

## Modes de construction du sens en corpus spécialisé

Anne Condamines\*

*Cet article se propose de montrer que l'analyse de corpus spécialisé relève d'une interprétation qui doit tenir compte de deux éléments : la situation de production des textes étudiés et l'objectif visé par l'interprétation. Loin de ne relever que d'une perspective applicative, ce type d'interprétation peut au contraire mettre en lumière des phénomènes linguistiques qui échappent à une approche seulement introspective. Le cas du repérage de relations sémantiques entre termes à l'aide de marqueurs est présenté ici. Dans ce cadre, les résultats de l'étude quantitative et qualitative du fonctionnement de avec ouvre des perspectives de description inédites.*

*This article aims to show that specialised corpus linguistics may be interpreted in terms of two factors: the context of production of the texts analysed and the aim of the analysis. Far from being part of a mere applied perspective, this kind of analysis can shed light on language phenomena which are impervious to a solely introspective approach. We present here the technique of pinpointing semantic relations between terms using cues. Within this context, the results of both a quantitative and a qualitative analysis of the preposition avec open up certain original perspectives for description.*

---

\* ERSS (UMR 5610), CNRS et Université de Toulouse-Le Mirail.

## 1. Introduction

La mise à disposition de textes au format électronique et d'outils pour les interroger constitue un tournant épistémologique majeur pour l'analyse sémantique. En effet, les données attestées peuvent être utilisées en lieu et place de l'introspection. Mais loin de ne constituer qu'une substitution, cette prise en compte suppose de s'interroger sur le rôle du contexte dans l'analyse sémantique, entendu à la fois comme contexte de production et contexte de réception. En effet, les textes qui nous parviennent sont forcément des textes situés alors que les exemples produits par introspection sont censés être totalement décontextualisés. Dans une perspective de systématisation linguistique, l'analyse sémantique de corpus doit à la fois tenir compte du contexte dans lequel les textes sont produits et évaluer les possibilités de généralisation, voire de neutralisation du contexte, l'objectif étant alors de proposer des descriptions valables quel que soit le contexte d'apparition. Compte tenu des besoins sociétaux (des entreprises en particulier), les corpus que nous sommes amenés à travailler relèvent souvent d'un contexte spécialisé. Il serait trop hâtif de dire que les descriptions que nous proposons relèvent par conséquent seulement d'une linguistique appliquée. Nous verrons que la nature spécialisée des corpus permet de s'interroger sur la connaissance mise en oeuvre pour les analyser, plus fondamentalement sur notre compétence de linguiste. La question de la construction de réseaux terminologiques, qui exemplifie le propos de cet article permet d'évaluer les possibilités de généralisation des descriptions des marqueurs de relations conceptuelles et propose de nouvelles perspectives pour décrire certaines prépositions (comme *avec*) en prenant en compte le genre textuel.

## 2. Corpus spécialisés et construction du sens

### 2.1. Construction du sens

76

Le titre de ce paragraphe nécessite tout d'abord que soit explicité ce que l'on peut entendre par « construction du sens ». En tout premier lieu, si le sens est construit, c'est parce qu'il n'est pas directement donné par la lecture d'un texte. En d'autres termes, il ne s'agit pas seulement de retrouver un sens préexistant à la mise en mots mais bien d'opérer une construction, c'est-à-dire une interprétation. De ce point de vue, la perspective que je propose s'inscrit dans une tradition herméneutique (Rastier, 1987) mais je m'intéresse plus particulièrement aux interprétations qui peuvent apparaître comme consensuelles, communes à des groupes de locuteurs qu'il faut parvenir à caractériser. Cette position peut (encore) paraître étonnante, en particulier parce qu'elle remet en question l'idée de la langue comme moyen de communication, c'est-à-dire comme moyen d'encodage et de médiatisation d'un contenu, censé parvenir à peu près intact à l'interlocuteur.

