

Les paliers de la sémantique

Patrice Enjalbert, Bernard Victorri

▶ **To cite this version:**

Patrice Enjalbert, Bernard Victorri. Les paliers de la sémantique. Patrice Enjalbert. Sémantique et traitement automatique du langage naturel, Hermès, pp.55-98, 2005. halshs-00009779

HAL Id: halshs-00009779
https://shs.hal.science/halshs-00009779

Submitted on 28 Mar 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Les paliers de la sémantique

Patrice Enjalbert et Bernard Victorri

Dans l'analyse sémantique il est commode de distinguer plusieurs niveaux, ou *paliers*, correspondant à une complexité croissante des « objets » linguistiques à prendre en compte. Pour esquisser rapidement cette « hiérarchie », reportons-nous à notre premier examen intuitif de la construction du sens au chapitre 1. Nous avions distingué un premier niveau qui est celui du *lexique*, c'est-à-dire des *mots*, en quelque sorte — et en première approximation — constituants élémentaires porteurs de sens ; puis celui de la *phrase* — ou plus généralement du *syntagme* — assemblage structuré de mots en un tout cohérent porteur d'un sens relativement autonome ; et finalement celui du *texte*, ou *discours*, suite de phrases plus ou moins fortement corrélées, et qui constitue la forme globale que prend, en général , le message linguistique. Pour exprimer ce jeu d'échelle — certains diront de granularité — nous pourrons adapter ici une terminologie due à F. Rastier[1] et distinguer une *microsémantique* pour la sémantique lexicale, une *mésosémantique* pour le syntagme et la phrase, et une *macrosémantique* pour le texte.

Ce point de vue « structurel » doit être complété par une perspective qui resitue le texte/discours dans un acte de communication, laissant toute sa place au libre jeu de l'interprétation du lecteur/auditeur, et que l'on désigne communément sous le terme de *pragmatique*. Ce que certains verront comme un « au delà » de la sémantique, et qui pour nous constitue plutôt un regard particulier opérant aux trois niveaux.
Certes la subdivision en trois paliers possède une portée linguistique très générale, mais c'est sous l'angle de la sémantique que nous la considérerons ici. Les questions que nous devrons nous poser sont : quelles « valeurs » sémantiques peuvent être associées aux objets linguistiques relevant de ces différents niveaux ? Quelles structures les organisent ? Comment décrire les unes et les autres, comment les *représenter* — en particulier d'un point de vue formel dans une optique de TALN ? Et d'un autre côté, quelles *opérations* permettent de créer ces valeurs et ces structures dans ce que nous avons appelé « la construction du sens » ? Cette double problématique représentation/opérations sera un fil conducteur de ce chapitre.

Finalement, il faut insister sur le fait que ces niveaux sont tout sauf étanches. La langue n'est pas un jeu de poupées russes ! Par exemple nous verrons que, *in fine*, la sémantique d'un item lexical se définit dans un contexte d'énonciation et un co-texte qui lèvent les ambiguïtés et en ajustent en quelque sorte le sens. Un certain nombre de mécanismes que nous choisirons ici de décrire au niveau du texte (par exemple ceux de la coréférence et de la temporalité) apparaissent en fait dès celui de la phrase. Etc. Disons que ces « paliers » sont plutôt des subdivisons de méthode que de fait, ou des sous-systèmes articulés du système global de la langue. Ces remarques étant posées, nous pouvons en examiner tour à tour les caractéristiques spécifiques.

[1] Cf. [RAS 94]. Adaptation libre, sous notre seule responsabilité.

1. Le mot

1.1. *Quelles unités minimales ?*

La première question sur laquelle il convient de revenir est celle des unités minimales porteuses de sens, de la caractérisation des signes linguistiques *élémentaires*. En première approximation nous avons répondu : ce sont les *mots*. A y regarder de près, la situation est plus complexe.

En premier lieu on trouve un certain nombre d'expressions plus ou moins figées dont la décomposition en mots est problématique ou cacherait un sens spécifique de l'expression dans son ensemble. Des exemples viennent immédiatement : *pomme de terre*, *river son clou* (à quelqu'un), *jusqu'à, en face, double faute* (au tennis)… La question est très concrète et importante pour le TALN, car si l'on définit séparément les mots *pomme, terre, river, son, clou…* il faudra élaborer des mécanismes permettant de calculer le sens des expressions elles-mêmes. Un simple examen montre la difficulté de l'entreprise. On préférera alors regrouper sous le terme de *lexie* à la fois les mots et un ensemble d'expressions figées, et en faire les entrées du dictionnaire. Mais on pourra aussi traiter certaines constructions spécifiques, suffisamment régulières, par un calcul compositionnel. Observons sur ce premier problème un phénomène intéressant, à savoir que le sens d'une lexie complexe n'est pas réductible à la « somme » du sens de ses parties : dans *double faute* il y a effectivement l'idée d'une faute double (ou redoublée) mais on ajoute un trait spécifique lié aux règles du service au tennis. Cet « ajout d'information » est une caractéristique assez générale de la *compositionnalité* sémantique, sur laquelle nous aurons l'occasion de revenir.

On est donc conduit à considérer comme « élémentaires » des unités plus complexes que le mot. Mais inversement, on est aussi amenés à analyser les mots en unités plus petites. Observons-les dans un texte par exemple cet extrait de constat d'accident :

> *J'étais à un stop avec 2 voitures devant moi (...) Alors que la première voiture passait ce stop je fis mon contrôle à gauche et je démarrais mais je percutais la deuxième voiture (...)*

Les verbes apparaissent évidemment sous une forme conjuguée : *passait, fis, démarrais, percutais, …* tandis que les noms et adjectifs sont accordés en genre et nombre : *2 voitures, la première voiture…* On doit donc analyser ces mots en constituants, un radical qui marque le mot proprement dit et une désinence qui indique le temps verbal, le genre, le nombre etc. : *pass-ait, f-is, démarr-ais, percut-ais, voiture-s, premièr-e.* On parlera de *morphème*[2] pour désigner les unités véritablement minimales porteuses de sens. On distinguera en particulier, comme mis en évidence dans notre exemple, des morphèmes *grammaticaux* (comme *-ait, -is, -é, -s, -e*) qui, du point de vue sémantique, portent des valeurs grammaticales de temps, genre, nombre etc. ; et des morphèmes *lexicaux* relatifs à une certaine « réalité conçue » : *pass-* (pour le concept général de 'passer'), *démarr-* (pour 'démarrer'), *premier-* (pour 'premier'), etc.

Mais on peut aussi considérer que ce que nous avons appelé « radical » est susceptible lui-même, dans certains cas, de décomposition. Prenons le mot *agriculteur*[3]. On peut le décomposer en *agri* (renvoyant selon l'origine latine à *ager*, champ), *cult-* (renvoyant à 'culture', 'cultiver'), et le suffixe *-eur* (indiquant l'auteur d'une action, comme dans : *aviateur, percepteur, menteur…*). Toutefois, on voit bien que le sens actuel en français de ce mot va au delà de la culture de champs, incluant

[2] Ou *monème* selon Martinet. Cf. [DUC 72], p. 261.
[3] Exemple emprunté à F. Rastier [RAS 94], p. 56.

notamment la pratique de l'élevage : nouvel exemple de la limite d'une compositionnalité sémantique « stricte ».

De manière générale *l'analyse morphologique* consiste à analyser un mot en morphèmes (quelle que soit finalement la notion retenue). En pratique en TALN on se limitera le plus souvent à établir, pour un mot donné, un *lemme*, c'est-à-dire une forme canonique telle qu'une entrée dans un dictionnaire (*passer, faire, voiture, premier*…), et une *flexion* caractéristique d'une forme grammaticale. Ainsi *passais* sera analysé en un lemme : *passer*, et une flexion (avec ambiguïté) : indicatif + imparfait + première ou seconde personne du singulier. On dira que *passais* est une forme fléchie de *passer*. On considérera alors de même que *le, la, les* d'une part, *un, une, des* sont des formes fléchies en genre et nombre du même article défini (lemme : *le*) ou indéfini (lemme : *un*)[4]. Du point de vue sémantique, on aura des mécanismes de compositionnalité : le sens d'une forme particulière résulte des valeurs sémantiques associées aux morphèmes qui le constituent — cette « résultante » ne devant pas être conçue trop strictement, comme l'ont montré les exemples précédents.

Il faut souligner ici que les « formes grammaticales » sont elles-mêmes signifiantes[5]. Il peut s'agir de *mots* tels que les déterminants (*le, un, ce*…), prépositions (*à, de, par*…) ou conjonctions diverses (*quand, où, et, car*…). On parlera alors de *lexique grammatical*, par opposé à un *lexique plein* (ou lexique « tout court ») contenant les noms verbes, adjectifs… qui renvoient à des objets ou concepts. Mais les *morphèmes flexionnels* dont nous venons de parler doivent également être considérés comme porteurs de sens. Ce point sera particulièrement développé dans le présent ouvrage à propos du temps grammatical (chapitre 5) : nous verrons en effet comment le temps d'un verbe, joint à d'autres indices, permet de calculer effectivement, sinon la position temporelle précise de l'événement auquel il réfère, du moins un ensemble de contraintes très fortes sur cette position.

Nous allons toutefois nous concentrer dans cette section sur le lexique *plein* et nous considérerons les *lexies* comme unités élémentaires (sans prendre donc en considération la décomposition en morphèmes). Le lexique grammatical possède en effet des caractéristiques spécifiques, liées à son usage général en langue et à son caractère fortement polysémique, qui en font un objet d'étude particulièrement complexe[6]. Notre exposé portera principalement sur la *représentation du sens lexical*, en particulier de manière formelle exploitable dans des procédures informatiques. Nous aborderons également le problème de son *calcul* en contexte ; ce qui pose notamment le problème de la *polysémie* (multiplicité de sens), un fait majeur qui fait l'objet du chapitre 3.

1.2. *Représentations du sens lexical*

Il convient immédiatement de noter qu'il n'y a pas de modèle unifié, mais plutôt différents modes de représentation, répondant à des options théoriques et pratiques différentes. Sans aucune prétention à une vue universelle et univoque de ce vaste champ linguistique[7], nous proposerons ici de distinguer

[4] D'un point de vue pratique il est très remarquable que nous disposons aujourd'hui d'analyseurs morphologiques capables d'effectuer cette reconnaissance de manière très fiable et efficace. Voir par exemple [GIG 97] [SCH 94].

[5] Contrairement à la tradition en Recherche Documentaire qui considère comme « vides » les mots grammaticaux (cf. chapitre 9). Un point de vue sans doute justifié par les objectifs spécifiques de cette technique, mais qui ne saurait être généralisé.

[6] Voir notamment la distinction opérée par Talmy [TAL 00], p. 21, qui définit deux « sous-systèmes » ayant des fonctions sémantiques nettement différenciées : le sous-système grammatical, qui détermine la structure de la représentation cognitive évoquée par un énoncé, et le sous-système lexical, qui en détermine le contenu.

[7] [NYC 98] en propose une vue assez complète. Les principaux modèles utilisés en I.A. sont décrits et détaillés dans [KAY 97] et [SAB 89].

trois « types » de modèles (comme toujours, sans barrière étanche) : les deux premiers sont de loin les plus fréquents en TALN et le troisième, aujourd'hui d'intérêt plus « linguistique » fera office de contrepoint et d'ouverture. Pour aider à situer ces approches les unes par rapport aux autres, nous discuterons particulièrement trois questions spécifiques:

– *Rapport entre sens « linguistique » et représentations cognitives*. Le problème est celui de la « nature » des représentations sémantiques : renvoient-elles à un univers de signifiés « proprement linguistiques » ? ou à l'univers général des « connaissances » et « représentations cognitives » du locuteur/interprète ?

– *La question des primitives*, abondamment discutée en sémantique lexicale : existe-t-il quelques *éléments de sens* premiers, à partir desquels tous les termes peuvent être définis, caractérisés ? si oui, comment les définir ? La réponse a des répercussions évidentes en termes de méthode d'élaboration des lexiques sémantiques.

– *Instanciation et référence*. Les mots renvoient par eux-mêmes à des « concepts » généraux, alors que le discours en présente (en général) des « instances » particulières, entités ou événements. Nous reviendrons dans la prochaine section sur cette importante opération, et la question sera ici de savoir comment le modèle peut la « prévoir ».

1.2.1. Des relations entre concepts

Une première manière, assez naturelle, de décrire le sens lexical est de considérer que les mots ou lexies renvoient à un réseau de « concepts » inter-reliés et interdépendants. Une première illustration est donnée par les relations élémentaires et universellement admises que sont :

– *L'hyperonymie et l'hyponymie*. Un mot A est hyponyme d'un mot B s'il désigne en quelque sorte un cas particulier : *canari* est un hyponyme de *oiseau* qui l'est de *animal* ; *vélo* est hyponyme de *deux-roues*, lui même de *véhicule*. La *conduite* est une *action* particulière, donc de nouveau dans un rapport d'hyponymie, de même que les verbes *conduire* et *agir*. *L'hyperonymie* est la relation symétrique : *oiseau* est hyperonyme de *canari*, etc.

