LA GRAMMAIRE CATÉGORIELLE COMBINATOIRE APPLICATIVE APPLIQUÉE AU FRANÇAIS

Ismaïl BISKRI^(1,2), Jean Pierre DESCLÉS⁽²⁾, Christophe JOUIS^(3,2)

(1) Laboratoire de l'Analyse Cognitive de l'Information, Université du Québec à Montréal, Canada; (2) CAMS-LaLIC, Université de Paris Sorbonne, France;

(3) IDIST, Université de Lille 3, France

1. LE MODÈLE DE LA GRAMMAIRE APPLICATIVE ET COGNITIVE

La Grammaire Applicative et Cognitive (Desclés 1990) est une extension de la Grammaire Applicative Universelle (Shaumyan 1987; Desclés 1990). Elle postule trois niveaux de description des langues :

a- le niveau phénotypique (ou le phénotype) où sont représentées les caractéristiques particulières des langues naturelles (par exemple, l'ordre des mots, les cas morphologiques, etc.). Les expressions linguistiques de ce niveau sont des unités linguistiques concaténées, la concaténation est notée par : 'u₁-u₂-...-u_n';

b- le niveau génotypique (ou le génotype) où sont exprimés les invariants grammaticaux et les structures sous-jacentes aux énoncés du niveau phénotypique. Le niveau génotypique est structuré comme un langage formel appelé «langage génotype»; il est décrit par une grammaire appelée «grammaire applicative»;

c- le niveau cognitif où sont représentées les significations des prédicats lexicaux par des schèmes sémantico-cognitifs.

Les trois niveaux font appel à des formalismes applicatifs typés où l'opération d'application d'un opérateur à un opérande est considérée comme primitive. Les niveaux deux et trois s'expriment dans le formalisme de la logique combinatoire typée de H.B. Curry (1958). Cette logique fait appel à des opérateurs abstraits — appelés «combinateurs» — qui permettent de composer intrinsèquement des opérateurs plus élémentaires entre eux (Desclés 1990). Les combinateurs sont associés à des règles d'introduction et d'élimination. Ceux que nous utilisons 1 sont 1 8, 1 9, 1 9, avec les règles suivantes (1 9, 1 9, 1 9, sont des expressions applicatives typées):

¹ Le combinateur C* est souvent noté T.

Ces règles conduisent à une autre formulation sous forme de règle de réduction (β -réduction) ou d'expansion (β -expansion) :

$$((\mathbf{B} \ \mathbf{U}_1 \ \mathbf{U}_2) \ \mathbf{U}_3) \ge (\mathbf{U}_1 \ (\mathbf{U}_2 \ \mathbf{U}_3))$$

$$((\mathbf{C}_* \ \mathbf{U}_1) \ \mathbf{U}_2) \ge (\mathbf{U}_2 \ \mathbf{U}_1)$$

$$((\Phi \ \mathbf{U}_1 \ \mathbf{U}_2 \ \mathbf{U}_3) \ \mathbf{U}_4) \ge (\mathbf{U}_1 \ (\mathbf{U}_2 \ \mathbf{U}_4)(\mathbf{U}_3 \ \mathbf{U}_4))$$

Dans ce qui suit, nous nous intéresserons aux relations entre les deux premiers niveaux (le phénotype et le génotype) en mettant en oeuvre un système d'analyse formelle — appelée : Grammaire Catégorielle Combinatoire Applicative (GCCA) — qui relie explicitement les expressions phénotypiques à leurs représentations sous-jacentes dans le génotype². Ce système consiste en :

- (i) l'analyse syntaxique des expressions concaténées du phénotype par une Grammaire Catégorielle Combinatoire;
- (ii) la construction à partir du résultat de l'analyse syntaxique d'une interprétation sémantique fonctionnelle des expressions phénotypiques.

1.1 Les Grammaires Catégorielles

Des chercheurs par leurs travaux ont marqué l'évolution des Grammaires Catégorielles :

- (i) Husserl (1913) : Catégories de significations.
- (ii) Lesniewski (1922): Catégories sémantiques.
- (iii) Ajdukiewicz (1935), Bar-Hillel (1953): Grammaires Catégorielles.
- (iv) Lambek (1958, 1961): Calcul de Lambek.
- (v) Steedman (1989): Grammaire Catégorielle Combinatoire.