Les approches centrées sur la mise au jour introspective d'un contenu dont on pense pouvoir contrôler la stabilité quels que soient les contextes ont fait

florès ; le générativisme et les approches cognitives se sont inscrits dans cette perspective et ont produit des résultats intéressants. Mais, d'une part la mise à disposition de textes informatisés, en particulier via internet, et d'autre part, la demande sociétale de plus en plus pressante font qu'on ne peut plus faire l'impasse sur la prise en compte de faits langagiers attestés. Cependant, prendre en compte la réalité des usages, c'est aussi prendre en compte la réalité des contextes dans lesquels ils apparaissent et, par conséquent de la variation dans les modes de production et de réception qui peuvent affecter grandement les modes d'interprétation. De ce fait, des approches qui revendiquent la nécessaire prise en compte du contexte pour construire le sens (analyse de discours, sociolinguistique, ethnolinguistique...) se trouvent confortées par la prise en compte de données textuelles attestées. Désormais, la consultation de ces données permet tout d'abord d'élaborer des hypothèses (par une consultation rapide d'un ensemble d'attestations « grapillées » dans différents textes) et, dans un second temps de les mettre à l'épreuve des faits grâce à l'étude d'un corpus spécifiquement constitué. En effet, comme l'a montré la linguistique de corpus, (Kennedy, 1998), un corpus n'est pas un ensemble de textes aléatoirement rassemblés mais un ensemble de textes rassemblés pour un objectif particulier. On voit ainsi que l'introspection ne disparaît pas complètement de cette approche puisque la construction du corpus est précédée de l'élaboration d'une hypothèse, élaboration qui se fait dans un premier temps en faisant intervenir l'intuition.

Considérons donc que l'analyse de corpus relève d'une interprétation. Se pose alors la nature de l'interprétation mise en œuvre ; dans une perspective linguistique en particulier, est-il possible de repérer des régularités interprétatives (qui seraient partagées par tout linguiste qui se trouverait dans la même situation d'interprétation) ? En d'autres termes, est-il possible de décontextualiser les interprétations, c'est-à-dire de considérer les fonctionnements comme propres à la langue (en tout cas partagés par tous les locuteurs) ou bien de proposer des descriptions qui soient liées à des types de situations ou bien encore de rendre toujours les descriptions dépendantes des situations (approche ethnolinguistique par exemple) ?

Ma proposition est qu'il est possible de repérer des régularités linguistiques et, parfois, de les désolidariser de leur situation d'apparition, c'est-à-dire de considérer qu'il s'agit de régularités de la langue. Mais il est très difficile de prédire *a priori* les fonctionnements qui pourront être considérés comme propres à une langue et ceux qui vont être dépendants d'une situation particulière. La difficulté pour ce second cas sera de parvenir à caractériser les textes auxquels s'adressent les descriptions proposées. La notion de genre textuel est alors particulièrement pertinente. Mise en œuvre à la fois par l'analyse de discours et l'analyse littéraire mais aussi par la linguistique de corpus, la notion de « genre textuel » pourrait constituer un palier de description pertinent des phénomènes langagiers. En effet, le genre textuel

permet de rendre compte d'affinités entre caractéristiques extra-linguistiques et fonctionnements langagiers ; elle peut constituer un mode de catégorisation des textes qui peut être utile pour préciser les descriptions (Bouquet, 2004).

## 2.2. Corpus spécialisé

Le cas des corpus spécialisés est particulièrement intéressant pour un linguiste. En effet, la dimension interprétative, moins facile à repérer pour des corpus non-spécialisés, se pose d'emblée puisque nous savons que nous n'avons pas de compétence linguistique (de compétence de locuteur) pour ces corpus, en tout cas pas toute la compétence linguistique. En revanche, nous pensons que notre compétence de linguiste (de professionnel de la langue) peut être utile et pertinente pour produire des interprétations. Mieux, cette situation particulière nous aide à nous interroger à la fois sur nos modes d'interprétation et sur notre compétence de linguistes.

La question de la nature spécialisée des corpus a fait l'objet de nombreux débats (Lerat, 1995). La notion de spécialité suppose l'existence d'une langue générale par rapport à laquelle les langues spécialisées se définiraient. En d'autres termes, il y aurait une langue commune aux locuteurs et des langues variables en fonction des compétences plus ou moins spécialisées des uns et des autres. Ce point de vue constitue une façon d'essayer de créer un lien entre la langue-système et les usages réels. Il suppose d'identifier des domaines auxquels on pourrait associer des usages stables. Or, la variation dans les usages peut être très présente et ne pas correspondre à des domaines d'activités identifiés *a priori*. Ainsi, il peut y avoir plus de similitudes dans les usages d'un groupe d'ingénieurs de différentes entreprises qui travaillent sur un même projet que dans les usages des ingénieurs d'une même entreprise. Cet état de fait peut compliquer considérablement l'objectif de stabilisation des descriptions.