– *La synonymie et l'antonymie*. Deux mots synonymes désignent des notions ou des concepts proches (avec en général une nuance de sens ou d'usage) : *salaire*, *appointements* et *traitement*, *voiture*, *bagnole* ou *automobile*, *redouter*, *craindre* et *avoir peur de*, etc. Deux antonymes ont au contraire des sens opposés sous un rapport ou un autre : *chaud* et *froid*, *cru* et *cuit*, *pur* et *impur*, *lever* et *baisser*…

De telles relations sont à peu près incontournables. On peut cependant les voir comme un cas particulier de *relations entre concepts*. Un modèle plus général, proposé dans les années soixante à la charnière de la Psychologie Cognitive et de l'Intelligence Artificielle, est celui des *réseaux sémantiques*. Devenus par la suite mode de représentation des connaissances « universel », les réseaux sémantiques sont avant tout et à l'origine un modèle de description du lexique, proposé à cet effet en 1966 par Quillians [QUI 68] (voir aussi [CAR 89] p. 103 sqq.). De nombreuses variantes ont été proposées, le caractère commun étant une représentation sous forme de graphes dont les nœuds sont étiquetés par des mots et les arcs par des relations entre les concepts associés à ces mots. A titre d'illustration, nous donnons dans la figure 1 un petit fragment d'un réseau relatif à un champ lexical « déplacements urbains ». On observe notamment que les liens d'hyponymie sont inclus dans le graphe sous la dénomination *est-un*. D'autres liens portent des étiquettes plutôt linguistiques (relations actancielles) : MOYEN, AGENT, OBJET…

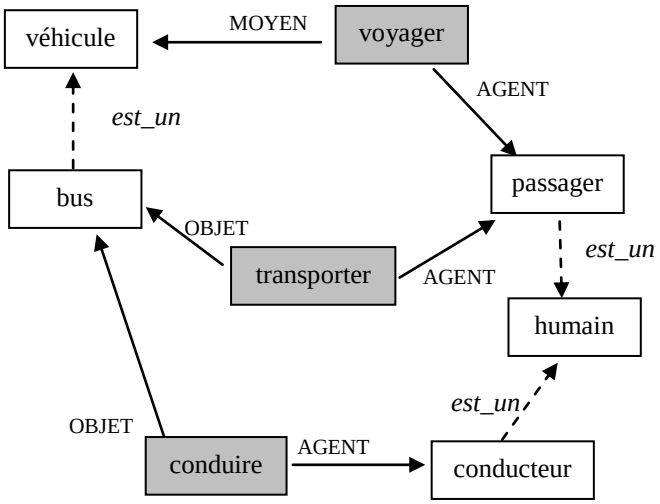


Figure 1. *Extrait d'un réseau sémantique « déplacements urbains »*

L'aspect relationnel de ce type de description, mis en évidence par la représentation graphique elle-même, doit être souligné. Les premiers réseaux ont été conçus comme modèle de *mémoire sémantique*, avec par exemple une préoccupation d'économie de stockage. Pour le valider, Collins et Quillian ont eu recours à un paradigme expérimental qui a fait flores, celui de la vérification de phrase. On propose à un sujet des questions du type *un A est un B* (ou *un A est B*) et on mesure le temps qu'il met à répondre par *vrai* ou *faux*. Ce temps est censé être proportionnel au chemin parcouru dans le réseau. Ainsi la réponse pour *un canari est jaune* sera plus rapide que pour *un canari a des plumes*, parce que le lien est direct alors que pour *avoir des plumes* il faut « passer » par l'hyperonyme *oiseau* auquel la propriété est liée. La réponse à *un canari mange* sera encore plus longue car il faut transiter par un nouveau degré d'hyperonymie : *animal*.

Ce type de description « relationnelle » a été très populaire, et en Psychologie Cognitive et en I.A. pour la représentation des connaissances[8]. On le retrouve dans des outils informatiques tels que *WordNet* [MIL 90] (initialement d'orientation psycholinguistique, utilisé couramment en TALN depuis), dont le nom anglais est assez éloquent, ou dans le *Sémiographe* développé par D. Dutoit avec dans ce dernier cas des calculs de proximité complexes entre nœuds [DUT 03].

DISCUSSION –. Dans la tradition des réseaux sémantiques (comme dans toutes les approches cognitives) il n'y a pas de coupure entre des « niveaux » linguistiques et cognitifs. L'expérience évoquée autour du concept « canari » fait évidemment appel à des connaissances encyclopédiques et les réseaux sémantiques, nous l'avons dit, ont évolué vers un modèle général de représentation des connaissances. D'un autre côté, nous avons déjà observé la nature « linguistique », pour ainsi dire « grammaticale » de certains liens, tandis que les liens d'hyperonymie/hyponymie sont souvent considérés comme relevant de la langue.

Concernant la question des primitives, en première approche on dira qu'il n'y en a pas pour les nœuds (mots-concepts) alors que les liens ont tendance à faire partie d'un inventaire fini. Ceci est très cohérent avec la représentation en réseau : il ne s'agit pas de définir « individuellement » les items lexicaux par rapport à des « concepts primitifs », mais par leurs *interrelations* : c'est le point de vue

[8] Voir le chapitre 7 pour une présentation des Graphes Conceptuels, une des formes les plus abouties pour la représentation des connaissances [SOW 84].

global, systémique, qui prime ; par contre ces relations doivent être caractérisées pour elles-mêmes. Toutefois, certains auteurs ont imaginé des ensembles de primitives telles que : CHANGEMENT, DEPLACEMENT, CAUSE etc., un nœud pouvant alors s'expanser en un réseau faisant appel à ces concepts/opérateurs élémentaires [SCH 75].

Enfin, notons que l'opération d'instanciation n'était pas considérée dans les premières versions, pour les raisons évoquées plus haut : il s'agit au départ d'un modèle d'organisation associative de la mémoire. Elle a été introduite dans les développements ultérieurs, avec une problématique générale de représentation des connaissances, notamment dans les Graphes Conceptuels de Sowa. On peut toutefois se demander si l'on ne perd pas alors (comme d'ailleurs avec l'introduction de primitives) l'idée première de réseau *associatif* qui en fait l'attrait, et la notion de *proximité sémantique* (la « topologie » du sens) qui en découle.

1.2.2. Entités, propriétés et relations

Une autre manière naturelle, et qui viendra sans doute spontanément à l'esprit du lecteur, de donner le sens du mot *oiseau* est de dire qu'il désigne un animal ayant des plumes, un bec, des ailes, capable de voler. *Doubler* désigne une action, consistant à passer d'un état 1 où l'on se trouve derrière un certain mobile, à un état 2 où on se trouve devant lui. Etc. Le schéma général est : le mot M désigne les entités (éventuellement abstraites) d'une surclasse C caractérisées par tel ensemble de propriétés, éventuellement dynamiques. Ce mode de description — que l'on pourra appeler la méthode des *Conditions Nécessaires et Suffisantes* (*CNS*) — se retrouve dans des traditions assez diverses :

– c'est un mode de définition courant des dictionnaires ;

– en psychologie cognitive, il est lié au problème de la *catégorisation* : comment l'esprit humain est-il capable de regrouper, de rapprocher, un ensemble d'expériences singulières et, en l'occurrence, de les symboliser par un mot ?

– c'est un mode traditionnel des approches logiques en philosophie du langage (dans la lignée de Russel ou Carnap) et en I.A.[9]

C'est un mode naturel mais rigide. Le fait de voler est certainement associé linguistiquement au mot *oiseau*, mais ne peut faire partie des CNS : que dire d'une autruche ou d'un oisillon ? Tout le mode associera sans doute au corbeau la noirceur de son plumage, mais que dire d'un corbeau tombé dans un sac de farine (Carnap) ou atteint d'une maladie qui le décolore ? Etc. Aussi la description en entités et propriétés ne doit-elle pas être étroitement associée à la notion de Condition Nécessaire et Suffisante. La psychologie cognitive a par exemple développé la théorie alternative du *prototype* : est associé dans notre esprit au mot *oiseau* l'image d'un oiseau typique, auquel peuvent être comparées diverses images perçues. Ou, plus abstraitement, un ensemble de propriétés dont aucune n'est absolument nécessaire, mais qui sont comparées globalement à celles de l'objet perçu (Voir [NYC 98] p. 303, [CAR 89] p. 99, [DUB 97]). La notion de prototype a été souvent adoptée par les sémantiques cognitives [VIC 04] et développée également en Intelligence Artificielle.

Une vision plus dynamique, intégrant les relations avec d'autres entités (nécessaire notamment dans le cas de verbes) peut aussi être proposée. On remarque d'abord que la sémantique d'un item lexical *mobilise, présuppose, convoque*… un certain nombre d'entités relevant d'un certain *domaine notionnel*. Ce domaine peut être plus ou moins abstrait, générique ou spécifique à un usage particulier. Et les entités présupposées par l'item devront se trouver « actualisées » dans l'énoncé ou le contexte d'énonciation. Dans ce contexte, à la fois appelé par lui-même et actualisé en situation,

[9] Voir également le chapitre 7.

l'item va *installer, poser, évoquer*… un ensemble de relations ou propriétés, éventuellement dynamiques. On appellera *schéma* associé à un item lexical l'ensemble de cette construction. Prenons un exemple lié à un usage bien spécifique : *rouler* (dans un véhicule). Cet item convoque ou présuppose l'univers du « monde de la route » et dans ce contexte un véhicule à roues (ni un avion ni une luge…) avec, de manière prototypique, son conducteur et ses passagers éventuels. Est alors évoqué le déplacement de ce véhicule, selon les procédures « standard » (pas de dérapage ou d'enlèvement à la fourrière…). Mais le modèle permet aussi de traiter certains phénomènes d'abstraction et de polysémie ; le chapitre 6 en donnera une illustration dans le domaine de la sémantique spatiale. Les *sémantiques cognitives* précédemment évoquées opèrent des descriptions de ce type, attribuant une place importante aux relations topologiques entre objets ainsi qu'à des opérations très générales apparentées à la perception (pour une version formelle, voir les travaux de J.-P. Desclés [DES 85]).

DISCUSSION–. La notion même de CNS comme celle de prototype ou de schéma fait évidemment référence à un univers de connaissances et de représentations cognitives générales. La question d'une distinction entre représentations sémantiques « linguistiques » et « cognitives » ne se pose donc pas[10]. La notion de primitive apparaîtra en TALN au niveau de la *représentation des connaissances* sur le domaine de référence (ou « modèle du monde » selon [ENJ 94]). Dans les sémantiques cognitives, on pourra avoir des *principes « génératifs »* (par exemple, les *archétypes cognitifs* de Desclés) analogues aux primitives « à la Schank ». Enfin, puisqu'elles intègrent une notion *d'entité*, ces approches sont évidemment appropriées à la description d'opérations d'instanciation et se prêtent bien à une représentation en logique —aisément opératoire pour le TALN.

1.2.3. Sémantique différentielle (ou componentielle)

Le troisième mode de représentation que nous présenterons ici se situe dans le grand courant du structuralisme. Il a été relativement peu considéré en TALN, et en particulier paraît largement étranger à la Compréhension Automatique. Pourtant il nous paraît porteur d'un point de vue et de méthodes d'analyse qu'il est difficile d'ignorer, y compris pour des traitements automatiques.

Conformément à un principe général du structuralisme (description de la langue en terme de système(s) d'oppositions entre « valeurs ») l'idée est de repérer dans un champ sémantique les différences, les traits sémantiques ou *sèmes* qui opposent les items lexicaux. Une définition plus précise du sème est proposée par B. Pottier[11] : *« Le sème est le trait distinctif du sémème [contenu, sens, d'un morphème] relativement à un petit ensemble de termes réellement disponibles et vraisemblablement utilisables chez un locuteur dans une situation de communication »*. Pour continuer un thème précédent, voyons comment B. Pottier illustre la méthode en comparant une série d'unités du champ lexical de ce qu'il appelle *« l'expérience du citadin voyageur »* (figure 2)[12]. Pour chaque mot, on indique la présence (+) ou l'absence (-) de chaque trait, ou une indétermination (≈). Autre exemple classique : le lecteur est invité à imaginer les traits différentiant les mots désignant des pièces de mobilier : *fauteuil, chaise, tabouret, table, lit*…, traits dans lesquels on pourra trouver des indications de forme, mais aussi de fonction (*pour manger*, *pour dormir*…).

[10] Voir, par exemple, les prises de positions d'un Langacker [LAN 86].
[11] Cité dans [RAS 87], p. 33.
[12] Cité dans [VIC 96] p. 46.

	sur terre	**sur rail**	**deux roues**	**indivi-duel**	**payant**	**4 à 6 pers.**	**intra-urbain**	**transport de pers.**
voiture	+	-	-	+	-	+	≈	+
taxi	+	-	-	≈	+	+	≈	+
autobus	+	-	-	-	+	-	+	+
autocar	+	-	-	-	+	-	-	+
métro	+	+	-	-	+	-	-	+
train	+	+	-	-	+	-	-	+
avion	-	-	-	≈	+	≈	-	+
moto	+	-	+	+	+	-	≈	+
bicyclette	+	-	+	+	-	-	≈	+

Figure 2. *Description sémique : l'expérience du citadin voyageur*

Il est important de bien voir le caractère relatif de telles descriptions. Les sèmes ne sont pas des « universaux » mais dépendent d'un corpus, ou comme dans la citation de Pottier, d'une situation de communication. Ainsi, dans un contexte gastronomique, on pourra opposer *bœuf* et *cheval* d'une part à *veau* et *porc* de l'autre par les traits 'viande rouge' vs 'viande blanche'. F. Rastier (op. cit.) montre aussi comment un item lexical peut être influencé par son « entourage » dont il héritera des sèmes, ou influencer en retour d'autres mots. C'est un des aspects séduisants de sa théorie. Notons encore que toutes sortes de « valeurs » peuvent constituer des sèmes, d'allure non nécessairement référentielle, concrète, comme dans l'exemple donné. Par exemple le trait 'luxe' est certainement caractéristique de *caviar*, permettant de l'opposer à *œufs de lump* ou *hareng*, provoquant l'étrangeté de *gauche caviar*, *gauche* (dans le sens politique) étant en principe porteur d'un sème 'populaire'.

DISCUSSION –. La tradition structuraliste revendique une autonomie et une spécificité du « linguistique ». Un François Rastier insistera sur le caractère proprement linguistique de l'analyse sémique, clairement différenciée pour lui du niveau « cognitif général », et de la représentation des connaissances, tout en y étant articulé subtilement : renvoyons le lecteur à la lecture des ouvrages cités de cet auteur pour une discussion fine de ce point.