Les Grammaires Catégorielles assignent des catégories syntaxiques à chaque unité linguistique. Les catégories syntaxiques sont des types orientés engendrés à partir de types de base et de deux opérateurs constructifs '/' et '\'.

- (i) N (syntagme nominal) et S (phrase) sont des types de base.
- (ii) Si X et Y sont des types orientés alors X/Y et X\Y sont des types orientés³.

² Dans le phénotype, les expressions linguistiques sont concaténées selon les règles syntagmatiques du français. Dans le génotype, les expressions sont agencées selon l'ordre applicatif.

³ Nous choisissons ici la notation de Steedman (1989) : X/Y et X\Y sont des types orientés fonctionnels. Une unité linguistique 'u' avec le type X/Y (respectivement X\Y) est considérée comme un opérateur (ou une fonction) dont l'opérande de type Y est positionné à droite (respectivement à gauche) de l'opérateur.

Une unité linguistique u de type orienté X sera désigné par '[X : u]'.

Les deux règles d'application (avant et arrière) sont notées :

Les prémisses dans chaque règle sont des concaténations d'unités linguistiques à types orientés considérées comme étant des opérateurs ou des opérandes, la conséquence de chaque règle est une expression applicative avec un type orienté.

La Grammaire Catégorielle Combinatoire (Steedman 1989) généralise les Grammaires Catégorielles classiques en introduisant des opérations de changement de type et des opérations de composition des types fonctionnels. Les nouvelles règles proposées visent une analyse pseudo-incrémentale (de gauche à droite) pour éliminer le problème de la pseudo-ambiguïté (Haddock 1987; Pareschi, Steedman 1987).

1.2 La Grammaire Catégorielle Combinatoire Applicative

Nous considérons dans la GCCA que les règles de la Grammaire Catégorielle Combinatoire de Steedman introduisent les combinateurs \mathbf{B} , \mathbf{C}_* , \mathbf{S} dans la séquence syntagmatique. Cette introduction permet de passer d'une structure concaténée à une structure applicative.

Les règles de la GCCA sont :

Les prémisses des règles sont des expressions concaténées typées; les résultats sont des expressions applicatives (typées) avec éventuellement introduction d'un combinateur. Le changement de type d'une unité u introduit le combinateur \mathbf{C}_* ; la composition de deux unités concaténées introduit le combinateur \mathbf{B} et \mathbf{S} . Avec ces règles nous pouvons analyser une phrase au moyen d'une stratégie quasi-incrémentale «de gauche à droite». Le choix d'une telle stratégie est motivé par le contrôle du problème de la pseudo-ambiguïté (Pareschi, Steedman 1987; Steedman

aime

Exemple:

Jean
-

[N : Jean] ---->**T**

1989).

 $[S/(S\N): (C_* Jean)]$ - $[(S\N)/N: aime]$

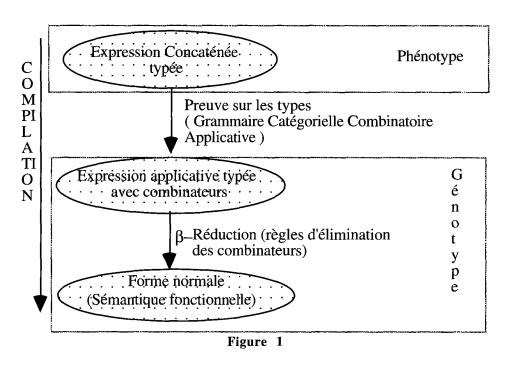
 $[S/N : (B (C_* Jean) aime)]$

La première règle (>T) appliquée à l'unité typée [N: Jean] transforme l'opérande en opérateur. Elle construit une structure applicative (\mathbf{C}_* Jean) ayant pour type S/(S\N). L'introduction du combinateur \mathbf{C}_* illustre dans la représentation applicative le changement de type : (\mathbf{C}_* Jean) fonctionne comme un opérateur avec son type fonctionnel. La règle (>B) combine les unités linguistiques typées [S/(S\N): (\mathbf{C}_* Jean)] et [(S\N)/N: aime] avec le combinateur \mathbf{B} de façon à pouvoir composer les deux unités fonctionnelles (\mathbf{C}_* Jean) et aime.