Si l'on est d'accord pour considérer que l'objet de la linguistique, ce sont d'abord les usages dont on veut décrire les éventuelles régularités, le problème se pose de manière différente. En effet, ainsi que nous l'avons vu, les usages nous arrivent accompagnés d'éléments de connaissance sur leur contexte de production. Parmi ces éléments, certains peuvent concerner le fait que ces textes ont été produits dans un contexte spécialisé. Bien souvent, il faut entendre par là un contexte qui a une visée praxéologique : créer, définir, construire, que ce soit un satellite, une œuvre d'art, un plat ou... une théorie. En d'autres termes, les corpus spécialisés sont associés à des actions, physiques ou intellectuelles et c'est cette perspective commune à plusieurs locuteurs qui entraîne la spécificité d'un corpus (Fillietaz et Bronckart, 2005). D'une certaine façon, les régularités qui se créent dans les usages langagiers viennent du fait que les locuteurs ont une visée commune ; dans la plupart des cas, ces locuteurs ne sont pas conscients des régularités qu'ils mettent en place.

La mise au jour de ces régularités constitue la première façon d'encadrer et de comprendre les modes d'interprétation. Une seconde façon de baliser l'interprétation, tout aussi importante et tout aussi présente dans les corpus spécialisés, concerne la prise en compte de l'objectif de l'étude.

On ne décide pas d'étudier un corpus au hasard ; comme nous l'avons vu, l'élaboration d'un corpus est précédée de la construction d'une hypothèse. Dans les cas dont nous allons parler en 3, l'hypothèse qui permet la construction du corpus est travaillée à partir d'une demande extérieure, non-linguistique. En effet, la demande en lien avec la nécessité de gérer, interroger, maintenir, traduire la documentation, tout particulièrement dans les entreprises est de plus en plus importante. C'est à cette demande que nous sommes souvent confrontés. Il ne s'agit pas pour nous simplement de mettre en œuvre une connaissance préexistante mais de comprendre comment cet objectif, en lien avec une demande particulière, vient prendre place dans l'interprétation que nous faisons. Ainsi que nous le verrons dans les exemples présentés dans le paragraphe 3, loin de ne relever que d'une linguistique appliquée, les études que nous menons ouvrent au contraire des perspectives inédites pour l'étude du fonctionnement de la langue.

### **2.3. Modes d'analyse**

Un constat s'impose : l'analyse sémantique de corpus ne peut se résumer au simple appariement d'une connaissance linguistique, supposée détenue par le linguiste, avec des faits langagiers. L'élaboration d'une interprétation s'ancre profondément dans les usages attestés dans les corpus ; elle fait aussi intervenir la compétence linguistique de deux façons : via sa compétence de locuteur (tout locuteur d'une langue « comprend » au moins en partie un texte dans cette langue) et via sa compétence de linguiste (métaconnaissance sur le fonctionnement de la langue). Grâce à cette double compétence, le linguiste élabore progressivement une interprétation en tenant compte de l'objectif de l'analyse.

Concrètement, on peut dire que, dans le cadre d'une analyse de corpus, l'approche mise en œuvre est toujours de type distributionnel. D'une part parce que, comme dans l'approche harissienne, le corpus (ou une partie) sert de cadre de référence et de limite de validité à l'analyse. D'autre part parce que, qu'elle soit ou non assistée par des outils, l'analyse d'un corpus passe par l'examen des contextes d'apparition de formes langagières, la nature des formes étudiées pouvant varier en fonction de l'étude (morphèmes, lexique, énoncés...). Ainsi, l'idée que des régularités existent dans un corpus et qu'elles peuvent être mises au jour par l'étude des contextes sous-tend nécessairement le travail d'analyse sémantique des corpus. Si nos analyses divergent avec la méthode strictement harissienne, c'est dû au fait que, dans l'approche harissienne, le sens émerge de la seule combinaison de formes : le sens, c'est la distribution. Or, il est clair que sur le constat de régularités de

distribution, une interprétation doit s'élaborer et c'est cette interprétation qui donne le sens. Ce point semble maintenant acquis par les informaticiens qui mettent en œuvre une approche distributionnelle (Habert et Zweigenbaum, 2002).