La question est d'importance dans la perspective de cet ouvrage, car la coupure opérée dans la tradition différentielle nous paraît un obstacle à sa prise en compte en TAL. On remarquera ici que, sous le terme de *sémantique componentielle*, l'analyse en traits sémantiques a été portée par des courants de la Psychologie Cognitive, avec cette fois une vision beaucoup plus homogène des niveaux linguistique et cognitif. On défendra donc l'hypothèse selon laquelle aucune « muraille de Chine » ne se dresse là et que les « connaissances » que l'on dit reléguer en quelque sorte hors de l'analyse sont en fait bien présentes dans la définition des sèmes. Les *sèmes* sont empruntés à un certain « espace notionnel », la formation de concepts (au sens le plus large ici) par différenciation dans un tel espace de pensée étant un mécanisme cognitif (y compris perceptif) extrêmement général. Ce que fait la sémantique différentielle, c'est repérer de manière fine une articulation entre ce mécanisme et une activité spécifiquement langagière. Le reconnaître n'enlèverait rien à son apport spécifique en termes de méthode linguistique d'analyse, mais lui permettrait en quelque sorte de « communiquer » avec des modes de représentation « cognitive » d'une part, et des modèles de « représentation des connaissances » informatiques plus riche que des « assemblages discrets » de sèmes de l'autre.

Un autre problème pour le TALN est l'absence de représentation véritablement élaborée des « valeurs » constituant les traits sémantiques. Dans la tradition linguistique, ceux-ci sont conçus comme des objets métalinguistiques et exprimés eux-mêmes en langue. Dans les travaux plus formels, ils sont en général représentés par un simple jeu d'étiquettes « symboliques » pris dans un

inventaire fini. Une représentation plus fine est esquissée par F. Rastier sous la forme de Graphes Conceptuels, mais sans que le statut de ce mode de représentation soit très clair[13]. Ceci est problématique, car une représentation du « contenu » en termes exclusivement d'opposition, de différence, ne pourra qu'attirer le scepticisme de toute personne ayant tâté de la représentation des connaissances.

Enfin il est clair que les opérations *d'instanciation* et de *référenciation* sont étrangères à l'approche différentielle. Cette question est discutée par F. Rastier, qui revendique clairement cette « extériorité » au profit (pour faire vite) d'une simple « impression référentielle » créée par la reconnaissance d'isotopies[14]. Nous considérerons pour notre part que c'est là un « angle mort » de la théorie et qu'un travail d'explicitation de l'articulation entre sémantique différentielle et référence serait possible et nécessaire.

Pourtant, le modèle ne devrait pas être dédaigné par le TALN, dans la mesure précisément où il propose *un niveau d'abstraction* pertinent pour la description d'un certain nombre de phénomènes linguistiques. Il est par exemple à l'origine du concept d'isotopie, qui constitue un aspect important de structure textuelle, sur lequel nous reviendrons à propos de l'indexation thématique au chapitre 9. Il apporte également une réponse originale à la question des primitives. Un jeu de *sèmes* constitue bien un ensemble de « valeurs » élémentaires, auquel est rapportée la signification lexicale. Mais ce jeu est relatif à une situation de communication. L'idée nous paraît intéressante. En fait toute analyse concrète est contrainte de fixer une certain nombre d'éléments « premiers », faute de quoi on se trouve plongé dans un vaste « réseau sémantique universel » ingérable. La sémantique différentielle nous propose des critères linguistiques, à partir d'observables, pour les déterminer. Une orientation vers une *sémantique de corpus*, donc. Ajoutons qu'il existe quelques tentatives pour exploiter ce cadre théorique[15] dans des réalisations informatiques, encore peu nombreuses mais stimulantes ; citons ici sans plus de détail, comptant sur la curiosité du lecteur [PER 04], [TAN 97] ou [VAL 04].

1.3. *Calcul du sens lexical*

A priori, la question du calcul du sens lexical ne devrait pas se poser. En effet, si l'on choisit de prendre les unités lexicales comme unité minimales de sens, comme nous l'avons proposé au § 1.1, le sens de ces unités doit être le point de départ des calculs, et non un de leurs résultats. Les seules questions pertinentes à ce niveau devraient donc être les problèmes de représentation que nous venons de traiter, puisque ce sont ces représentations qui devraient être stockées dans le lexique pour être utilisées par les opérations ultérieures de calcul du sens, au niveau du syntagme et de la phrase. Mais en fait, les choses ne sont pas aussi simples. Comme nous l'avons dit, les paliers de la sémantique ne sont pas étanches, et des problèmes de calcul du sens lexical se posent à la fois en amont et en aval du niveau lexical.

D'abord, en amont de ce niveau, une analyse morphologique peut être utile dans certains cas, certes limités mais non négligeables pour autant. Nous avons vu, avec des exemples tels que *agriculteur* ou *pomme de terre*, que l'absence de compositionnalité sémantique intégrale rendait illusoire et inefficace l'analyse morphologique dans le cas général, et qu'il n'y avait pas d'autre choix que de consigner le sens de ces termes dans le lexique. Mais tout ne peut pas être dans le lexique. Notamment un certain nombre de préfixes et de suffixes restent très productifs dans la langue et sont

[13] A notre avis. Voir l'auteur pour plus de détail, notamment : [RAS 91], p. 203 et [RAS 94], p. 56.
[14] Où *récurrence de sèmes* (cf. § 3.5)
[15] Dans certains développements, plus élaborés, dont il est impossible de rendre compte ici.

à l'origine de la création permanente de mots nouveaux, non répertoriables en temps réel dans un dictionnaire. Par exemple, le préfixe *re-* ou le suffixe *–able* peuvent être appliqués à pratiquement n'importe quel verbe, produisant des effets tout à fait calculables[16]. Il faut noter que cette systématicité est limitée aux nouveaux composés : si *reschtroumpfer* veut dire de toute évidence 'schtroumpfer de nouveau', *regarder* ne veut pas dire 'garder de nouveau' ni *retourner* 'tourner de nouveau', de même qu'un *tour pendable* n'est pas un 'tour qui peut être pendu'… Les mêmes remarques valent pour les autres affixes productifs du français, comme *anti-*, *hyper-*, *-isme*, *-eur*, *-age*¸ *-ment*, etc.[17]

Mais c'est surtout en aval du palier du mot que les problèmes de calcul lexical vont se poser. En effet, il n'y a pas de correspondance biunivoque entre mots et sens, contrairement à une idée naïve, que l'exposé précédent a pu entretenir[18]. Il suffit d'ouvrir un dictionnaire pour prendre conscience que la règle est au contraire, pour chaque item lexical, d'une *multiplicité de sens*, décrite par les différentes rubriques d'une entrée de dictionnaire. C'est ce que l'on appelle la *polysémie*[19]. Ainsi, pour reprendre l'exemple du verbe *rouler*, que nous avons décrit avec son sens de déplacement dans un véhicule à roues, le Petit Robert lui consacre une vingtaine de subdivisions, correspondant chacune non pas à un sens mais à un ensemble de sens différents. Le sens 'avancer, voyager dans un véhicule à roues' se trouve décrit à l'intérieur d'une de ces subdivisions, en compagnie de plusieurs autres. On trouve d'abord 'avancer au moyen de roues, de roulettes' (ex. : *la voiture roule à 100 à l'heure*), mais aussi deux sens dit « familiers » : le sens de 'ça va' dans *ça roule*, et celui de 'travailler pour quelqu'un' que peut prendre l'expression *rouler pour quelqu'un*. Cet exemple montre d'ailleurs à quel point les descriptions des dictionnaires sont inadaptées au traitement automatique. Autant cela fait sens, dans une optique de modélisation et de formalisation, de regrouper les deux premières définitions dans une même subdivision, autant le regroupement avec les deux autres usages paraît relever d'une logique lexicographique difficilement compatible avec les impératifs du traitement automatique.

Pour en revenir à la polysémie, il faut noter qu'elle est omniprésente : elle touche toutes les classes grammaticales, depuis les noms, les verbes et les adjectifs jusqu'aux mots outils comme les prépositions ou encore les morphèmes liés comme les marques de temps verbal. Il faut aussi souligner que ce sont les unités les plus fréquentes, celles qui font partie du vocabulaire de base des locuteurs et qui sont apprises le plus tôt par l'enfant, qui sont les plus polysémiques, ce qui montre bien qu'il ne s'agit pas d'un défaut accidentel et marginal, mais bien d'une propriété inhérente au fonctionnement même de la langue.

La polysémie pose bien évidemment un problème de calcul : il faut être capable de trouver le sens adéquat de chaque occurrence de chaque mot polysémique dans les textes que l'on analyse. Et ce problème dépasse le niveau du mot, puisque c'est le contexte du mot qui permet de trouver le sens correct de chacun de ses emplois. Ce contexte peut être plus ou moins large. Ainsi pour *rouler* la donnée d'un complément d'objet suffit à déterminer son sens dans des emplois comme *rouler les hanches*, *rouler les 'r'*, *rouler un tapis* ou *une cigarette*. En revanche, dans la phrase *Je roulais sur la partie droite de la chaussée*, le verbe reste *a priori* ambigu entre le sens de déplacement dans un véhicule et celui de mouvement de rotation du corps sur le sol (rouler sur soi-même). Comme on le

16. En témoigne le mot *répertoriable* que nous venons d'utiliser et qui n'est justement pas répertorié dans les dictionnaires…
17. Pour approfondir ces questions, voir par exemple [DAL 03] et [FRA 03].
18. L'illusion peut provenir de la description d'un item dans un domaine particulier, comme *supra* dans l'exemple de réseau sémantique « transports urbains ». La réduction à un domaine de connaissance est évidemment un facteur de réduction de la polysémie. Est-elle suffisante pour se ramener au schéma 1–1 ? La question est ouverte et sera discutée au chapitre 3.
19. Du grec *polus*, *plusieurs*, et *semainen*, *signifier*.

verra plus loin dans ce chapitre, ce sont en fait des éléments contextuels extra-linguistiques qui permettent de trancher.

La polysémie des unités grammaticales présente des difficultés du même ordre. Prenons l'exemple d'un adverbe aspectuel comme *encore*, qui peut prendre des sens très divers :

- duratif : *A cette heure-ci, il est sûrement encore dans son bureau*.
- répétitif : *Il a encore raté son examen*.
- quantitatif : *Il y a encore d'autres livres dans la chambre*.
- renchérissement : *Cela complique encore davantage le problème*.
- modal : *Il est à peine au niveau, et encore !*
- notionnel : *Se taire est encore une façon de parler, quoi qu'on pense*.

etc.

Comment calculer son sens précis dans une phrase donnée ? Une étude linguistique très détaillée ([VIC 96], chapitres 5 et 6) montre que les indices co-textuels qui permettent ce calcul peuvent être répartis dans toute la phrase : type de verbe, valeur du temps verbal, nature du déterminant du groupe nominal complément, position syntaxique et portée de l'adverbe, voire même parfois l'ordre des constituants en surface. En somme, le calcul du sens d'une unité grammaticale comme *encore* ne peut pas être dissociée du calcul du sens de l'énoncé dans son ensemble.

Ainsi le calcul du sens lexical traverse d'une certaine manière tous les paliers et tous les domaines de la sémantique. On aura donc l'occasion d'y revenir à de nombreuses reprises dans cet ouvrage, non seulement au chapitre 3, qui est entièrement consacré à la modélisation de la polysémie lexicale, mais aussi aux chapitres 5 et 6 où sont traités respectivement la sémantique du temps et de l'aspect et celle de l'espace et du mouvement.

1.4. *Conclusion : quels modèles pour le TALN ?*

On le voit, la sémantique lexicale est en soi déjà un immense chapitre de la sémantique tout court. La première remarque qui résulte de notre exposé concerne la diversité des *modèles de représentation*. Celle-ci nous paraît légitime : nous sommes dans un domaine d'une grande complexité, où se côtoient et s'interpénètrent le système de la langue, des modes de fonctionnement et de représentations cognitives (penser ne serait-ce qu'au problème de la catégorisation), l'organisation des connaissances… On pourrait aussi dire que c'est tout le problème de la symbolisation, entre langue et concept, qui est en jeu.

Il est donc inévitable que plusieurs types de modèles aient été élaborés, selon différents angles, objectifs ou niveaux d'abstraction. Le travail du modélisateur, comme du concepteur d'applications TALN, est donc de choisir le mode de représentation adéquat, quitte d'ailleurs à le développer ou l'adapter. C'était l'objectif de cet exposé de développer l'éventail des modes de représentation existants, de manière significative sinon exhaustive. Reste que si cette multiplicité est légitime, un travail plus important de rapprochement, de mise en relation, d'explicitation et d'articulation des points de vue, serait le bienvenu — notamment en regard des applications, qui gagneraient à tirer parti de manière à la fois éclectique et ordonnée des différents modèles.

A cette question s'ajoute un problème de *calcul*, par *compositionnalité* d'une part et par actualisation en contexte du *« potentiel polysémique »* des items lexicaux. Ici, il sera bon de préciser les buts recherchés : modélisation ou/et application. Dans le premier cas on cherchera avant tout une

finesse d'analyse et une couverture large des phénomènes lexicaux, le TALN apportant comme toujours des moyens de formalisation, d'opérationnalisation, et donc d'expérimentation. Cette voie a été illustrée par diverses références bibliographiques dans notre texte. Mais on peut aussi avoir en vue les applications, et l'on est alors souvent conduit à opérer, implicitement ou explicitement, un certain nombre de restrictions. En particulier, on s'intéressera souvent à l'organisation conceptuelle spécifique d'un *domaine* particulier. C'est un sujet de débat — qui sera repris en détail dans le chapitre 3 — de savoir si, dans ces conditions, le phénomène de la polysémie disparaît. En tout état de cause, elle est considérablement réduite, avec de grandes conséquences pratiques : c'est ce que nous appellerons ici *l'hypothèse des microdomaines*. Mais on peut aussi, sans assumer pleinement cette hypothèse, restreindre les tâches et les corpus pour qu'elle devienne en pratique valide. C'est clairement l'option choisie en général en Compréhension Automatique de textes, et dans certains de ses dérivés en Recherche d'Information : une démarche développée et illustrée dans la troisième partie de l'ouvrage.