Un traitement complet basé sur la Grammaire Catégorielle Combinatoire Applicative s'effectue en deux grandes étapes :

- (i) la première étape s'illustre par la vérification de la bonne connexion syntaxique et la construction de structures prédicatives avec des combinateurs introduits à certaines positions de la chaîne syntagmatique;
- (ii) la deuxième étape consiste à utiliser les règles de β -réduction des combinateurs de façon à former une structure prédicative sous-jacente à l'expression phénotypique. L'expression obtenue est applicative et appartient au langage génotype. La GCCA engendre des processus qui associent une structure applicative à une expression concaténée du phénotype. Il nous reste à éliminer les combinateurs de l'expression obtenue de façon à construire la «forme normale» (au sens technique de la β -réduction) qui exprime l'interprétation sémantique fonctionnelle. Ce calcul s'effectue entièrement dans le génotype.

Le traitement que nous proposons donc prend la forme d'une compilation dont les étapes sont résumées dans la figure 1 :



Traitons un exemple simple : Jean aime Marie $[N : Jean] - [(S \setminus N)/N : aime] -$ [N:Marie] Structure concaténée typée du phénotype 2 $[S/(S\setminus N):(C_* Jean)]-[(S\setminus N)/N:aime]$ -(>T) $[S/N : (B (C_* Jean) aime)] - [N : Marie]$ (>B)Compilation $[S:((\mathbf{B}(\mathbf{C}_{\star} Jean) aime) Marie)]$ (>) 5 $[S:((\mathbf{B}(\mathbf{C}_{\star} Jean) aime) Marie)]$ Structure applicative typée du génotype 6 $[S:((\mathbf{C}_{\star} Jean) (aime Marie))]$ **(B)** 7 [S: ((aime Marie) Jean)] (C,) Forme normale du génotype

Le changement de type (>T) affectant l'opérande Jean permet d'engendrer l'opérateur (C_*Jean) que la règle fonctionnelle (>B) compose avec l'opérateur aime. L'opérateur complexe (B (C_*Jean) aime) s'applique à l'opérande Marie pour former

l'expression applicative du génotype ((\mathbf{B} (\mathbf{C}_* Jean) aime) Marie). La réduction des combinateurs dans le génotype construit l'interprétation sémantique fonctionnelle sousjacente à l'expression phénotypique en entrée.

Les règles catégorielles combinatoires applicatives ne sont pas conçues pour réaliser un simple calcul. Un principe purement applicatif motive la conception de telles règles. Rappelons qu'avec les Grammaires Catégorielles, les unités linguistiques sont considérées comme des opérateurs avec leurs opérandes. De ce fait, avec les règles d'applications fonctionnelles nous formalisons l'application d'un opérateur à son opérande, avec les règles de composition nous composons deux foncteurs pour construire un foncteur complexe.

2. LA RÉORGANISATION STRUCTURELLE

L'analyse syntaxique «de gauche à droite» soulève le problème du nondéterminisme introduit par la présence dans la langue, de modifieurs arrières qui sont des opérateurs qui s'appliquent à l'ensemble ou à une partie d'une structure préalablement construite.

Si dans le premier cas l'utilisation d'une règle d'application permet la poursuite de l'analyse⁴, il en est autrement pour le second cas où l'analyse «patine». Pour une phrase comme Jean aime Marie tendrement l'analyseur produit dans un premier temps le constituant [S:((B(C*, Jean) aime) Marie)]. Ce dernier n'est pas combinable avec tendrement, de type (S\N)\(S\N). En effet, tendrement est un opérateur qui a comme opérande (aime Marie) positionné à sa gauche. Une analyse quasi-incrémentale de «gauche à droite» favorise l'application d'une règle combinatoire dès que possible. Ce facteur a pour conséquence directe de «noyer» aime et Marie dans ((B (C*, Jean) aime) Marie), ce qui évidemment ne nous permet pas de construire directement l'opérande (aime Marie).