Grosso modo, deux types de phénomènes langagiers apparaissent lors de l'étude :

– Ceux qui semblent conformes à l'intuition du linguiste, c'est-à-dire à la fois proches de phénomènes qu'il a déjà rencontrés en tant que locuteur et proches de descriptions qu'il a faites ou qui ont été faites par des collègues. Ces phénomènes sont précieux car l'interprétation va pouvoir prendre appui sur eux.

– Ceux qui ne sont pas conformes à l'intuition du linguiste, soit qu'ils échappent à sa compétence de locuteur soit qu'ils ne soient pas conformes aux descriptions linguistiques qui ont pu être faites précédemment.

Une des difficultés consiste à repérer les éléments qui fonctionnent sur le mode attendu et ceux dont le fonctionnement est spécifique au corpus. L'analyse est terminée lorsqu'une stabilité s'est instaurée, c'est-à-dire quand tous les éléments pertinents ont pris leur place dans l'interprétation (en fonction d'un objectif, rappelons-le).

### **3. Exemple d'interprétation sémantique : la construction d'ontologies à partir de textes**

#### **3.1. Construction d'ontologies**

La notion d'ontologies est apparue de manière assez récente en ingénierie des connaissances (une quinzaine d'années environ) et a fait depuis l'objet de nombreux travaux (Bachimont, 2000). Une ontologie est une représentation formelle de la connaissance sous la forme de nœuds reliés par des arcs. Il s'agit d'un mode de représentation particulièrement adapté à l'informatique ; il est d'ailleurs utilisé par plusieurs modèles informatiques (graphes conceptuels, logiques terminologiques). Les nœuds et les arcs étant étiquetés par des noms (plus rarement des verbes), le lien avec la langue est très présent et la nature sémantique des relations entre les termes est travaillée par différentes disciplines (Green *et al.*, 2002). Avec le développement des applications informatiques qui utilisent ce type de ressources (recherche d'information, système d'aide au raisonnement...), ce lien s'est encore renforcé et désormais, un courant de recherche interdisciplinaire étudie les modes de construction possible de ces représentations à partir de textes. La plupart du temps, ces textes proviennent d'entreprises, qui ont des besoins très importants en matière de gestion de leur documentation. Ce nouveau courant de recherche a accéléré le rapprochement avec la sémantique de corpus, particulièrement avec la sémantique de corpus spécialisés. La demande est claire : il s'agit d'évaluer les modes d'élaboration possible d'une représentation de la connaissance d'un texte ou d'un corpus sous la forme

d'un réseau de relations. En d'autres termes, il s'agit d'étudier comment passer d'un discours dont la caractéristique est d'être linéaire, continu et syntaxique à une représentation qui est discontinue, spatiale et iconique (les flèches reliant les nœuds venant se substituer à des structures syntaxiques).

Clairement, ce type de représentation relève d'une interprétation. Il ne s'agit pas de retrouver un réseau qui préexisterait et qui serait simplement mis en mots lors de la production du discours mais bien de construire une représentation qui prenne appui sur des passages du texte et qui tienne compte de l'objectif, c'est-à-dire du besoin identifié (construction d'un outil d'aide au raisonnement, aide à la traduction...). Ce besoin intervient d'ailleurs dès le moment de la constitution du corpus.