2. La phrase

2.1. *Caractérisation du niveau: intégration sémantique*

Deux faits caractérisent la phrase du point de vue sémantique :

– A partir de l'ensemble des « ingrédients » que sont les mots d'une phrase et le « contenu » qui leur est associé, est constitué un « tout » cohérent, structuré, possédant une sorte d'unité de sens, relativement autonome du reste du texte. Il faut aussi noter que ce calcul implique des connaissances générales sur le domaine de l'énoncé. C'est ce que nous appellerons ici *l'intégration sémantique*[20].

– La phrase fait l'objet d'une *énonciation*, par un locuteur ou scripteur. Cette énonciation n'est pas « désincarnée », mais répond à une certaine *visée*, dans un certain *contexte*. Elle réfère à une certaine « réalité extra-linguistique »[21] avec laquelle l'énonciateur entretient une certaine relation de « croyance » ou plus généralement « d'adhésion », et se situe dans une intention et un rapport communicationnels.

En fait ces deux caractéristiques *commencent* à se manifester au niveau du syntagme et se poursuivent au niveau textuel. Mais c'est bien au niveau de la phrase qu'elles jouent pleinement leur rôle, comme on va maintenant l'illustrer. Prenons l'exemple suivant :

Je roulais sur la partie droite de la chaussée quand un véhicule arrivant en face dans le virage a été complètement déporté.

– *Des liens* sont établis entre les valeurs sémantiques et référentielles associées aux différents mots, pour former une *structure* sémantique porteuse d'une *information complexe*. Dans *Je roulais*, *Je* désigne le locuteur comme **acteur** de l'action de rouler et *sur la partie droite de la chaussée* désigne le **lieu** où s'effectue ce déplacement. Et *la partie droite de la chaussée* répond également à un schéma général avec une **entité globale** (ici : la chaussée) et une sorte de **fonction de sélection** (ici : la droite, par opposition à que la gauche) d'une **partie** de cette entité.

On observe que ces liens sont en relation étroite avec des fonctions syntaxiques (sujet, complément de lieu, complément de nom, fonction adjectivale…). D'où l'idée d'une procédure exploitant ces relations syntaxiques pour constituer la structure sémantique globale à partir des mots. C'est ce que l'on appelle le principe de *compositionnalité syntaxique* sur lequel nous allons revenir.

– Il y a une différence importante entre la phrase elle-même et les autres syntagmes qu'elle contient, c'est l'idée de *saturation*. Par exemple l'expression *la partie droite* attend un complément : partie droite de quoi ? *sur la partie droite de la chaussée* fait déjà une certaine unité de sens,

[20] Expression librement empruntée à J. Caron ([CAR 89], p. 158).
[21] Conçue, non « objective » ou « ontologique ».

désignant une localisation, mais on se demande : que se passe-t-il dans ce lieu ? La proposition subordonnée *quand un véhicule…* ne peut évidemment exister sans la principale. Seule la phrase dans son ensemble ne porte pas ce type d'attente. En fait, cette remarque doit être prise *cum grano salis* : une phrase ne se suffit pas forcément à elle-même pour être complètement compréhensible (à cause des phénomènes d'anaphore, en particulier), et, à l'inverse, une partie de phrase peut se suffire à elle-même (certains groupes circonstanciels ne sont pas nécessaires à la compréhension de la proposition sur laquelle ils portent).

– On retrouve les effets de contexte sur les valeurs lexicales. Par exemple *sur* est interprété comme spatial à cause du syntagme nominal qu'il commande. Plusieurs choix sont déterminés par le contexte « accident » (connu ou inféré pour des raisons de plausibilité), à commencer par celui du sens de *roulais*. Il y a là aussi une forme de *cohérence*.

– Des connaissances générales sont mobilisées, et diverses inférences sont effectuées au cours de la compréhension du texte. Par exemple *la partie droite* peut être *la partie située à droite* ou *la partie rectiligne*. Le plus plausible est la première interprétation. Ici intervient sans doute la connaissance du code de la route et le fait que l'on imagine bien que le locuteur veut mettre en avant son respect dudit code. De plus, si notre interprétation est la bonne, la phrase exprime effectivement l'idée de conformité au code, qui n'est absolument pas dite explicitement dans le texte ! En fait, dans le sens produit, nous avons donc des « ingrédients » donnés à la fois par le texte et par des « connaissances » générales. Et il peut être difficile de les démêler.

C'est donc tout un ensemble d'*opérations de construction du sens*, de nature syntaxico-sémantique, qui interviennent au niveau de la phrase et qui contribuent à cette intégration sémantique, que nous allons maintenant détailler quelque peu. En TALN, ces opérations seront réalisées dans le cadre d'un des modèles de *représentation et de traitement des connaissances* déjà évoqués à propos des représentations lexicales, le plus souvent les *Graphes Conceptuels* ou la *Logique*. Des méthodes générales, permettant de décrire l'enchaînement de ces traitements sémantiques ont été élaborées,que nous présenterons au chapitre 7.

2.2. *Instanciation*

Nous aurons d'abord ce qui peut être caractérisé comme des opérations *d'instanciation*, bien connues des informaticiens et des logiciens. Alors que les éléments lexicaux évoquent des catégories (ou des *types*, en langage plus informatique), les syntagmes complets qui constituent les phrases évoquent généralement des exemplaires de ces catégories (des *instances*). Ainsi les noms *voiture* et *barrière* désignent des catégories d'objets et le verbe *percuter* une catégorie d'actions, alors que dans la phrase :

Une petite voiture blanche a percuté violemment la barrière de sécurité.

les syntagmes *une petite voiture blanche* et *la barrière de sécurité* désignent des objets, et le syntagme *a percuté violemment* un événement, particuliers[22].

Ces opérations d'instanciation se produisent à certains niveaux bien précis de l'arbre que l'on associe généralement à la structure syntaxique de la phrase : ainsi dans l'embranchement *voiture* → *voiture blanche* → *petite voiture blanche* → *la petite voiture blanche*, ce n'est qu'avec l'ajout du déterminant *la* que se produit l'instanciation, *voiture blanche* et *petite voiture blanche* évoquant des catégories, au même titre que le nom *voiture* lui-même. De même, c'est avec l'ajout de la marque de temps verbal que s'opère l'instanciation dans la série : *percuter* → *percuter violemment* → *a percuté violemment*.

[22] Cette opération d'instanciation pose donc, du point de vue du TAL, des problèmes spécifiques de représentation et de calcul qui ne se posaient pas vraiment au niveau précédent, celui du mot.

Ainsi, l'instanciation est généralement le résultat de la combinaison d'unités linguistiques de deux sortes[23] :

– un ensemble d'unités lexicales (*petite* + *voiture* + *blanche*, *percuter* + *violemment*) qui contribuent à préciser et à qualifier la classe sémantique dont un exemplaire va être extrait. Ici intervient un premier calcul, de type compositionnel, sur lequel nous reviendrons au § 2.3.

- une unité grammaticale (déterminant, marque verbale) qui est le support de l'opération d'instanciation proprement dite, permettant ainsi au syntagme de jouer son rôle référentiel d'évocation d'une entité ou d'un événement extralinguistiques.

A vrai dire, les choses ne sont pas toujours aussi simples. Ainsi, les déterminants ne servent pas toujours à construire un exemplaire d'une classe, notamment quand ils prennent des valeurs dites *génériques*, comme dans *L'homme est un roseau pensant* ou encore *Un caniche est moins encombrant qu'un saint-bernard*. Inversement, certains mots, comme les pronoms personnels et les noms propres, désignent à eux seuls des instances, qu'ils soient utilisés isolément ou têtes de syntagmes plus importants (comme dans *Ce pauvre Albert* ou dans *Je, soussigné Dupont Albert*).

Indépendamment de cette question des *génériques*, le problème majeur — en particulier en TALN — consiste à *identifier* les entités et l'événement[24] pertinents, et de caractériser leurs relations. Pour chaque entité évoquée, il faut notamment déterminer s'il s'agit d'une entité nouvelle, introduite pour la première fois dans le discours, ou si, au contraire, elle a déjà été construite dans une phrase précédente. Autrement dit, comme nous le verrons plus en détail au chapitre 4, il faut maintenir une liste des entités déjà créées, et, à chaque phrase, mettre à jour cette liste. Une entité déjà présente dans le « contexte du discours » peut en effet être *reprise*, *évoquée*, une nouvelle fois, par un pronom ou une description définie (le X, ce X…). On parlera *d'anaphore pronominale* et *nominale* respectivement. Exemple :

Une petite voiture blanche a percuté violemment la barrière de sécurité. **Elle** a été endommagée mais **la barrière** est intacte.

Cette question est celle du *calcul de la coréférence* : sa portée se situe clairement au niveau du texte (cf. § 3) et nous l'étudierons en détail au chapitre 4.

Notons également que la décision de *création* d'une entité à partir d'une unité nominale n'est pas toujours très simple : reprenons la phrase *Je roulais sur la partie droite de la chaussée.* Cette phrase comporte trois unités nominales : *Je*, *partie* et *chaussée*. Cela ne veut pas dire que chacune d'elle va automatiquement conduire à créer une entité. Bien entendu, à *je* va correspondre une entité de type humain représentant le locuteur. Mais, on peut décider, pour des raisons de cohérence de typage, d'introduire une deuxième entité présente dans *je roulais*, bien que non exprimée explicitement : la voiture du locuteur. Ce n'est en effet pas le locuteur qui roule, mais sa voiture (phénomène de *métonymie*). En revanche, pour *la partie droite de la chaussée*, on peut légitimement se demander s'il faut créer deux entités, la partie droite et la chaussée reliées entre elles par une relation partie-tout, ou s'il ne vaut pas mieux s'en tenir à une seule entité, la chaussée, et spécifier qu'il s'agit de la partie droite dans la relation de localisation qui relie cette chaussée à la voiture du locuteur.

---

[23] Les théories syntaxiques prennent généralement ce fait en compte. Ainsi, dans la théorie chomskyenne dite *Xbar*, *la petite voiture blanche* est un syntagme nominal de type N" (N-bar-bar), alors que *voiture blanche* et *petite voiture blanche* sont de type N' (N-bar). Dans certaines variantes du générationalisme, c'est même le déterminant qui est considéré comme la tête du syntagme nominal complet, renforçant ainsi l'importance de cette distinction.

[24] Il faut souligner que nous nous limitons dans cette présentation à des procès, « événementiels », les plus typiques. Il existe bien d'autres formes de phrases, qui ne mettent pas en œuvre des événements, mais d'autres types de procès, par exemple des *états*, conduisant à d'autres formes de représentation. Ainsi dans la phrase *La voiture était blanche*, il n'y a ni création d'entité, ni construction d'événement, mais simplement l'assignation d'un attribut 'couleur' à une entité déjà existante.

REMARQUE —. Il faut noter que le contexte joue dans certains cas un rôle déterminant. Pour reprendre l'exemple de *Je roulais*, il n'y a pas à proprement parler de conflit de type, puisqu'une personne peut aussi rouler sur elle-même. Il faut donc faire appel au contexte pour éviter cette interprétation, et à un contexte assez fin : il ne suffit pas d'invoquer le « domaine » des accidents de la route, puisqu'il peut arriver justement à la suite d'un accident, qu'une personne (conducteur de deux roues ou piéton) roule effectivement sur elle-même sur la chaussée. On notera ici *l'anticipation* du lecteur, qui, selon toute vraisemblance, n'envisagera même pas cette hypothèse : c'est qu'il s'attend à ce que le rédacteur décrive d'abord la situation *avant accident*, et s'attend donc à ce que le locuteur se décrive comme conducteur d'un véhicule.

2.3. *Relations argumentales et compositionnalité*

Le deuxième type d'opérations concerne les relations entre les entités et les événements évoqués dans la phrase. C'est en effet une des fonctions sémantiques essentielles de la syntaxe de la phrase que de distribuer les rôles des différents acteurs de la *« saynète »* qui est décrite par la phrase, pour reprendre une métaphore qui remonte au moins à Tesnière.

Ce sont d'abord et avant tout les constructions verbales qui sont le support de ces opérations, en permettant de répondre à des questions du type : *qui fait quoi*, *à qui*, *avec quoi*, etc. La détermination de ce que l'on appelle les *relations actancielles* constituent donc une part importante de la sémantique de la phrase. Elle repose sur toute une série de marques grammaticales et syntaxiques : prépositions, ordre des mots, diathèse (voix active, passive, réflexive, …), mais aussi sur les éléments lexicaux. Il n'y a en effet pas de correspondance biunivoque entre une fonction grammaticale (sujet, objet, …) et une relation actancielle donnée, même à voix constante. Par exemple, dans les trois phrases suivantes :

Je n'avais pas fermé la porte de ma voiture

Ce dispositif électronique ferme toutes les portes de la voiture

La porte avant droite fermait mal

on trouve en position sujet trois rôles actanciels différents : agent, instrument et siège du procès. Il s'agit pourtant dans les trois cas du même verbe à la voix active : c'est, de fait, l'analyse sémantique du groupe nominal sujet et la présence ou non d'un groupe objet qui permettent de déterminer le rôle actanciel en question.

Des relations argumentales du même genre peuvent aussi porter sur d'autres unités que les verbes : des noms (*Le transport des voitures par camion*, *la bicyclette de Pierre*, etc.), des adjectifs (*rouge de colère*, *un bidon plein d'essence*, etc.). Là encore, la correspondance entre construction syntaxique et rôle sémantique est loin d'être simple. Ainsi dans *l'agression de Pierre*, Pierre peut être aussi bien l'agresseur que l'agressé : seul le contexte, et parfois un contexte très éloigné, peut permettre de lever l'ambiguïté.