Le problème posé revient à la possibilité d'un retour arrière. Mais ce retour arrière est de nature à accroître le coût «computationnel» (mémoire et temps d'exécution) d'une analyse syntaxique. Cependant, un retour arrière «intelligent» (que nous allons proposer plus loin) peut nous permettre de réduire considérablement ce coût tout en construisant des analyses sémantiques correctes et en éliminant les pseudo-ambiguïtés.

Ainsi un tel retour arrière décomposera le constituant déjà construit en deux composantes dont une se combine forcément avec le modifieur arrière.

Formellement, cette opération de réorganisation structurelle s'effectue par les deux étapes successives suivantes :

⁴ Prenons l'exemple de la phrase Jean frappa Marie hier où le modifieur arrière hier opère sur l'ensemble de la phrase Jean frappa Marie; hier étant de type syntaxique S\S, il suffit, pour poursuivre l'analyse, d'appliquer hier à Jean frappa Marie par la règle (<).

⁵ C'est à dire que Marie n'apparaît pas comme l'opérande directe de l'opérateur aimer.

a- La réorganisation du constituant déjà construit isole à chaque fois deux souscatégories, et teste si le modifieur arrière se combine à gauche⁶, ou pas, avec une de ces deux sous-catégories. Nous procédons ensuite à la réduction des combinateurs jusqu'à ce que le test nous donne une valeur positive. À la fin du processus nous récupérons une nouvelle structure applicative typée «équivalente» à la première.

Exemple : Dans le cas de l'énoncé Jean aime Marie tendrement, les étapes de la réorganisation sont :

Le constituant construit : [S : ((B (C* Jean) aime) Marie)]

Les deux sous-catégories sont : [S/N : (B (C* Jean) aime)]; [N : Marie]

Test: $[S/N : (B(C_* Jean) aime)]$ ne se combine pas à gauche avec

 $[(S\N)\(S\N) : tendrement]$

[N: Marie] ne se combine pas à gauche avec

 $[(S\N)\(S\N) : tendrement]$

Réduction du combinateur \mathbf{B} : [S : ((\mathbf{C}_* Jean) (aime Marie))]

 $Les \ deux \ sous-catégories \ sont: \qquad [S/(S\N):(C_*\ \textit{Jean})];$

 $[S\N : (aime\ Marie)]$

Test: $[S/(S\N) : (C_* Jean)]$ ne se combine pas à gauche avec

 $[(S\N)\(S\N) : tendrement]$

[S\N: (aime Marie)] se combine à gauche avec

 $[(S\N)\(S\N) : tendrement]$

Arrêt du processus de réduction des combinateurs. Nous récupérons en sortie la catégorie :

$$[S:((C_* Jean) (aime Marie))].$$

b- La décomposition réalisée grâce aux deux règles⁷ :

Nous lisons ces règles comme suit :

⁶ Dans notre terminologie, u_1 se combine à gauche avec u_2 si une des règles suivantes <, <B, <Bx, <S, <Sx, peut composer les types de u_2 et u_1 ou si une des règles combinatoires peut composer les types de $(C_* u_2)$ et u_1 .

⁷ Notons que les deux règles (>dec) et (<dec) sont respectivement inverses des règles d'application fonctionnelle (>) et (<).

- Pour (>dec): Si nous avons une structure applicative (u1 u2) de type X, avec u1 de type X/Y et u2 de type Y, alors nous pouvons construire une nouvelle expression concaténée formée des deux catégories [X/Y:u1] et [Y:u2].
- Pour (**<dec**): Si nous avons une structure applicative (u₁ u₂) de type X, avec u₁ de type X\Y et u₂ de type Y, alors nous pouvons construire une nouvelle expression concaténée formée des deux catégories [Y:u₂] et [X\Y:u₁].