Il faut noter que ce mode de représentation n'est pas toujours adéquat pour représenter le sens. Même si on peut considérer que le mode de représentation peut toujours s'adapter au sens à représenter, la représentation sous forme relationnelle introduit une discontinuité et un figement dans un discours qui se caractérise par une continuité dynamique. Ainsi, ce mode de représentation n'est pas toujours une façon de rendre explicite une connaissance qui ne l'était pas sous sa forme discursive. La suppression des éléments syntaxiques peut, au contraire, créer de l'ambiguïté, en tout cas si le cadre d'interprétation n'est pas clairement explicité (Cusin-Berche, 1997). Par ailleurs, le discours fonctionne parfaitement avec des flous alors que la représentation sous forme de réseau oblige à préciser les contours, c'est-à-dire à donner une interprétation plus restrictive. C'est le cas avec les nominalisations déverbiales par exemple qui, très souvent, peuvent être, en contexte interprétées à la fois comme un processus et comme le résultat de ce processus comme dans :

*Le chef de projet est responsable de la documentation*

Dans cette phrase, *documentation* renvoie à la fois au processus (constitution de documents qui accompagne le projet) et au résultat (ensemble des documents produits). Dans d'autres parties du texte, elle renvoie soit à un sens soit à l'autre.

Le choix de modélisation oblige à « lever cette ambiguïté » et à ne sélectionner que l'un ou l'autre sens.

Malgré ces difficultés, que l'on doit avoir à l'esprit en permanence, il reste que la représentation sous forme relationnelle permet la mise en œuvre par des outils, ce qui présente un double avantage. En effet, ce mode de représentation est utile non seulement pour permettre la construction d'outils d'aide au raisonnement mais aussi pour aider à repérer des incohérences ou des incomplétudes lors même de la construction des réseaux terminologiques (par exemple, une tête qui n'aurait qu'un fils dans une hiérarchie). Ces

incohérences ou incomplétudes peuvent provenir d'un oubli (il faut alors par exemple rechercher un autre fils) ou bien d'un corpus insuffisamment représentatif ou bien encore d'un point de vue qui a sa légitimité et sur lequel il faut interroger un expert du domaine. Ainsi, même si ce mode de représentation n'est pas toujours adapté pour représenter le sens d'un texte (encore qu'il soit mieux adapté pour les textes spécialisés dans lesquels le sens se construit de manière plus consensuelle), il présente un certain nombre d'avantages qui justifient qu'on le mette en œuvre. Enfin, c'est un mode de représentation qui est très répandu (sciences de l'information, terminologie, didactique...).

### 3.2. Les marqueurs de relation

La construction de réseaux de relations à partir de textes s'appuie majoritairement, de façon plus ou moins consciente, sur le repérage d'éléments lexicaux ou syntaxiques (voire même typographiques ou dispositionnels) que l'on sait pouvoir utiliser pour repérer des portions qui peuvent être représentées sous forme relationnelle. Par exemple dans :

*Tous les invertébrés sauf les insectes se rangent dans les vers* (manuel de biologie)

On trouve le marqueur [tous les N1 sauf les N2] qui permet de repérer une relation d'hyponymie entre *invertébrés* et *insectes*.

Ces éléments sont appelés de différentes façons selon les auteurs : *diagnostic frames*, chez Cruse (1986), *knowledge rich contexts* chez Meyer (Meyer, 2000)... Nous les appelons *marqueurs de relations conceptuelles*.

L'analyse de corpus pour décrire les marqueurs de relation met en évidence différents éléments.

– Tout d'abord, les marqueurs ne sont pas des indices qui pointent directement sur une interprétation. Ils sont des éléments qui constituent un déclencheur d'une interprétation possible mais ils ne donnent pas l'interprétation. Si l'on prend le cas de *chez*, par exemple, on se rend compte que cette préposition peut déclencher une interprétation méronymique dans des corpus didactiques de sciences naturelles. Ainsi, dans :

*Chez les colobinés, le nez fait saillie sur la lèvre supérieure*

Il y a bien une relation méronymique entre *nez* et *colobinés* et cette interprétation est associée à la présence de *chez* (Condamines, 2000). Mais, il serait tout à fait inexact de penser que c'est *chez* qui donne l'interprétation méronymique. Cette interprétation fait en effet appel à une connaissance pré-existante qui est convoquée par une structure particulière qui met un être vivant (qui est introduit par *chez*) en position thématique, le reste de la phrase

présentant une connaissance nouvelle sur un élément de cet être vivant : son anatomie, son habitat, son mode de reproduction. Il se trouve que le plus souvent, c'est l'anatomie qui est évoquée, d'où cette impression que *chez* « marque » la méronymie.