Ces relations argumentales jouent un rôle clé dans la construction du sens de la phrase. Sous sa forme la plus simple, ce « sens » consiste en une « configuration » dont le centre est une instance d'événement, auquel sont reliées des entités par des relations actancielles de type agent, patient, instrument, etc. Par exemple pour la proposition *Un camion a percuté la petite voiture blanche*, nous aurons une structure comprenant :

– un événement E instance de la classe 'percuter' ;

– une entité C instance de 'camion', dans la relation 'agent' avec E ;

– une entité V, instance de 'voiture', dans la relation 'patient' avec E ;

– l'attribution de propriétés 'être blanc' et 'être petit' à V.

Notons qu'il ne s'agit là bien sûr que de l'ossature de la représentation, qui devra être complétée par des relations entre propositions — par exemple de simultanéité, voire de causalité, dans *Un camion a percuté la petite voiture blanche qui arrivait à grande vitesse* —, de coréférence (ou identification des entités, précédemment évoquée), ou encore de cadrage, exposées *infra*.

Le calcul opère essentiellement par compositionnalité : à partir des relations actancielles et syntaxiques d'une part, des représentations associées aux mots de l'autre, une représentation est calculée pour des syntagmes de plus en plus complexes : à partir de 'voiture', 'blanche' et 'petite, on aura 'petite voiture blanche', concept type, instancié dans 'la petite voiture blanche' ; puis, à partir de 'un camion' (instance de 'camion') et 'a percuté', les relations actancielles conduisent à la structure ci-dessus. Nous verrons au chapitre 7 comment formaliser ce calcul (complexe) dans un cadre de Compréhension Automatique de textes.

S'il s'agit bien du mécanisme « central » au niveau de la phrase, les limites de la compositionnalité doivent cependant — ici comme à propos du lexique —être relevées[25]. L'exemple de l'introduction d'une entité 'voiture' pour représenter *Je roulais* en témoigne. Comme on peut le constater, le tout (*Je roulais*) est plus que la somme des parties (*Je + roulais*). Il s'agit d'un problème très général, qui se pose pour des combinaisons de toute nature. Prenons un autre exemple. On trouve dans un constat la phrase *J'ai glissé à cause d'un brutal verglas* : c'est clairement la perception du rédacteur, qui est brutale, pas le verglas. Une méthode pour résoudre ce type de problème a été bien systématisée par Pustejowski [PUS 95] : on décèle un conflit de type entre deux unités reliées entre elles (ici : *brutal* et *verglas*), et on le résout en cherchant dans la « situation » évoquée un « attribut » susceptible de recevoir la qualification problématique.

Il faut aussi souligner le lien étroit entre ces problèmes de compositionnalité au niveau de la phrase, et la question de la polysémie abordée en section 1. Dans de nombreux cas, comme justement celui de *Je roulais*, on peut considérer que le problème est celui de la désambiguïsation d'une unité lexicale (en l'occurrence le verbe *rouler*, dont on peut distinguer, entre autres les deux acceptions 'rouler sur soi-même' et 'se déplacer à l'aide d'un véhicule à roues'. Notons également de nouveau des effets importants du *contexte* dans l'interprétation des liens entre unités lexicales. Par exemple, l'expression *la voie de droite,* doit être interprétée comme *la plus à droite, par un mécanisme dont on peut discuter la nature exacte, mais clairement d'ordre pragmatique (cf. section 4).*

2.4. *Repérage, cadrage et opérations énonciatives*

Il s'agit d'un ensemble d'opérations qui ont en commun de commencer au niveau de la phrase, mais d'être généralement d'une portée plus grande que la phrase, et que l'on retrouvera donc en section 3 (le niveau du texte). Les unes servent à situer la « saynète » évoquée par la phrase, que ce soit dans l'espace et dans le temps, s'il s'agit d'une scène décrivant un événement, ou dans un espace notionnel particulier, dans le cas d'une proposition plus abstraite. Les autres servent à préciser le point de vue de l'énonciateur par rapport à son dire : comment présente-t-il la scène en question, quels éléments met-il en relief, dans quelle mesure prend-il en charge sa véracité, etc. Chez Culioli[26] les deux types sont appelés des opérations de *repérage*, avec un repère spatio-temporel pour les premières, et un repère intersubjectif centré sur l'énonciateur pour les secondes. Toutes ces opérations sont portées principalement par des tournures syntaxiques et des morphèmes grammaticaux :

[25] Voir [NAZ 98] pour différentes études sur ce thème, sous l'angle du TALN.
[26] Cf. [CUL 90], p. 127-130.

marqueurs aspectuo-temporels, marqueurs modaux, marqueurs de thématisation et de focalisation, etc.

En ce qui concerne le repérage spatio-temporel, il faut noter une dissymétrie entre l'espace et le temps. Comme on l'a vu ci-dessus, l'opération d'instanciation d'un événement s'effectue, en règle générale, par l'intermédiaire de la marque de temps verbal. On a donc, presque automatiquement, un repérage aspectuo-temporel au niveau de la phrase, qui s'opère en même temps que l'instanciation : en quelque sorte, on ne peut évoquer un événement sans commencer à le repérer temporellement, même si ce premier repérage, comme on aura l'occasion de le voir en détail (chapitre 5), est généralement très partiel. En revanche, on n'a rien de tel pour le repérage spatial : on peut, bien entendu, situer dans l'espace l'événement décrit par une phrase à l'aide notamment de circonstanciels (groupes prépositionnels ou propositions subordonnées), mais il n'y a pas de marque grammaticale obligatoire dans chaque phrase comme pour le repérage temporel[27].

On peut faire la même remarque pour les opérations de repérage intersubjectif, qu'il s'agisse des modalités ou des relations de thématisation et de focalisation. Ces opérations se caractérisent par le fait que des marques de toute nature peuvent contribuer à les spécifier. Par exemple, une modalité épistémique comme la plausibilité peut s'exprimer de manières très diverses :

– *La voiture a vraisemblablement/sans doute/certainement dérapée* : adverbe dit « de phrase ».

– *Il est probable/vraisemblable que la voiture ait dérapé* : construction propositionnelle complexe.

– *La voiture a dû déraper* : verbe modal.

etc. De même pour la thématisation :

– *Quant à Jean, il n'a heureusement pas été blessé* : syntagme prépositionnel

– *En ce qui concerne Jean, il n'a heureusement pas été blessé* : proposition subordonnée

– *Jean, lui, n'a heureusement pas été blessé* : pronom en apposition

etc. Et la focalisation :

– *C'est Pierre qui conduisait la voiture* : construction clivée

– *La voiture était conduite par Pierre* : voix passive

etc.

C'est d'ailleurs souvent la combinaison de plusieurs marques, réparties un peu partout dans la phrase, qui produit la propriété globale en question. Ainsi, c'est la combinaison de l'imparfait, d'un verbe exprimant un événement « ponctuel », et un complément de temps de type prospectif qui fait que la phrase suivante exprime le mode 'irréel' :

Une minute plus tard j'étais écrasé par la chute de l'échafaudage

Pour chacun des grands champs mentionnés (aspectuo-temporalité, modalité, thématisation, focalisation, etc.), il faut donc mener des calculs spécifiques sur l'ensemble de la phrase : il est impossible de s'en tenir à un calcul local centré sur le verbe ou tel autre syntagme. Nous verrons tout cela en détail en ce qui concerne l'un de ces domaines : l'aspectuo-temporalité (chapitre 6). Et soulignons de nouveau que ces phénomènes ne se jouent pas complètement au niveau de la phrase, mais au niveau du texte.

[27] Nous limitons ici nos observations au français. La situation n'est pas la même dans toutes les langues : s'il semble qu'aucune langue n'impose un repérage spatial, beaucoup de langues, comme le chinois, n'ont pas non plus de marques verbales temporelles. La seule contrainte universelle porte sur l'aspect [COM 76], qui semble devoir toujours accompagner l'opération d'instanciation d'un événement.

3. Le texte

3.1. *Structure et cohésion textuelle*

Pas plus que les mots n'existent de manière isolée, mais (presque) toujours au sein d'unités structurées que sont les syntagmes — et singulièrement de cette unité primordiale que constitue la phrase — les phrases n'apparaissent seules et solitaires, mais au sein d'unités plus vastes que sont les textes[28]. Et les textes, loin d'être de simples successions de phrases, possèdent leurs propres formes d'organisation. C'est, avant même tout discours théorique, une évidence empirique : un « bon » texte possède une « structure » interne qui permet au lecteur d'assembler un ensemble d'idées ou d'impressions en un tout cohérent.

Une métaphore chimique, à prendre avec la distance requise, nous permettra d'ancrer cette idée d'une structure « au delà de la phrase ». Nous avons déjà, de manière imagée, dit que les morphèmes portaient en quelque sorte des atomes de sens. Le regroupement en mots puis en phrases constitue alors des molécules de sens selon divers procédés évoqués dans les sections précédentes, dont les sous-syntagmes seraient des radicaux plus ou moins saturés. Un texte serait alors comme une macromolécule de la chimie organique, un très long assemblage de ces molécules sémantiques ; une molécule qui, comme un ADN, possède ses plis et replis, sa disposition spatiale et ses liaisons secondaires, responsables pour une grande part d'une *fonction* biochimique[29].

La seconde remarque qu'il convient de faire avant d'entrer dans le vif du sujet, est que cette structure est en fait multiforme et, à la vérité, particulièrement riche. On parlera souvent de *cohésion textuelle* et de ses divers *facteurs*. Certains apparaissent déjà au niveau syntagmatique, mais souvent se déploient véritablement au sein du texte. Il faut bien voir également qu'au-delà d'une simple « impression de cohérence » c'est la question proprement sémantique de l'organisation de « l'information » portée par le texte, et de son appréhension par le lecteur, qui est ici en ligne de mire. Et il faut se souvenir que le texte est *l'unité fonctionnelle* par excellence : son organisation dépendra donc fortement du genre (narratif, discursif, poétique, avec toutes sortes de ramifications) et contribuera de manière décisive à la fonction spécifique de ce genre. La diversité des structures et de leur agencement en découle. Nous envisageons maintenant quelques-unes de ces structures ou facteurs de cohésion.

3.2. *Coréférence*

C'est un des facteurs de structuration les plus évidents. Nous nous plaçons ici dans la perspective selon laquelle des expressions linguistiques — en particulier mais non exclusivement nominales — évoquent, réfèrent à des « objets »[30] (entités ou événements, concrets ou abstraits…) extralinguistiques, relevant d'un certain domaine de discours. On dira alors que deux expressions *coréfèrent* si elles renvoient au même « objet ». La figure 3 illustre ce mécanisme. On voit en particulier qu'une même entité peut être reprise plusieurs fois, sous diverses dénominations : se forme ainsi ce que l'on appelle une *chaîne de coréférence*. L'usage est de marquer par un même

[28] En accord avec un usage discutable mais répandu, nous parlerons indistinctement de « texte » ou « discours » pour évoquer ce palier de la sémantique. Avec une nuance qui serait de désigner plutôt par le premier les productions et par le second la manière dont elles sont produites.

[29] On trouvera une élaboration plus fine de la métaphore biochimique dans [VIC 96], chapitre 9.

[30] Objets d'un monde conçu, et non directement du monde « sensible », peuplé de Père Noël, de beauté convulsive et de tiers-mondisme autant que de tables et de chaises.

indice (i, j, k…) les éléments d'une chaîne. On notera les deux modes principaux de reprise : par un pronom ou un groupe nominal. Observons également que l'opération de coréférence peut lier un syntagme nominal et un procès verbal, comme dans le constat (*le choc*).

Du point de vue textuel, la présence de telle chaîne assure une forme de suivi du texte, à travers « les objets dont on parle », chaque nouvelle mention d'une entité apportant de nouvelles informations : cumulatives en quelque sorte, dans le cas d'un texte informatif (comme la dépêche d'agence), ou temporellement organisées dans une narration. À ce titre c'est un élément essentiel dans le processus de compréhension. Le lecteur peut imaginer l'impression d'incohérence, de « décousu » qui émanerait d'un discours sans suite référentielle, un discours où une pléthore d'objets et d'événements seraient introduits sans jamais être repris. Pour ces raisons, la détection des chaînes de coréférence est une tâche considérée aujourd'hui comme cruciale pour les applications TAL (cf. chapitres 7 et 8). C'est aussi un joli problème de modélisation linguistique auquel nous consacrerons le chapitre 4.

Le matin du 16 Avril, le docteur Bernard Rieux $_i$ sortit de son $_i$ cabinet et buta sur un rat mort $_j$, au milieu du palier. Sur le moment, il $_i$ écarta la bête $_j$ sans y prendre garde et descendit l'escalier. Mais, arrivé dans la rue, la pensée lui $_i$ vint que ce rat $_j$ n'était pas à sa place et il $_i$ retourna sur ses $_i$ pas pour avertir le concierge $_k$. Devant la réaction du vieux M. Michel $_k$, il $_i$ sentit mieux ce que sa $_i$ découverte avait d'insolite. La présence de ce rat $_j$ lui $_i$ avait paru seulement bizarre tandis que, pour le concierge $_k$ elle constituait un scandale. La position de ce dernier $_k$ était d'ailleurs catégorique : il n'y avait pas de rats dans la maison.

a. *La Peste : chaînes sur B. Rieux, le rat et le concierge.*

Je venais de freiner derrière B lorsque la 205 (532 JAL 75) m'a percuté $_i$ l'arrière. Bien qu'à l'arrêt lors du choc $_i$ j'ai été poussé sur B.

b. *Un constat d'accident de la route*

Bridgestone Sports Co. $_i$ said Friday it $_i$ has set up a joint venture $_j$ in Taiwan with a local concern and a Japanese trading house to produce golf clubs to be shipped to Japan. The joint venture $_j$, Bridgestone Sports Taiwan Co. $_j$, capitalized at 20 million nex Taiwan dollars, will start production in January 1990 with production of 20,000 iron and "metal wood" clubs a month.

c. *Une dépêche d'agence*

Figure 3. *Chaînes de coréférence*[31]

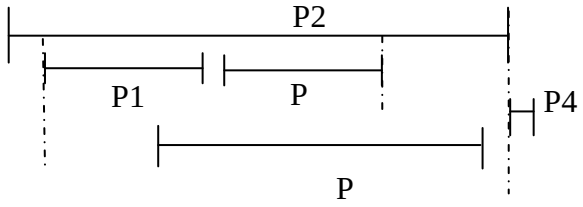
3.3. *Temporalité*

Un autre type de structure aisément repérable, particulièrement dans les textes narratifs, est constituée des rapports de succession/simultanéité entre les événements relatés. Ce type de relations a déjà été évoqué au niveau de la phrase, mais se poursuit à l'évidence au niveau du texte dans son

[31] Les textes proposés comme exemples dans cette section ont été présentés dans le premier chapitre.

entièreté. Pour les représenter, nous pouvons associer à chaque *procès* (expression d'un événement ou d'un état stable) une *période* (ou *intervalle*) pendant lequel « il se produit ». La structuration temporelle d'un texte sera caractérisée par un *chronogramme* visualisant les relations temporelles entre ces périodes. Une théorie, permettant effectivement de calculer des chronogrammes à partir du texte même, sera présenté au chapitre 5.