Ces deux règles nous permettent de reconstruire un nouvel agencement concaténé de la structure opérateur/opérande issue de la réorganisation.

Pour la phrase *Jean aime Marie tendrement* la décomposition est appliquée à la structure qui résulte de la réorganisation :

$$[S:((C_* Jean) (aime Marie))]$$

Avec la règle (>dec), nous produisons l'agencement concaténé :

$$[S/(S\N) : (C_* Jean)] - [S\N : (aime Marie)].$$

Ces deux étapes entrent dans l'analyse complète de la phrase *Jean aime Marie tendrement* comme suit (étape 5 pour la réorganisation et étape 6 pour la décomposition) :

Structure concaténée typée du phénotype

1	$[N: Jean] - [(S\N)/N: aime] - [N: Marie] - [(S\N) \setminus (S\N): tendrement]$	_
4	$[S:((\mathbf{B}(\mathbf{C}_{*} Jean) \ aime) \ Marie)]-[(S\N)\(S\N): \ tendrement]$	
5	$[S:((C_*Jean)(aime\ Marie))]-[(S\N)(S\N):tendrement]$	(a)
6	$[S/(S\N): (C_* Jean)] - [S\N: (aime\ Marie)] - [(S\N)\(S\N): tendrement]$	(b)
7	$[S/(S\N): (C_* Jean)]-[S\N: (tendrement (aime Marie))]$	(<)
8	[S: ((C _* Jean) (tendrement (aime Marie)))]	(>)

Structure applicative typée du génotype

9	[S: $((C_*Jean)$ (tendrement (aime Marie [S: $((tendrement (aime Marie)))$	e)))]	
10	[S: ((tendrement (aime Marie)) Jean)]	(C _*) Forme normale du génotype	

3. LA COORDINATION

La coordination est l'action de joindre deux mots ou deux expressions du même genre ou ayant même fonction. Dans le cadre des Grammaires Catégorielles, Steedman (1989), Barry et Pickering (1990) considèrent que deux unités linguistiques peuvent être coordonnées pour donner une unité linguistique de type X si et seulement si chaque unité est de type X. Même si cette définition reste incomplète, sachant que la coordination se

présente sous différentes formes, elle indique de manière idéale la voie à suivre pour établir une solution fiable.

Présentons quatre types d'exemples de coordination par ET. Nous pouvons coordonner 8 :

- a) Deux segments de même type, de même structure et contigus à ET : [Jean aime]_{S/N} et [Paul déteste]_{S/N} ces tableaux
- b) Deux segments dans une construction elliptique : Jean aime [Marie tendrement] et [Sophie sauvagement] [Jean] aime [Marie] et [Paul Sophie]
- c) Deux segments de structures différentes : Marie marche [doucement] et [avec élégance] Jean [chante] et [joue du violon]
- d) Deux segments, sans distributivité:
 Le drapeau est [blanc] et [rouge]
 (≠ Le drapeau est blanc et le drapeau est rouge).

À la conjonction 'ET' on associe le type polymorphique $(X\setminus X)/X$. Le contexte donne cependant plus de spécifications pour assigner un type à 'ET'.

Les hypothèses 1 et 2 permettent d'assigner un type à ET en tenant compte du contexte.

Hypothèse1 : La catégorie construite qui suit immédiatement la conjonction ET détermine le type de la coordination.

Cette hypothèse nous amène indirectement à introduire une interruption dans l'analyse quasi-incrémentale : dès que nous rencontrons la conjonction ET nous interrompons momentanément l'analyse quasi-incrémentale pour construire le second membre de la coordination.

Nous revenons ensuite en arrière pour déterminer le premier membre de la coordination.

Nous proposons la seconde hypothèse :

Hypothèse 2 : Quand nous avons une coordination de type X définie par l'hypothèse 1, le premier membre de la coordination est la catégorie de type X qui précède immédiatement la conjonction.