– Ensuite, les marqueurs sont moins spécialisés que les termes qu'ils relient. Pour le dire autrement, le métalangage des marqueurs de relations relève d'une compétence linguistique moins experte que celle requise pour interpréter les termes. C'est d'ailleurs pour cette raison que des terminologues ou ingénieurs de la connaissance peuvent construire des réseaux terminologiques même dans des domaines pour lesquels ils ne sont pas experts.

– Contrairement à une vision que l'on retrouve dans bon nombre de travaux sur les marqueurs, on se rend compte que les marqueurs permettent de repérer des relations n-aires et pas seulement des relations binaires. Ainsi dans :

*Chaque subdivision transmet une fiche aux chantiers qui se terminent*

On peut repérer des relations interdépendantes concernant le scénario de la communication : X envoie Y à Z (X = *subdivision*, Y = *fiche*, Z = *chantiers*). On peut rendre compte de ce schéma sous la forme de deux relations séparées : X envoie Y et X envoie à Z. Mais on ne rend pas compte de la totalité de la connaissance.

– Enfin, les marqueurs ne constituent pas un système unique et stable. Il existe des variations que des études à partir de corpus peuvent mettre en évidence et qui font intervenir la notion de genre textuel (Condamines, 2002). Les types de variations possibles sont présentées dans le prochain paragraphe.

### **3.2.1. Variation et marqueurs de relations conceptuelles**

Tous les marqueurs de relations conceptuelles ne fonctionnent pas de manière stable quel que soit le corpus. Trois types de fonctionnement peuvent être décrits.

– Certains marqueurs sont stables quel que soit le corpus. Il faut entendre par là que, quelle que soit la nature du corpus, ils peuvent être utilisés et produiront des résultats pertinents, c'est-à-dire que les passages repérés de cette façon pourront être représentés sous la forme d'une relation. Ces marqueurs semblent assez peu nombreux et concernent majoritairement les relations d'hyponymie (Rebeyrolle et Tanguy, 2000) et de méronymie.

– Certains marqueurs semblent étroitement liés au corpus dans lequel ils apparaissent. Par exemple, dans un manuel de spécification en génie logiciel, nous avons pu trouver le marqueur : [(phase, étape), déverbal] + (lorsque, dès que) + V au passif] qui permet d'identifier une relation de condition comme dans :

Anne Condamines

*La phase d'intégration du composant peut commencer lorsque l'ensemble des éléments logiciels ont été codés*

qui doit être comprise comme : la phase *d'intégration du composant* ne peut commencer que lorsque *l'ensemble des éléments logiciels ont été codés*, le codage de l'ensemble des éléments logiciels étant la condition du début de la phase d'intégration.

Ces marqueurs sont imprédictibles et des stratégies d'identification particulières doivent être mises en œuvre pour les repérer (Condamines et Rebeyrolle, 2001).

– Enfin, certains marqueurs sont liés au genre du texte dans lequel ils sont utilisés (voir 3.2.2).

Cette dépendance avec le genre textuel se manifeste d'un point de vue quantitatif et/ou qualitatif. Cela signifie que ces marqueurs ne sont pas équiprobables dans tous les corpus et/ou qu'ils ne fonctionnent pas de la même façon selon les corpus.

### 3.2.2 Avec comme marqueur de méronymie

Prenons l'exemple de la préposition *avec* qui peut, dans certains cas, être considérée comme un marqueur de méronymie (*une robe avec des dentelles*). Dans le cas de cette préposition, la différence en fonction du genre se manifeste à la fois du point de vue quantitatif et du point de vue qualitatif. Le fonctionnement de *avec* en lien avec la méronymie a été étudié dans cinq textes, relevant de cinq genres textuels différents : un roman (*Germinal*), un manuel de géomorphologie, un catalogue de jouets, un ensemble de petites annonces immobilières trouvées sur le web, un ensemble de descriptifs d'itinéraires (issues d'une enquête psycholinguistique).

L'analyse des résultats a mis en évidence les fonctionnements suivants.

84

#### Point de vue quantitatif

La répartition des *avec* méronymiques dans les différents corpus parle d'elle-même.

