantique au niveau pragmatique implique résoudre simultanément deux problèmes de la construction du sens. Il s'agit d'un côté de rendre compte, au niveau sémantique, des transformations de la langue en un discours. Ces transformations, impliquent le calcul du sens implicite véhiculé par les constructions prédicatives, par rapport à la progression du discours. Ce calcul s'appuie sur une double opposition entre individu vs. relation, et au sein des relations, entre action vs. propriété. De l'autre côté, il s'agit de mettre en relation la progression du discours avec le monde extra-linguistique de l'application. Le repérage du référent mondain des séquences de surface se fera en appareillant la représentation pragmatique de leur sens discursif avec les connaissances sur le domaine d'application. Par exemple, si la construction prédicative renvoie à l'accomplissement d'un processus alors il sera nécessaire de calculer les effets concrets de cet accomplissement sur l'objet visé. Par exemple, *la tondeuse coupe l'herbe* implique une discrétisation de *l'herbe*. En revanche, si le (TP) référent n'implique pas un accomplissement, e.g. *devant mon objection il a coupé court*, alors ce calcul ne sera pas nécessaire.

### III - Structure du modèle sémantique

Comme précédemment indiqué (cf. § II) nous considérons que le modèle de sémantique linguistique a pour vocation de fournir une interprétation de la sortie du traitement syntaxique, qui soit exploitable par le traitement pragmatique. La modélisation présentée se fonde sur deux objets théoriques tel que le *schéma de lexis* (Culioli, 1990), dont nous nous inspirons assez librement, et les *types de procès* (Fuchs, 1992).

#### III-1 - Le méta-modèle linguistique

Comme indiqué dans le § I-2, le calcul du sens d'une séquence de surface implique en premier lieu l'identification du type d'énoncé noyau auquel elle peut être associée. D'où la nécessité de disposer d'une méta-structure linguistique qui permette de représenter de façon homogène et de calculer les ROLES joué par les différents constituants syntaxiques d'une construction prédicative, par rapport à son sens implicite.

La *théorie de la lexis* formulée par Culioli (1990) nous a paru le cadre théorique qui pouvait mieux répondre à nos préoccupations. D'après cet auteur, une relation prédicative permet d'appréhender

différents domaines notionnels pré-linguistiques. Il s'agit d'une relation primitive non saturée, formée et ordonnée, selon le *schéma de lexis*  $\langle 1, \text{ , } 2 \rangle$  dans lequel 1 et 2 sont deux variables d'arguments et une variable d'opérateur de prédication. La *lexis* est l'instanciation d'un schéma à partir de la formation d'un faisceau de relations primitives entre les éléments constitutifs de la relation de prédication. Toute énonciation d'un locuteur peut donc être représentée comme une lexis  $\langle A, R, B \rangle$ , où R (relateur) a un statut prédicatif et A (notion source) et B (notion but) ont un statut non-prédicatif. Les *lexis* correspondent à la fois à des contenus propositionnels *dicibles* et à des formes pré-assertées de relations prédicatives primitives. Le problème est que Culioli ne précise pas la nature de ces relations, d'où la nécessité d'opérer une série d'aménagements afin de rendre plus opérationnelles les modalités de construction d'une lexis.

Nous considérons que le type d'énoncé noyau associé à une séquence de surface corespond à une *lexis*. Le sens implicite véhiculé par le relateur (R) est identifié au type de procès référent, tel qu'il résulte des données lexicales. A partir des critères formulés par Fuchs et Léonard (1979), nous avons retraduit l'opposition action vs. propriété dans les termes de trois types de procès. A savoir, les processus résultatifs (PR), qui induisent une manipulation des propriétés des arguments à droite du verbe, les processus non-résultatifs (PNR) et les états (E) qui affectent plutôt les arguments à gauche du verbe (Gallo, Rouault, 1992).

Le fonctionnement de base d'un prédicat est défini dans le cadre d'une construction de type sujet - verbe - complément d'objet. Pour les constructions prédicatives plus complexes, ce référent premier du relateur peut être transformé par l'occurrence d'un ou plusieurs des trois éléments suivants :

- a) La combinatoire des compléments régis (rang >1) qui peuvent être calculés automatiquement, selon qu'ils soient obligatoires ou facultatifs.
- b) Les traits sémantiques des arguments occupant les différentes places syntaxiques, à droite et à gauche du verbe
- c) Les opérations d'énonciation (temps, aspect, modalités, déterminants, ...) et certains des compléments circonstanciels (« temps », « lieu », etc.) que nous considérons comme relevant des choix du locuteur.

Ces trois éléments contribuent à fixer et/ou à transformer le référent premier (processus vs. état) d'un prédicat, selon les intentions communicationnelles du locuteur. Le modèle que nous présentons concerne les marques du co-texte mentionnées aux points a) et b) et les interactions avec les opérations d'énonciation portées par les déterminants (point c).