Les règles que nous devons dégager à travers ces deux hypothèses partent donc de l'idée que les deux membres de la coordination ont des types syntaxiques X identiques correspondant à des interprétations sémantiques fonctionnelles différentes. Le résultat de

⁸ Les catégories à coordonner sont entre crochets.

l'application de ces règles conserve le même type syntaxique X. Nous établissons deux types abstraits pour la conjonction. Le premier concerne la conjonction distributive, nous le noterons CONJD. Le deuxième type concerne la conjonction non distributive et nous le noterons CONJN.

Nous appliquons la règle **<CONJD>** aux cas de la coordination distributive. Pour prendre en compte la distributivité au niveau de la structure applicative, nous utilisons le combinateur Φ . Nous appliquons la règle **<CONJN>** aux cas de la coordination non distributive (voir exemple E3).

Avec l'analyse quasi-incrémentale, lors de l'application de l'hypothèse 2, deux cas de figure se présentent :

- Le constituant produit avant de rencontrer la conjonction est du même type que le constituant déterminé par la coordination. Ce constituant est alors le premier membre de la coordination. Par exemple, l'analyse de la phrase : [Jean aime]_{S/N} et [Paul déteste]_{S/N} ces tableaux construit [S/N : (B(C_{*}Jean) aime)] avant de rencontrer la conjonction. Le type est identique au type du second membre [S/N : (B(C_{*}Paul) déteste)], le constituant déterminé par la première hypothèse. Le constituant [S/N : (B(C_{*}Jean) aime)] est alors le premier membre de la coordination.
- Le constituant déterminé avant de rencontrer la conjonction n'est pas du même type que le constituant déterminé par la coordination. Il est nécessaire de modifier la structure de ce constituant. Par exemple, l'analyse de la phrase : Jean aime [Marie tendrement] et [Sophie sauvagement] construit [S:((C*, Jean) (tendrement (aime Marie)))] avant l'analyse de la conjonction. Le second membre de la coordination étant [(S\N)\((S\N)/N) : (B* sauvagement (C*, Sophie))]⁹.

Dans ce second cas le processus de réorganisation structurelle nous permet :

- soit d'isoler directement le premier membre de la coordination¹⁰.
- soit d'isoler la structure binaire «opérateur/opérande» qui contient le premier membre de la coordination. Dans cette deuxième situation, il est nécessaire d'adjoindre à la réorganisation structurelle l'utilisation des équivalences logiques¹¹ de la logique combinatoire (e,f,g,h)¹²:

(e)
$$(u1 (u2 u3))$$
 <==> $((B u1 u2) u3)$

⁹ L'énoncé Jean aime Marie tendrement et Sophie sauvagement est ambigu. Dans notre exemple, nous considérons que Sophie est objet.

¹⁰ Voir les étapes 6 et 7 de l'exemple E1.

¹¹ Ces équivalences sont les conséquences directes de l'introduction et l'élimination des combinateurs ${\bf B}$ et ${\bf C_*}$.

¹² Voir l'étape 8 de l'exemple E2.

- (f) $((u1 \ u2) \ u3)$ $<==> <math>((B \ (C_* \ u3) \ u1) \ u2)$
- (g) $(u1 (u2 u3)) \iff ((\mathbf{B} u1 (\mathbf{C}_{*} u3)) u2)$
- $\text{((u1 u2) u3)} \qquad <=> \qquad \qquad ((\textbf{B} \ (\textbf{C}_* \ \text{u3}) \ (\textbf{C}_* \ \text{u2})) \ \text{u1})$

4. LES MÉTARÈGLES

Nous enrichissons notre formalisme par différentes métarègles qui contrôlent le changement de type.

Ces métarègles d'une part nous indiquent qu'une règle de changement de type doit être appliquée, et d'autre part choisissent le changement de type particulier à effectuer.

Nous ne concevons pas ces métarègles comme un outil purement informatique, nous leur donnons une pertinence logique et linguistique. Elles peuvent avoir une interprétation si nous tenons compte des facteurs prosodiques.

Nous présentons dans ce qui suit trois métarègles parmi celles que nous avons conçues. Ces dernières sont au nombre de dix¹³.