La figure 4 donne un exemple de chronogramme[32] pour un constat d'accident extrait de notre corpus. Cinq procès, notés P1…P5 sont définis (les propositions complétives *pour repartir* et *pour prendre de l'essence* ne sont pas prises en compte) et des contraintes temporelles peuvent être établies : P1 et P3 se situent pendant P2. P4 est « ponctuel » (c'est-à-dire qualitativement plus court que les autres) et termine P2. P5 s'achève avant P4 et débute après le début de P1. Un certain nombre de positions relatives et les durées restent indéterminées.



Je m'engageais dans une file de station service (P1). La pompe étant en panne (P2), je reculais (P3) pour repartir lorsque j'ai heurté le véhicule B (P4) qui s'était engagée également dans la même file (P5) pour prendre de l'essence.

Figure 4. Cohérence temporelle : chronogramme

3.4. *Structures du récit*

La question est ici de repérer des formes d'organisation du *récit*. Une structure complémentaire en quelque sorte des relations temporelles, dans laquelle interviennent des schémas culturels « de haut niveau », fortement « codés ». Un exemple classique et emblématique est fourni par l'analyse des contes populaires (cf. [PRO 28]). On y trouve en effet des schémas assez stéréotypés associant :

– des personnages, des figures : le bon, le méchant, le sauveur, la princesse, le roi (le détenteur du pouvoir)…

– une intrigue répondant à une succession du type : situation initiale paisible - épreuve (matérielle, morale…) – lutte – résolution (succès) – récompense – retour une situation paisible.

Mais ce que l'on pourrait appeler cette « macro organisation narrative » n'est pas spécifique des contes et apparaît, de manière éventuellement un peu moins figée, dans d'autres productions littéraires romanesque[33] (Voir par exemple [ECO 85]). On la retrouve jusque dans des textes aussi simples et courts que des constats d'accidents ! Avec un schéma finalement assez proche de celui des contes : situation initiale (je roulais dans telles circonstances) – événement imprévu (dû à l'adversaire) – choc – éventuellement, suites : dégâts, venue des gendarmes… Sur l'exemple de la

[32] Dans une version simplifiée par rapport aux représentations décrites au chapitre 5.

[33] Et autres formes culturelles et artistiques, par exemple dans toute une production cinématographique hollywoodienne, avec notamment une variante « dramatique » du schéma ci dessus : la 'résolution' du conflit est en général apparente et provisoire, et on a un cycle de deux ou trois occurrences de la chaîne 'épreuve – lutte – résolution (apparente)'.

figure 4 : P1 — P2 & P3 — P4 & P5 (pas de phase 'suite'). Le schéma complet apparaît dans d'autres constats. Malgré la difficulté évidente, liée au caractère éminemment culturel de telles structures, il existe diverses tentatives de formalisation et d'opérationalisation informatique, dont [SAB 00] se fait l'écho.

On pourra considérer qu'il s'agit d'une question « pragmatique » plutôt que « sémantique ». Nous discuterons dans le prochain chapitre de cette dichotomie, mais notons ici qu'est effectivement en jeu, selon nous, une effet sémantique que l'on pourrait qualifier « de second niveau », dans la lignée d'un Barthes ou d'un Eco[34]. La compréhension du texte inclus la reconnaissance du schéma culturel sous-jacent.

3.5. *Isotopies*

Nous abordons maintenant un type de structure assez différente dans la mesure où il s'appuie directement sur une sémantique lexicale componentielle. De manière générale, *l'isotopie sémantique*[35] est définie comme récurrence de sèmes, c'est à dire la reprise, au long du texte ou d'un fragment, d'un même trait sémantique ou ensemble de traits.

Les textes poétiques fournissent de bons exemples de ce phénomène[36]. Considérons par exemple le poème de Pierre Jean Jouve déjà évoqué au chapitre 1 (figure 5). Il est difficile de se faire une idée très précise, de ce qui serait « le sens » du texte : la dimension « référentielle » apparaît de prime abord assez décousue, et il faut un ensemble de lectures, avec un jeu assez libre d'association pour l'approcher. Par contre on perçoit immédiatement certains facteurs de cohérence — qui sont autant d'indices pour une compréhension plus approfondie — à travers différents champs lexicaux que l'on peut effectivement analyser en termes d'isotopies.

Ainsi, on pourra relever la récurrence d'un sème (dit *générique*) 'couleur', spécialisé en deux sèmes (*spécifiques*) 'vert' (*vert*, répété plusieurs fois, *plante, herbu*) et 'rouge' (*rouge, sang*, répété, *brique*), d'où ressort une impression visuelle de « contraste coloré », complétée toujours dans le domaine « physique » par la récurrence d'un trait 'matière' (*émail, épaisseur, velours, brique*). Un autre type de récurrence porte sur un sème 'religieux – magique', particulièrement dans la seconde strophe (*magies*…), ou encore 'histoire – pouvoir' dans la première : *illustres*… Etc. Autant de départs pour des parcours interprétatifs enchevêtrés.

D'UNE VILLE

Un objet ***rouge*** en émail **vert**.
Un émail de ***sang rouge*** et de **plante** et de larme
Sur *d'illustres* canaux de magique épaisseur
Où des *tours* éperdues dressent des beautés *d'armes*

[34] Cf. [BAR 85], particulièrement le texte intitulé *Introduction à l'analyse structurale des récits*), pour le premier, le § 4.4 pour le second.
[35] D'autres notions d'isotopie, récurrence d'autres éléments ou valeurs linguistiques, peuvent être définies.
[36] Mais le phénomène ne saurait en aucun cas être assigné à ce seul genre. Nous allons voir qu'il joue un rôle essentiel dans la structure thématique de tous types de textes.

Règnent sur le corps vieux des silences **herbus**

Des magies de **vert** noir de **vert** vieux de **vert vert**
Des velours de ***sang*** précieux du Christ et de charmes
Des cryptes ivres du cadavre très pieux
Des soleils espagnols sur une unique ***brique***
Et musiques de ponts et anges chaleureux :

Des meurtres des amours les résidus *antiques*.

Figure 5. *Isotopies*

3.6. *Structuration thématique*

La problématique dont il va être maintenant question est fortement motivée par des applications documentaires. Les textes prioritairement concernés ici sont de type « expositif » : ouvrages didactiques, articles journalistiques de synthèse, rapports administratifs, *etc*. Une question importante lorsque l'on consulte ou prend connaissance de textes de ce type est de savoir « de quoi cela parle », et si possible plus finement : « de quoi parle telle ou telle partie, où trouver tel type d'information ou d'analyse ». On voit immédiatement les enjeux en termes d'ingénierie documentaire, dans le grand champ de la « recherche d'information ».

Plus précisément, la question est de repérer une organisation en « unités thématiques », c'est-à-dire en segments textuels relativement homogènes quant au sujet, ou thème, traité : définir précisément ce qu'est exactement un thème ou sujet est évidemment problématique, et nous nous contenterons ici de l'idée intuitive que peut s'en faire tout lecteur, en constatant *qu'un texte bien formé doit faire apparaître ce type d'unités*. D'un point de vue théorique, la question est de savoir comment se manifeste cette homogénéité, et d'un point de vue pratique, comment faire pour la repérer en associant des « valeurs » caractérisant ce contenu dans sa généricité. Plusieurs types de marques de segmentation et d'organisation coexistent, certaines étant posées intentionnellement par le rédacteur, d'autres étant la simple conséquence de la structuration logique de son exposé. Voici quelques exemples.

– *L'organisation logique et la mise en forme matérielle matérielle (spatiale) du texte* : titres et sous-titres, sections et sous-sections, jusqu'au simple découpage en paragraphe, jeux de retraits, de « puces », de fontes… [VIR 01]

– *Densité lexicale* : l'unité thématique va fréquemment se manifester par une récurrence forte (« densité ») de certains mots, ou de mots corrélés dans un certain champ lexical. Des méthodes statistiques seront ici évidemment pertinentes, mais on peut aussi clairement réinterpréter ces phénomènes en termes *d'isotopie*.

– *Marques linguistiques explicites*. Il existe un certain nombre de dispositifs linguistiques pour introduire un thème « local ». Les exemples les plus évidents sont des expressions prépositionnelles telles que : *en ce qui concerne, pour ce qui est de, quant à, à propos de*… Mais un simple positionnement en tête de phrase et de paragraphe va donner un poids thématique à une expression. Et on constate que les expressions ainsi mises en relief ont une portée qui va souvent être de quelques phrases : c'est la notion de « cadre de discours » de M. Charolles [CHA 97].

L'exemple de la figure 6[37] illustre plusieurs de ces mécanismes : densité de termes politiques dans le premier paragraphe, structure énumérative marquée par des changements de paragraphes et des

[37] Il s'agit du résultat d'un traitement automatique [WID 04].

expressions spatiales introductrices de cadres. Nous reviendrons sur ces questions et leur application en Ingénierie Documentaire dans les chapitres 9 et 10. A cette structure thématique, il conviendrait d'adjoindre la structure *rhétorique* — son dual en quelque sorte — qui vise à mettre en relation argumentative, plutôt que de segmenter. Dans la figure 6, on repèrera une figure *loi générale – cas spécifiques* entre le premier paragraphe et l'énumération qui suit, ainsi qu'une relation *similarité* (opposée à *contraste*) entre ces trois fragments. L'article du *Monde* donné en exemple au chapitre 1 présente également une argumentation nettement marquée par des dispositifs textuels, et nous ne développerons pas cet aspect plus avant.

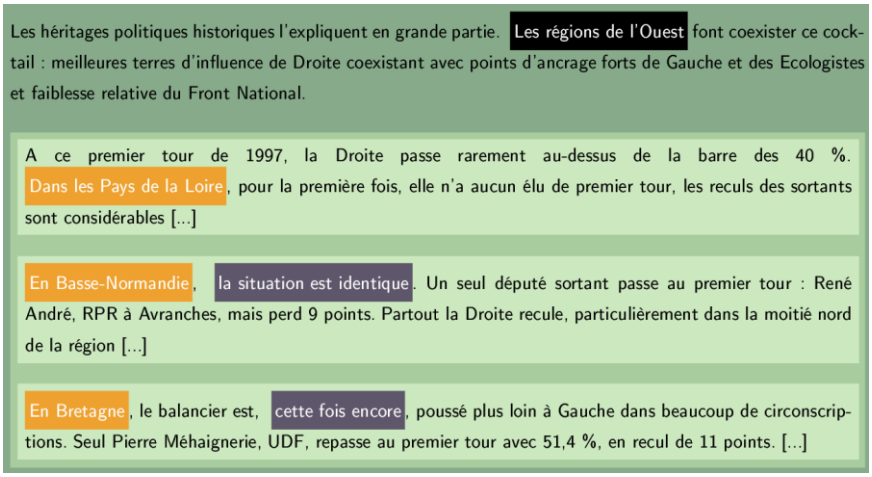


Figure 6. *Structure thématique énumérative*

4. La Pragmatique

4.1. La pragmatique comme étude de l'usage de la langue

Dans les sections précédentes, notre exposé a été organisé à partir des *productions langagières* elles-mêmes et, en quelque sorte de leur *structure interne*. La distinction de différents paliers : mot ou morphème au niveau le plus élémentaire, syntagme et phrase comme structures de base, et pour finir texte comme entité la plus vaste, est révélatrice de l'ancrage de notre présentation sur le « matériau » linguistique donné à comprendre ou interpréter.

Or on pourrait objecter à cette présentation qu'elle fait l'impasse sur les situations réelles de communication, et plus généralement sur *l'usage* des productions langagières dans l'interaction humaine. Elle s'arrêterait en quelque sorte aux représentations sémantiques associées, sans dire ce que l'on en fait et dans quelles activités (cognitives, communicationnelles…) elles s'inscrivent : là se situeraient les limites, les bornes, du domaine de *la sémantique*. C'est au contraire sur ces questions que *la pragmatique* va mettre l'accent. De manière générale, elle se définit comme la discipline qui étudie *tous les types d'actes auxquelles l'usage de la langue donne lieu*[38]. Dans la démarche très concrète qui est la nôtre, placée dès le premier chapitre sous le signe de *l'activité* de compréhension du langage, il est clair que l'étude de ces actes est pertinente : soit comme complément (ce qui est fait

[38] Le radical grec *pragma-* signifie précisément *action*.

des représentations produites) soit, plus fondamentalement, comme conditionnant (intervenant dans, interagissant avec) la « construction du sens » elle-même.

En réalité, ce partage un peu radical entre sémantique et pragmatique est discutable. Dans les sections précédentes, notre présentation s'articule sur une dualité représentations / opérations dans laquelle des éléments de contexte, aussi bien que fonctionnalité du texte, interviennent. De notre point de vue, une partie de l'objet d'étude revendiqué par la pragmatique doit donc être intégrée dans une sémantique bien comprise. Nous ferons cependant dans l'immédiat abstraction de ce débat pour simplement décrire et mettre en valeur ce qui fait la spécificité et la cohérence de la pragmatique, soulignant sa contribution importante, et en certains points décisive, à l'étude du langage.