	GERMINA.	GEOMOR.	CATALO.	P. ANNO.	ITINE.
Nbre de mots	209 200	219 100	9 200	12 630	48 000
Avec	667 (0,3%)	432 (0,2%)	236 (2,6%)	185 (1,5 %)	116 (0,2%)
Avec « mérony. »	43 (6,4 %)	55 (12,7 %)	161 (68,2 %)	141 (76,2 %)	75 (67%)

La valeur méronymique de *avec* est bien plus probable dans le catalogue de jouets, dans les petites annonces immobilières et dans les descriptifs d'itinéraires que dans le roman ou dans le manuel.

**Point de vue qualitatif**

Une caractéristique commune aux trois corpus dans lesquels *avec* méronymique est fréquent peut être mise en évidence : dans tous les cas, la référence à une partie d'un élément (un méronyme) est le support d'une autre information qui fait intervenir la notion de saillance :

Saillance commerciale pour le catalogue et les petites annonces :

*Cuisine avec plaque de cuisson qui rougit vraiment* (catalogue)  
*Villa avec hall de nuit + placards et penderie* (petites annonces)

Saillance visuelle pour les descriptifs d'itinéraires :

*Vous repèrerez la place grâce à une église avec une coupole*

En revanche, un fonctionnement caractérise le catalogue de jouets qui concerne les méronymes introduits par *avec*. Les auteurs sont en général d'accord pour dire que *avec* introduit des méronymes non-essentiels (Cadiot, 1997). Par exemple dans :

*Buanderie avec évier*

Si l'évier est mentionné, c'est qu'il n'est pas toujours présent dans une buanderie. De même dans :

*Maison construction traditionnelle avec cuisine équipée*

La *cuisine équipée* constitue une partie non obligatoire d'une maison (alors que la cuisine ne serait pas mentionnée).

Il en va tout autrement dans les catalogues de jouets. En effet, beaucoup de présentations de jouets cherchent à donner un effet de réel en décrivant les parties que l'on s'attend à trouver dans la version «pour adulte » de l'objet :

*Bloc de cuisson avec hotte, four et plaque de cuisson*

Ces exemples échappent complètement aux descriptions habituelles faites par introspection. Ce fonctionnement apparaît ainsi comme spécifique aux catalogues de jouets ; on ne le retrouve dans aucun des autres textes.

Si le point de départ de l'analyse du fonctionnement des marqueurs de relations conceptuelles concerne l'objectif de la construction d'ontologies, on voit qu'il aboutit à des descriptions qui pourraient relever d'un objectif plus classiquement linguistique. Ainsi, la description de la préposition *avec*, en

tout cas lorsqu'elle est en lien avec la méronymie, apparaît enrichie par la notion de genre textuel (Condamines, à paraître).

#### 4. Conclusion

Les études à partir de corpus que je propose s'inscrivent dans la perspective de la construction d'un sens c'est-à-dire qu'elles supposent une interprétation. Mais les interprétations recherchées font intervenir une dimension collective dont il faut définir la nature. De ce point de vue, la prise en compte du contexte est primordiale : il s'agit de comprendre quelles régularités linguistiques sont dépendantes du contexte (entendu comme situation de production et de réception) et peuvent être décrites en faisant intervenir la notion de genre textuel par exemple et quelles régularités ne sont pas dépendantes du contexte. Dans ce dernier cas, la description peut être considérée comme relevant uniquement du système langue (donc non sensible au contexte). La dimension contextuelle concerne aussi la situation d'interprétation, c'est-à-dire la visée (applicative ou non) qui, tout autant que la situation dans laquelle les textes se sont élaborés, balise cette interprétation. Ainsi, étudier le fonctionnement des prépositions par exemple, dans leur capacité à jouer un rôle de marqueur de relations n'est pas équivalent à étudier le fonctionnement de ces mêmes prépositions « en langue » (c'est-à-dire par introspection), même si des résultats d'analyse peuvent être (fort heureusement) similaires.