## II-2 - Le modèle calculatoire

Le calcul du sens d'une séquence de surface implique en premier lieu l'identification du type d'énoncé noyau auquel elle peut être associée, hors incidents d'énonciation. Le but est d'établir à quel type de procès renvoie l'énoncé noyau afin de calculer l'opération prédicative induite, i.e. action vs. état, et les actants affectés par la prédication. Ce calcul implique l'intégration de plusieurs types d'informations. La première information concerne les éventuels glissements du référent d'une même unité verbale, selon les schémas syntaxiques qui peuvent lui être associés. Pour les processus, une deuxième information concerne l'axe temporel ou spatial sur lequel il évolue et qui dépend étroitement du rang du/des compléments régis. Enfin il s'agit de déterminer la stabilité du référent du prédicat aussi selon les caractéristiques de la notion source (A). C'est le cas de l'instrument (Manes Gallo, Rouault, 1998).

L'intégration de ces différentes informations vise à spécifier et à opérationnaliser la combinatoire des relations prédicatives primitives entre les trois constituants de la *lexis*. Cette combinatoire repose sur une double catégorisation. La première catégorisation coïncide avec une reformulation de la classification des compléments régis formulée par (Gross, 1975), qui fait coïncider le rang d'un complément avec le type de particule préverbale (ppv) qui peut le substituer. Par exemple, les particules *il(s)/elle(s)* renvoie à un complément de rang 0, tandis que *y* renvoie à un complément de rang 3. La reformulation vise à ajouter une dimension sémantique aux (ppv), à travers l'intégration des traits sémantiques des arguments pouvant occuper les différentes places syntaxiques. Cette intégration est nécessaire à cause de la dimension sémantiquement *incolore* des prépositions les plus fréquemment utilisées (Cadiot, 1989). Par exemple, la préposition *à*, selon les traits sémantiques *animé* vs. *inanimé* de l'argument, peut renvoyer soit à un complément de rang 2 *prêter à*, *téléphoner à* (*QUI*) soit à un complément de rang 3 *aller à*, *admettre quelqu'un à* (*QUOI*) (Manes Gallo, Rouault, 1997).

La deuxième catégorisation concerne une sous-catégorisation de notre classification primitive des types de procès. La distinction entre processus et état dépend de la possibilité de transformer le verbe à la forme progressive. Nous distinguons trois types d'état : Etat0 *être grand*, Etat1 *savoir*, Etat résultant *être*

*cassé*. La sous-catégorisation des processus dépend du rang du complément régi, selon que ce dernier fasse ou pas référence à l'axe spatial. Ainsi au sein des (PNR), qui sont repérables par rapport à l'axe du temps, e.g. *aider le voisin*, nous distinguons les processus résultatifs spatiaux (PRS) qui renvoient à l'accomplissement d'un changement de localisation, e.g. *aller à Rome, arriver de la gare*. Enfin les (PRT) sont repérables par rapport à l'axe du temps, e.g. *manger une pomme, connecter un terminal au serveur*.

L'intégration des informations relative au co-texte (cf. points a) et b)) de l'énoncé noyau se fait en deux étapes. A la sortie de la syntaxe on dispose d'une double structure syntaxique - en constituants et fonctionnelle - de la séquence de surface. Les dictionnaires (morphologique et des verbes) activés lors du traitement syntaxique permettent aussi d'associer aux syntagmes identifiés un certain nombre d'étiquettes. Les étiquettes sont relatives : à la temporalisation (tense), au caractère spatial ou non, aux traits sémantiques des actants potentiels et au type de procès lexical du prédicat, selon le rang des compléments régis. Ainsi l'emploi intransitif *manquer de* du verbe transitif (PRT) *manquer (QUI/QUOI)* le fait glisser dans la catégorie des (E1). Chaque type de procès établit des relations différentes entre les actants et le prédicat. La nature de ces relations détermine le RÔLE joué par les arguments occupant les différentes places syntaxiques. Le schéma syntaxique correspondant à la construction prédicative de surface est traduit en un schéma sémantique, en appliquant aux compléments les RÔLES déterminés par le (TP) lexical. L'identification de ce dernier permet donc de construire la *lexis* correspondante au type d'énoncé lexical (TPL) sous-jacent à la séquence de surface. L'intégration des caractéristiques du C0 (sujet grammatical) donne une première valeur TE1 à l'énoncé noyau. Les caractéristiques du C0 sont représentées en terme de *causalité intrinsèque* et/ou de *causalité extrinsèque* par rapport à la mise en oeuvre du procès (Manes Gallo, Rouault, 1998). Pour les (PRT) le cas de l'*instrument* est reconstruit en intégrant au TPL l'information sur la fonction de *causalité intrinsèque* de l'argument qui joue le RÔLE d'agent dans l'accomplissement du processus. Ensuite on prend en compte les marques d'énonciation, en particulier celles portées par les déterminants et on obtient une deuxième valeur améliorée TE2 du type d'énoncé correspondant à la séquence de surface. La structure sémantique définitive, transmise au modèle pragmatique, implique enfin la prise en compte des traits sémantiques des arguments 2 et 3 régis par le verbe. On aboutit ainsi à une troisième valeur TE3 qui est la représentation linguistique du type d'énoncé exploité par le modèle pragmatique.