Quel est donc cet « usage de la langue » dont la pragmatique dit faire son objet d'étude ? De quels « actes » s'agit-il ? Nous en présenterons rapidement deux aspects importants :

– Le premier s'ancre dans la situation de *communication*, avec la notion *d'acte de langage*, et plus généralement un ensemble de phénomènes liés à *l'énonciation* ;

– Le second s'intéresse à *l'interprétation* qui est faite par le récepteur d'un énoncé, et aux *inférences* réalisées, inévitablement, pour aller en quelque sorte « au delà du texte ».

4.2. Énonciation et actes de langage

L'idée générale est ici de constater que l'énonciation est une *action* réalisée par le locuteur, de l'étudier sous cet angle et d'en tirer toutes les conséquences. On pourra remarquer tout d'abord que le langage est véritablement le moyen par lequel nous réalisons un certain nombre *d'actes sociaux*, c'est-à-dire d'actes qui modifient les relations interpersonnelles au sein d'un groupe social. Par exemple, un officier d'état civil ou un prêtre prononçant la phrase :

(1) *Je vous déclare unis par le mariage*

modifie, de ce fait même, le statut social du couple auquel il s'adresse. De même :

(2) *La séance est ouverte*

prononcé par une personne qui en a le « pouvoir » modifie effectivement le statut de la réunion en cours. À une échelle interpersonnelle plus réduite,

(3) *Je te promets de venir demain*

installe un état d'obligation morale de son locuteur, un engagement à retrouver son interlocuteur.

En fait dans toute situation de communication le locuteur effectue en parlant un acte, dont il attend un effet sur son interlocuteur. Considérons les trois énoncés suivants :

(4) *Ferme la porte !*

(5) *La porte est-elle fermée ?*

(6) *La porte est fermée*.

(4) constitue un ordre, par lequel l'énonciateur attend que son interlocuteur agisse et ferme la porte. Dans (5), il attend une réponse constituant en une information sur l'état actuel de la porte. Mais (6) également est un acte, le locuteur escomptant que le récepteur accepte cette information et la range dans ses connaissances pour un usage ultérieur. On parlera *d'acte illocutoire* (ou *illocutionnaire*), consistant à poser une question, donner un ordre, informer etc. Et l'on appellera effet *perlocutoire* (ou *perlocutionnaire*) l'effet attendu d'un acte énonciatif, cet effet pouvant évidemment être réalisé ou non. C'est la théorie des *actes de langage*, initiée par Austin et Searle [AUS 70] [SEA 72] dans une problématique de philosophie du langage. Cette théorie a donné lieu à des formalisations logiques rigoureuses (voir par exemple [VAN 91]) et des applications au dialogue homme machine (par exemple dans les travaux sur le dialogue oral évoqués dans le chapitre 1 [SAD 94]).

Bien d'autres phénomènes sont liés à l'énonciation et mettent en évidence, au-delà du sens de la phrase énoncée, l'acte qu'elle constitue par elle-même. Considérons les énoncés suivants :

(7) *Il ne m'a pas demandé de sortir, il m'a mis dehors*

(8) *Pierre s'est marié, mais c'est un secret*

(9) *Es-tu libre ce soir ? parce qu'il y a un bon film au cinéma.*

Dans (7) l'effet de la négation porte non sur la proposition niée mais sur le fait même de **l'affirmer** : ce qui est faux, ce n'est pas « qu'il m'ait demandé de sortir », mais que ce soit une manière adéquate de « décrire son attitude ». Dans (8) la connexion porte non sur le **fait** que Pierre soit marié, mais sur le fait de le **dire**. Et dans (9) le connecteur *parce que* relie **la question** elle-même et non ce sur quoi elle porte : ***je te pose la question*** *parce qu'il y a un bon film au cinéma.* Ces exemples mettent en évidence le *caractère pragmatique* de certains usages de la négation et de certains connecteurs, non réductible à un contenu propositionnel au sens strict. Ce ne sont là que quelques exemples simples. On reconnaîtra sans peine que l'énonciation constitue un acte *proprio sensu* et peut être étudiée en tant que telle, la signification même de l'énoncé pouvant faire référence à cet acte.

4.3. *Interprétation (1) : Sens littéral et implicature*

Nous abordons maintenant une autre facette de l'analyse pragmatique. Bien souvent, dans notre compréhension d'un énoncé, nous allons au-delà de ce qui semble formellement, explicitement, dit ou écrit, comme dans les deux exemples (classiques) suivants :

(10) *Y a-t-il du sel sur la table ?*

(11) *Brrr. Il ne fait pas chaud ici.*

Tout le monde comprend immédiatement que, plus qu'une question, (10) est une demande du locuteur pour qu'un des convives lui transmette la salière. Et (11) n'est probablement pas qu'une simple constatation, mais une invite à fermer la porte ou la fenêtre. Mieux : il se peut dans les deux cas que la formule employée soit *principalement* une demande, la question ou l'assertion n'étant que de pure forme.

Suivant Moeschler [MOE 94][39] : « Une des tâches de la pragmatique est alors d'expliquer comment un auditeur peut arriver à comprendre une énonciation de manière non littérale et pourquoi le locuteur a choisi un mode d'expression non littéral plutôt qu'un mode d'expression littéral. » Et en effet le phénomène est beaucoup plus général et systématique que nos premiers exemples – formes assez conventionnelles assimilables à des formules de politesse – pourraient le faire croire. Voici quelques autres exemples. Considérons les énoncés :

(12) (a) *Jean a trois enfant.* (b) Jean a trois enfants. Il en a même cinq.

(13) (a) *Le drapeau est blanc*. (b) *Le drapeau est blanc et bleu*.

Si (12b) paraît étrange, bien que logiquement indiscutable, c'est que (12a) exprime en fait que Jean a *exactement* trois enfants, sans que cela soit dit explicitement. De même (13a) s'oppose à (13b) en ce qu'il spécifie implicitement que le drapeau est *uniquement* blanc. Il y a là un usage très général, inscrit dans la pratique de la langue, dans lequel on reconnaîtra quelque chose comme un *principe d'information maximum*. Si je parle des enfants de Jean ou de la couleur du drapeau, je dois donner toute l'information dont je dispose, donc respectivement le nombre et la couleur exacts. C'est ce qui permet à l'auditeur d'aller au-delà du sens littéral.

Soit maintenant la conversation suivante :

(14) A : *Je n'ai plus d'essence.*

[39] Envers lequel nous signalerons ici notre dette dans la rédaction de cette section, et vers lequel nous adressons le lecteur pour une introduction générale à la Pragmatique.

B : *Il y a un garage au coin de la rue*.

En premier lieu nous pouvons reconnaître dans l'assertion de A une question déguisée (*Peux-tu me dire où il y a de l'essence ?*) par le même mécanisme que dans nos premiers exemples. Mais la réponse de B est, elle aussi, indirecte. Il ne dit pas explicitement *où trouver de l'essence*, mais donne une information dont A doit pouvoir le déduire sans peine. Le mécanisme est complexe, mais nous pouvons reconnaître un principe très général de *pertinence* dans l'échange langagier : l'information apportée (ici : présence d'un garage) doit être pertinente par rapport au problème posé (trouver de l'essence), elle doit apporter un élément significatif pour sa solution.

Plusieurs « explications » de ces phénomènes langagiers ont pu être proposées. On citera ici Sperber et Wilson et leur *théorie de la pertinence* [SPE 89], présentée comme principe premier dont découle, par divers mécanismes, tout un ensemble d'effets dont les exemples ci-dessus constituent un aperçu. Et on s'attardera davantage sur un modèle particulièrement « populaire », la *logique de la conversation* de Grice [GRI 79]. Grice, et avec lui toute une tradition en pragmatique, part du postulat d'une distinction entre un sens *stricto sensu*, défini en termes référentiels ou vériconditionnels (le fait que la température est basse pour *il fait froid*) et un sens dérivé (*il faudrait fermer la fenêtre*) qui échappe aux conditions de vérité de la phrase. Cette partie de la signification de l'énoncé est nommée par Grice *implicature* : quelque chose qui est d'une certaine manière impliquée par le sens vériconditionnel, mais dans un sens non purement logique, faisant intervenir d'autres principes, relevant du bon usage de la communication langagière. Ces principes sont répertoriés par Grice sous forme de *maximes de conversation*, dont voici deux instances :

Maxime de quantité : Que votre contribution contienne autant d'information qu'il est requis et qu'elle ne contienne pas plus d'information qu'il n'est requis.

Maxime de relation (de pertinence) : Parlez à propos (soyez pertinent).

Il faut bien voir que ces principes et maximes ne sont pas des indications normatives et idéalistes : rien à voir avec un guide de savoir-vivre entre interlocuteurs bien intentionnés ! guide qui serait par ailleurs fréquemment violé, l'expérience le montre à loisir. Mais bien un ensemble de principes pour une théorie rationnelle de l'interprétation des énoncés dans l'échange langagier ; principes dont la violation autant que le respect sont à même de déclencher des implicatures. Ainsi, par la *maxime de quantité* (12a) a pour implicature : *Jean a **exactement** trois enfants*, et la maxime est violée par (12a), d'où son étrangeté. De même, (13a) a pour implicature que *Le drapeau est **entièrement** blanc*. C'est la maxime de pertinence qui est à l'œuvre dans (10), (11) et (14). Par exemple dans ce dernier cas, A est en droit de supposer que B a donné une information pertinente et donc a implicité : *Le garage a des pompes et est ouvert*.

4.4. *Interprétation (2) : « le lecteur modèle »*

Mais les mécanismes d'interprétation, permettant d'aller « au delà » de l'information *apparemment* portée par le texte, se manifestent dans d'autres circonstances. En premier lieu il se peut que l'on ne puisse pas vraiment attribuer un sens littéral précis à l'énoncé, parce que celui-ci est en effet ambigu. L'interprétation, empruntant en général des informations au contexte, va permettre de lever l'ambiguïté. Par exemple l'énoncé suivant :

(15) *Je roulais sur la partie droite de la chaussée quand un véhicule arrivant en face dans le virage a été complètement déporté. Serrant à droite au maximum, je n'ai pu éviter la voiture qui arrivait à grande vitesse.*

Le syntagme *la partie droite de la chaussée* est ambigu. Le rédacteur veut-il dire simplement qu'il roulait à droite ou qu'il était encore dans une partie rectiligne de la route, avant le virage. C'est

probablement la première interprétation qui est la bonne, celle que comprendra spontanément le lecteur[40]. Divers indices peuvent y conduire, dont la référence au code de la route et la volonté évidente du rédacteur de se disculper (nous allons y revenir). Quoi qu'il en soit, on devine ici un jeu complexe de connaissances et d'inférences.

Le lecteur ou auditeur est aussi souvent conduit à véritablement *ajouter de l'information* au texte. Un texte comme (15) possède une fonction argumentative : disculper le rédacteur et faire porter la responsabilité de l'accident sur l'adversaire. Un certain nombre d'indices conduisent le lecteur vers cette conclusion : le rédacteur roule, puis serre *à droite*, tandis que le véhicule adverse roule *à grande vitesse* au point d'être *complètement* déporté. Il n'est pas vraiment ici question de sens littéral et second (la description de l'accident est claire), mais d'une sorte de corrélat argumentatif. Alors est-il encore question de pragmatique, pour ne pas dire de sémantique ? Ou seulement de raisonnement à partir des informations exprimées par le texte ? Mais force est de constater que c'est bien le texte qui, par certains effets, certaines tournures, tend à induire le lecteur vers la conclusion de responsabilité de l'adversaire du rédacteur. Le texte « joue » en quelque sorte avec le lecteur pour l'amener à cette conclusion[41].

Cette interaction entre le texte et le lecteur, et entre sémantique et pragmatique, peut être beaucoup plus « intime » que dans les exemples précédents. C'est un point important de l'approche d'Umberto Eco et de sa notion de « lecteur modèle » développée notamment dans [ECO 85], dont nous allons donner maintenant un aperçu. Soit le fragment suivant [42] :

(16) *Jean entra dans la pièce. « Tu es revenu, alors », s'exclama Marie, radieuse.*

Entre autres choses, le lecteur doit comprendre : que Jean entre dans la pièce où se trouve Marie ; que celle-ci s'adresse à lui ; que Jean était parti ; que Marie ne s'attendait pas à son retour, ou n'osait l'espérer (le *alors* : il peut s'agir soit d'une information nouvelle, soit d'une confirmation); et qu'elle le désirait ardemment (combinaison *s'exclama* + *radieuse*). Etc.

Des mécanismes divers sont à l'œuvre. Mais il est clair que tout n'est pas dit explicitement : le lecteur doit combler de nombreux *vides* informationnels. C'est ainsi que fonctionne un texte, et c'est ainsi qu'il est produit par son auteur : comme « un tissu d'espaces blancs », dont « il prévoit qu'ils seront remplis par le lecteur ». À la fois pour des raisons d'économie (ne pas tout dire, si tant est que cela soit possible) et, dans le cas de la littérature, pour des raisons esthétiques : le lecteur doit avoir l'initiative de l'interprétation. Écrire un texte, c'est donc mettre en œuvre une stratégie qui prévoie les mouvements interprétatifs d'un *lecteur modèle*. Les prévisions peuvent s'avérer erronées : le texte ne fonctionne pas comme il devrait. L'auteur peut désirer guider étroitement le lecteur (texte « fermé »), ou aboutir au même résultat par lourdeur de style ; ou bien, dans un but esthétique, le laisser mener « sa libre aventure interprétative » (texte « ouvert »). Mais dans tous les cas, le lecteur doit être actif et coopératif pour comprendre le texte.