Le cas de l'analyse des textes spécialisés, qui s'accompagne souvent d'une situation de production des textes et d'une situation d'interprétation souvent bien identifiées et circonscrite est particulièrement propice pour étudier comment se met en œuvre notre compétence de linguistes ; en effet, nous savons d'emblée que, dans cette situation, nous ne pouvons pas simplement projeter notre connaissance langagière. Ainsi, nous pouvons prendre la mesure du mode d'interprétation que nous mettons en œuvre. Loin de ne constituer qu'une linguistique appliquée, longtemps comprise comme une linguistique de seconde zone, ce mode d'observation ouvre au contraire des perspectives de description nouvelles, qui prennent en compte la réalité des usages.

#### Références bibliographiques

- Ahmad, K. (1993), « Terminology and Knowledge Acquisition : A Text Based Approach », in K.D Schmitz (ed.), *Proceedings of TKE'93 : Terminology and Knowledge Engineering*, Frankfurt, Indeks verlag, pp. 56-70.
- Bachimont, B. (2000), « Engagement sémantique et engagement ontologique : conception et réalisation d'ontologie en ingénierie des

- connaissances », in J. Charlet, M. Zacklad, G. Kassel & D. Bourigault (éds), *Ingénierie des Connaissances, Evolution récentes et nouveaux défis*, Paris, Eyrolles, pp. 305-324.
- Bouquet, S. (éd.) (2004), *Langages* 153, *Les genres de la parole*.
- Cadiot, P. (1997), « Avec, ou le déploiement de l'éventail », in C. Guimier (éd.), *Co-texte et calcul du sens*, Caen, Presses Universitaires de Caen, pp. 135-155.
- Condamines, A. (2000), « Chez dans un corpus de sciences naturelles : un marqueur de méronymie ? », *Cahiers de Lexicologie* 77, pp. 165-187.
- Condamines, A. (2002), « Corpus Analysis and Conceptual Relation Patterns », *Terminology*, volume 8 number 1, pp. 141-162.
- Condamines, A. (à paraître), « Avec et l'expression de la méronymie : l'importance du genre textuel », in G. Kleiber et C. Schnedecker (éd.), *La partition en langue et en discours*.
- Condamines, A. & Rebeyrolle, J. (2001), « Searching for and Identifying Conceptual Relationships via a corpus-based approach to a Terminological Knowledge Base (CTKB) : method and results », in D. Bourigault, M.-C. L'Homme & C. Jacquemin (ed.) : *Recent Advances in Computational Terminology*, Amsterdam/Philadelphia : John Benjamins, pp. 127-148.
- Cruse, D.A. (1986), *Lexical Semantics*, Cambridge, Cambridge University Press.
- Cusin-Berche, F. (1997), « A la recherche de quelques caractéristiques linguistiques des textes spécialisés et de la rédaction technique », *Le langage et l'homme*, Vol. XXXII n°4, *Langues de spécialité et terminologie*, pp. 21-55.
- Fillietaz, L. & Bronckart, J.-P. (éd.) (2005), *L'analyse des actions et des discours en situation de travail. Concepts, méthodes et applications*, Louvain, Peeters.
- Green, R., Bean, C.A. & Hyon Myaeng, S. (2002), *The Semantics of Relationships, An Interdisciplinary Perspective*, Dordrecht/Boston/London, Kluwer Academic Publishers.
- Habert, B. & Zweigenbaum, P. (2002), « Contextual Acquisition of Information Categories : what has been done and what can be done automatically ? », in Nevin B. (ed.), *The Legacy of Zellig Harris : Language and Information into 21<sup>st</sup> century*, volume 2, Amsterdam, John Benjamins.
- Kennedy, G. (1998), *An introduction to Corpus Linguistics*, London and New York, Longman.
- Lerat, P. (1995), *Les langues spécialisées*, Paris, PUF.
- Meyer, I. (2000), « Extracting Knowledge-rich Contexts for Terminography : A Conceptual and methodological Framework », in D. Bourigault, M.-C. L'Homme & C. Jacquemin (ed.), *Recent Advances in Computational Terminology*, John Benjamins, pp. 279-302.

Anne Condamines

- Rastier, F. (1987), *Sémantique interprétative*, Paris, PUF.
- Rebeyrolle, J. & Tanguy, L. (2000), « Repérage automatique de structures linguistiques en corpus : le cas des énoncés définitoires », in A. Condamines (éd.), *Cahiers de Grammaire 25, Sémantique et Corpus*, pp. 153-174.