## IV - Travail en cours

De notre point de vue, les opérations d'énonciation reposent en partie sur les déterminants définis et/ou indéfinis associés aux arguments qui encadrent le verbe. L'accent mis sur les déterminants *LE, LA, LES, UN* et *DES* vise à satisfaire la contrainte méthodologique évoquée au § I-1, relative à l'opérationnalisation du modèle. L'hypothèse générale est que les autres déterminants (e.g. *CELUI-CI, CELUI-LA*) sont des sur-déterminations des opérations d'énonciation portées par les cinq déterminants précédents qui sont, eux, considérés comme primitifs.

La fonction des déterminants est d'un côté de signaler la référence à un individu (*il épluche UNE pomme de terres pour le dîner*) ou à une classe (*il épluche DES pommes de terre pour le dîner*) et de l'autre d'explicitier aussi la stratégie énonciative choisie par le sujet pour (re)introduire son/ses objets de discours. Les déterminants définis expriment des opérations de fléchage, ou de reprise d'un objet précédemment introduit, tandis que les déterminants indéfinis renvoient à une opération d'extraction d'une classe et/ou d'un individu (Culioli, 1990). Ainsi, la différence entre *il épluche DES pommes de terre pour le dîner* et *il épluche LES pommes de terre pour le dîner* dépend des deux modalités énonciatives mises en oeuvre pour se référer au même objet de discours, i.e. *pommes de terre*. Le (TE1) correspondant à ces deux séquences est le même, i.e. un (PRT) dans lequel *pommes de terre* joue le RÔLE de patient. Dans le premier cas, *DES* renvoie à une opération d'extraction d'une sous-classe extensionnelle, au sein de la classe extensionnelle maximale du type *pomme de terre*. Le RESULTAT-VISE *des pommes de terre sont épluchées* indique l'introduction d'un nouveau objet de discours. Tandis que, dans le deuxième cas, *LES* renvoie à une opération de fléchage sur la sous-classe intensionnelle *pomme de terre pour le dîner*, faisant partie du type intensionnel *pommes de terre*. Le RESULTAT-VISE *les pommes de terre pour le dîner sont épluchées* indique un changement d'état et/ou l'attribution d'une nouvelle propriété à un objet de discours soit précédemment introduit, soit acquis comme faisant partie des connaissances partagées entre les interlocuteurs.

Les transformations induites par les déterminants primitifs sur (TE1) contribuent à la prise en compte de l'évolution des objets du discours. Leur intégration permet de relier le schéma sémantique de l'énoncé au type d'objet intensionnel vs. extensionnel auquel il se réfère au niveau pragmatique. Surtout

dans le cas des (PRT) et des (PRS) la manipulation des propriétés de l'actant visé dépend du type d'objet auquel se réfère l'argument au niveau pragmatique. Ainsi l'argument extensionnel de *manger un boeuf* sera affecté par le (PRT) *manger* contrairement à l'argument intensionnel de *manger du boeuf*. Le déterminant *du* (amalgame entre *de* et *le* ) par une opération de fléchage contribue à l'introduction d'un objet intensionnel qui inhibe la visée résultative portée par le prédicat.

## V - Références bibliographiques

- Berrendonner A. & Rouault, J. (1991). Sémantique des objets et calcul des noms,  
*KMET 91*, Nice, 1-8
- Berrendonner A., Fredj M., Oquendo F. & Rouault, J. (1992). Un système inférentiel orienté objet pour  
des applications en langues naturelles, *COLING 92*, Nantes, 1-7
- Berrendonner A. La mémoire discursive  
(à paraître)
- Cadiot P. (1989). La préposition : interprétation par codage et interprétation par inférence,  
*Cahiers de Grammaire*, 14, 25-51
- Culioli A. (1990). *Pour une linguistique de l'énonciation*,  
Ophrys, Paris.
- Fuchs C. & Léonard A.M. (1979). *Vers une théorie des aspects*,  
Mouton, Paris
- Fuchs C. ed (1992). *Les typologies de procès*,  
Klincksiek, Paris
- Gallo M.C. & Rouault J. (1992). Schéma d'action et types de procès,  
*Intellectica*, 1/2, 105-127
- Gross M. (1975). *Méthodes en syntaxe*,  
Herman, Paris
- Manes Gallo M.C. & Rouault J. (1997). Rôles sémantiques et schémas d'énoncés, *Actes  
TALN97*, PUG, Grenoble, 15-29
- Manes Gallo M.C. & Rouault J. (1998). Connaissances véhiculées par les verbes dans le discours : les  
schéma d'action, *Verbum*, tome XX, n° 2, 189-210
- Rouault J. (1988). Apports, contraintes et limites du langage dans le traitement automatique des langues,  
*Actes du colloque « Perspectives méthodologiques et épistémologiques dans les sciences du  
langage »*, Peter Lang, Bern, 141-157