Eco identifie et décrit alors différents mécanismes à l'œuvre dans la coopération entre le lecteur et le texte. Ces mécanismes s'appuient sur des structures sémantiques, comme les *isotopies* identifiant un thème ; mais aussi des schémas narratifs évoqués dans la section précédente, nommés ici *fabula*, laquelle est *« le schéma fondamental de la narration, la logique des actions et la syntaxe des personnages, le cours des événements ordonné temporellement. »* Par exemple le lecteur d'*Oedipe roi* pourra construire une fabula très synthétique selon laquelle « *un coupable refuse de reconnaître l'histoire de sa culpabilité »*. Ou une version plus détaillée, qui suit la progression dramatique, avec

[40] Contrairement à : « Je roulais encore sur la partie droite de la chaussée, avant le virage ».

[41] Et on peut montrer que, dans certains constats du corpus, cette force argumentative a un effet en retour sur la sémantique même [EST 94].

[42] Issue de la nouvelle *Un drame bien parisien* d'Alphonse Allais.

des éléments d'interprétations du mythe. La fabula dépend donc elle aussi du lecteur qui la *construit*. A partir de là, le lecteur joue entre hypothèses, inférences, confirmations et remises en causes, dans ce qu'Eco appelle des *promenades inférentielles*.

4.5. *Conclusion. Un « au delà » de la sémantique ?*

Si nous reprenons la question par laquelle s'est ouvert cet ouvrage, « qu'est-ce que comprendre un énoncé, ou un texte ? » il est clair que la pragmatique met l'accent sur des aspects essentiels :

– mise en jeu de *connaissances générales partagées* (ou réputées telles),

– ainsi que de règles de *bon usage* dans la relation de communication (orale ou écrite),

– et prise en compte du *contexte d'énonciation* qui, tout en apportant des connaissances spécifiques, joue un rôle important pour mobiliser ces connaissances,

– dans des mécanismes de type *inférentiels*, conduisant à enrichir ou dépasser l'information « apparemment » portée par le sens texte, responsables du caractère essentiellement ouvert de *l'interprétation*.

Ces éléments sont clairement exogènes par rapport au texte. Pour autant sommes-nous « au-delà de la sémantique » ? Sémantique et pragmatique ont-elles des objets d'étude clairement distincts, sans recouvrement ? Ou au contraire s'agit-il de points de vue qu'il s'agit de croiser ? Les avis sont sur ce point partagés. Nous pourrons relever ici trois positions typiques.

Un premier point de vue, sans doute le plus répandu, s'ancre sur la distinction instituée par Morris entre la *syntaxe*, ou étude des relations et combinaisons entre unités de langue, la *sémantique*, intéressée par les relations entre les mots et expressions langagières et « le monde », et la *pragmatique* qui s'occupe des relations entre les signes linguistiques et les usagers du langage. Le *modèle linéaire* qui en découle spécifie une *suite* de traitements des énoncés séquentiellement, selon ces trois aspects de la langue. Nous ne discuterons pas ici la question des rapports entre syntaxe et sémantique. Concernant l'articulation sémantique/pragmatique, la « valeur ajoutée » de la pragmatique peut consister par exemple en l'attribution à l'énoncé d'une force illocutoire (acte de langage) ou dans le calcul d'un sens dérivé au-dessus du sens littéral, vériconditionnel comme il se doit dans la position de Morris.

Ce découpage strict et séquentiel pose de nombreux problèmes. Nous laissons au lecteur le soin de juger par lui-même — à partir notamment des exemples ci-dessus — s'il pense passer par « l'étape » du calcul d'un sens « littéral » et « factuel ». Le cas de l'ambiguïté (exemple 15) peut être ici particulièrement problématique, ou encore celui de l'interprétation des connecteurs « logiques » et du rôle de l'énonciation (exemples 7, 8, 9). On est alors amené à intégrer les deux « niveaux ». une manière de le faire est de distinguer une *signification* « purement linguistique » du *sens* de l'énoncé, qui intègre des « effets contextuels ». Il n'y a donc plus un sens « purement informatif et littéral », mais une signification portée par les éléments linguistiques seuls et en tant que telle incomplète, sur laquelle opère par un « composant rhétorique » intégrant des données de la situation d'énonciation pour produire le sens.

C'est par exemple la proposition la théorie de la *pragmatique intégrée* d'Anscombre et Ducrot [ANS 83] qui donnent l'exemple suivant. Si l'énonciateur de *J'ai un peu d'argent sur moi* sort une liasse de billets pour payer le barman, l'auditeur ne fait pas le détour par un sens littéral contredit par la situation et la maxime de qualité (« dire le vrai »), mais calcule directement le sens « j'ai suffisamment d'argent » grâce à une *loi de discours* de la litote qui permet d'interpréter dans certaines situations une expression de faible quantité (« un peu ») par une grande quantité. C'est

donc le sens de *un peu* qui est resté non totalement spécifié par le composant linguistique, en attente de spécification plus précise de par la situation. Voir aussi l'exemple de *la voie de droite* (cf. § 2.3) dans lequel à la sémantique courante de la description définie, en terme de « spécification univoque » est substitué un critère de « maximalité » par un procédé relativement normé en langue.

Le troisième courant pourrait être qualifié de « sémantique libéral » dans la mesure où ses tenants, refusant la coupure, « introduisent en quelque sorte des éléments pragmatiques dans le domaine sémantique » ([ECO 85] p. 303) et insistent, de multiples façons, sur l'imbrication des informations contextuelles (au sens le plus large) et linguistiques dans la compréhension des énoncés. Le modèle de la *pragmatique intégrée* peut en être vu comme une version « timide » car encore séquentielle. Le §.4.4 a donné quelques exemples de cette imbrication. Un autre exemple concerne la sémantique lexicale elle-même qui doit, selon cette vision, intégrer des éléments encyclopédiques (Eco insiste particulièrement sur ce point).

Notre position est du troisième type, les chapitres prochains en donneront diverses illustrations. Notre point de vue peut être schématisé à partir de l'idée d'une boucle d'interprétation, dans laquelle le lecteur/interprète intègre les « valeurs » sémantiques portées par les constituants du texte, à « l'univers de ses connaissances » en fonction d'attentes et d'anticipations. L'analyste tentera de démêler (comme nous venons de le faire dans ce chapitre) les différentes « instances », sémantique et pragmatique, responsables de « l'état de compréhension » auquel il parvient. Mais cette analyse nécessaire ne devra jamais être assimilée « en droit » à une séquentialité de traitements. Au contraire l'attention doit être portée sur les différentes formes de *coopération* entre ces instances. La séquentialisation pourra être rendue nécessaire, dans une certaine mesure, par les traitements informatiques, pour des raisons de modularité : nous le verrons au chapitre 7 ; mais le développeur devra rester conscient du caractère artificiel du découpage.

5. Bibliograhie

[ANS 83] ANSCOMBRE J.-C., DUCROT O., *L'argumentation en langue*, Mordaga, 1983.

[AUS 70] AUSTIN, J.L., *Quand dire, c'est faire*, Seuil, 1970.

[CAR 89] CARON, J., *Précis de psycholinguistique*, Presses Universitaires de France, Coll. Le psychologue, 1989.

[CHA 97] CHAROLLES M., « L'encadrement du dicours - Univers, champs, domaines et espace », *Cahier de recherche linguistique*, 6, p. 1-60, 1997.

[COM 76] COMRIE B., *Aspect*, Cambridge University Press, 1976.

[CUL 90] CULIOLI A., *Pour une linguistique de l'énonciation, Opérations et représentations*, tome 1, Ophrys, 1990.

[DAL 03] DAL, G., « La productivité en questions et en expérimentations », *Langue française*, 140, Décembre 2003.

[DES 85] DESCLES, J.-P., « Représentation des connaissances : archétypes cognitifs, schèmes conceptuels, schèmes grammaticaux », *Actes Sémiotiques*, Documents(VII), p. 69-70, 1985.

[DUB 97] DUBOIS, D. (DIR.), *Catégorisation et cognition*, Editions Kimé, 1997.

[DUC 72] DUCROT O., TODOROV, T., *Dictionnaire Encyclopédique des Sciences du Langage*, Seuil, 1972.

[DUT 03] DUTOIT D, NUGUES P., DE TORCY P., « The Integral Dictionary : a lexical network based on computational semantics », *ICCSA, International Conference on Computational Science and its Applications*, Calgary, Canada, Springer, Mai 2003,

[ECO 85] ECO U., *Lector in fabula*, Traduction Française, Le livre de poche, n° 4098, Grasset, 1985.

[ENJ 94] ENJALBERT P., VICTORRI B., « Du langage au modèle », in LEVY, F. (dir.), *Approches sémantiques, Revue T.A.L.*, Vol 35, n°1, p. 37-64, 1994.

[EST 94] EXTIVAL, D., GAYRAL, D. « Contexte et inférence », in LEVY, F. (dir.), *Approches sémantiques, Revue T.A.L.*, Vol 35, n°1,

[FRA 03] FRADIN B., KERLEROUX F., « Quoi de neuf en morphologie ? », *Langages*, 152 Décembre 2003.

[GIG 97] GIGUET E., VERGNE J., « From part of speech tagging to memory-based deep syntactic analysis », *Proceedings of International Workshop of Parsing Technologies*, Massachussetts, MIT, 1997.

[GRI 79] GRICE, H.P., « Logique et Conversation », *Communications* 30, p. 57-72, 1979.

[KAY 97] KAYSER D., *La représentation des connaissances*, Hermès, 1997.

[LAN 86] LANGACKER R. W., An introduction to Cognitive Grammar, *Cognitive Science*, 10, p. 1-40, 1986.

[MIL 90] MILLER G. A. , BECKWITH R., FELLBAUM C., GROSS D., MILLER K.J., « Introduction to WordNet: An on-line lexical database », International Journal of Lexicography, vol 3, n°4, p. 235-312, 1990. Voir aussi http ://www.cogsci.princeton.edu/~wn.

[MOE 94] MOESCHLER, J., REBOUL, A., *Dictionnaire encyclopédique de pragmatique*, Seuil, 1994.

[NAZ 98] NAZARENKO A. (ED.), *Compositionnalité, Revue T.A.L.*, Vol 39, n°1, 1998.

[NYC 98] NYCKEES, V, *La sémantique*, Belin Coll. Sujets, 1998.

[PER 04] PERLERIN, V. Sémantique légère pour le document – Assistance personnalisée pour l'accès au document et l'exploration de son contenu. *Doctorat en informatique de l'université de Caen / Basse-Normandie*, Décembre 2004.

[PRO 28] PROPP, V. J., *Mophologie du conte*, Seuil, 1928.

[PUS 95] PUSTEJOVSKI, J., *The generative lexicon*, MIT Press, 1995.

[QUI 68] QUILLIAN R., « Semantic memory », *Semantic information processing*, Minsky M. (Ed.), p. 227-270. M.I.T. Press, 1968.

[RAS 87] RASTIER F., *Sémantique interprétative*, Presses Universitaires de France, 1987.

[RAS 91] RASTIER F., *Sémantique et recherches cognitives*, PUF, 1991.

[RAS 94] RASTIER F., *Sémantique pour l'analyse*, Masson, 1994.

[SAB 00] SABAH, G., GRAU, B., « Compréhension automatique de textes », in PIERREL J.M. (DIR.), *Ingénierie des langues*, Hermès, 2000.

[SAB 89] SABAH, G., *L'intelligence artificielle et le langage* (2 volumes), Hermès, 1988-1989.

[SAD 94] SADEK D., « Towards a theory of belief reconstruction: application to communication », *Speech Communication Journal* 15, n°3-4 (numéro spécial tiré de Int. Symp. on Spoken Dialogue, Tokyo, Japon, Novembre 1993), p. 251-263, 1994.

[SCH 75] SCHANK R.C., *Conceptual Information Processing*, North-Holland, 1975.

[SCH 94] SCHMIDT H., « Probabilistic Part-of-Speech Tagging Using Decision Trees », *Proc. Int. Conference on New Methods in Language Processing*. Manchester, UK, 1994.

[SEA 72] SEARLE, J.R., *Les actes de langage*, Hermann, 1972.

[SOW 84] SOWA J. F., *Conceptual Structures: Information Processing in Mind and Machine*, Addison-Wesley, 1984.

[SPE 89] SPERBER, D., WILSON, D., *La pertinence. Communication et cognition*, Ed. de Minuit, 1989.

[TAL 00] TALMY Leonard, *Towards a Cognitive Semantics*, vol. 1, Cambridge, MIT Press, 2000.

[TAN 97] TANGUY, L., Traitement automatique de la langue naturelle et interprétation : contribution à l'élaboration d'un modèle théorique informatique de la sémantique interprétative, Thèse de l'université de Rennes 1, 1997.

[VAL 04] VALETTE, M., « Sémantique interprétative appliquée à la détection automatique de documents racistes et xénophobes sur internet », *Actes Conf. Int. sur le Document Numérique, CIDE 7*, p. 215-230, 2004.

[VAN 91] VANDERVEKEN, D., *Meaning and Speech Acts* , Volume 1: *Principles of Language Use*, Volume 2: *Formal Semantics of Success and Satisfaction* , Cambridge University Press, 1990-1991.

[VIC 96] VICTORRI, B., FUCHS, C. *La polysémie : construction dynamique du sens*, Hermès, 1996 .

[VIC 04] VICTORRI, B., *Les grammaires cognitives*, in C. FUCHS (ed.) *La linguistique cognitive*, Ed. de la Maison des Sciences de l'Homme & Ophrys, 2004.

[VIR 01] VIRBEL, J. LUC, Ch. « Le modèle d'architecture textuelle : fondements et expérimentation », *Verbum* , XXIII, 1, 103-123, 2001.

[WID 04] WIDLÖCHER, A., FAUROT, E., BILHAUT, F., « Multimodal Indexation of Contrastive Structures in Geographical Documents ». In *Proceedings of RIAO 2004*, p. 555-570, Avignon, France, 2004.