Soient u₁ et u₂ dans l'expression concaténée 'u₁-u₂':

Métarègle 1: Si u_1 est de type N et u_2 de type $(Y\setminus N)/Z$, alors nous appliquons le changement de type avant (>T) à u_1 :

```
[\mathrm{N}:\mathfrak{u}_1 ==> \mathrm{Y}/(\mathrm{Y}\backslash\mathrm{N}):(\mathbf{C}_*\,\mathfrak{u}_1)]
```

Exemple: Jean mange la pomme

 $[N : Jean]-[(S\backslash N)/N : mange]-[N/N : la]-[N : pomme]$

 $[S/(S\N) : (C_* Jean)]-[(S\N)/N : mange]-[N/N : la]-[N : pomme]$

Dans ce cas Y = S; Z = N.

Métarègle 2: Si u1 est de type N (u1 précédé de et) et u2 de type N, alors nous appliquons le changement de type avant (>T) à u1:

```
[ N : u1 ==> S/(S\N) : (C_* u1) ]
```

Exemple: Jean aime Marie et Paul Sophie ...-[CONJD: et]-[N: Paul]-[N: Sophie]

 \dots -[CONJD : et]-[S/(S\N) : (\mathbb{C}_* Paul)]-[N : Sophie]

Métarègle 3 : Si u2 est de type N et u1 de type Y/X (u1 précédé de et), alors nous appliquons le changement de type arrière (<T) à u2 :

 $[N: u2 ==> X\setminus(X/N): (C_* u2)]$

Exemple: Jean aime Marie et Paul Sophie

¹³ Voir Biskri (1995).

```
...-[CONJD : et]-[S/(S\N) : (\mathbb{C}_* Paul)]-[N : Sophie]
...-[CONJD : et]-[S/(S\N) : (\mathbb{C}_* Paul)]-[(S\N)\((S\N)/N) : (\mathbb{C}_* Sophie)]
Dans ce cas X = S\N; Y = S.
```

5. EXEMPLES

E1 : Jean aime Marie et déteste Sophie

Phénotype

1	$[N:Jean]-[(S\N)/N:aime]-[N:Marie]-[CONJD:et]-[(S\N)/N:déteste]-[N:Sophie]$	
4	[S:((B (C _* Jean) aime) Marie)]-[CONJD:et]-[(S\N)/N:déteste]-[N:Sophie]	
5	[S:((B (C _* Jean) aime) Marie)]-[CONJD:et]-[S\N:(déteste Sophie)]	(>)
6	$[S:((C_* Jean) (aime Marie))]-[CONJD:et]-[S\N:(déteste Sophie)]$	(B)
7	$[S/(S\N):(C_*\ Jean)]-[S\N:(aime\ Marie)]-[CONJD:et]-[S\N:(déteste\ Sophie)]$	(>dec)
8	$[S/(SN):(C_* Jean)]-[SN:(\Phi \ et \ (aime \ Marie) \ (déteste \ Sophie))]$	(<conjd>)</conjd>
9	$[S:((C_* Jean) (\Phi \ et \ (aime Marie) \ (déteste Sophie)))]$	(>)

Génotype

10	$[S:((C_*Jean)(\Phi\ et\ (aime\ Marie)\ (déteste\ Sophie)))]$	
11	[S : ((Φ et (aime Marie) (déteste Sophie)) Jean)]	(C _*)
12	[S: (et ((aime Marie) Jean) ((déteste Sophie) Jean))]	(Φ)

E2 : Jean aime Marie et Paul Sophie

Phénotype

1	[N:Jean]-[(S\N)/N:aime]-[N:Marie]-[CONJD:et]-[N:Paul]-[N:Sophie]	
4	[S:((B (C * Jean) aime) Marie)]-[CONJD:et]-[N:Paul]-[N:Sophie]	
5	$ \hbox{-} [\hbox{CONJD} \hbox{:} et] \hbox{-} [\hbox{S/(S\N)} \hbox{:} (\hbox{\mathbb{C}_* } Paul)] \hbox{-} [\hbox{N} \hbox{:} Sophie]$	(>T), M2
6	$[CONJD:et]-[S/(S\backslash N):(\mathbf{C_*}\ Paul)]-[(S\backslash N)\backslash(S/(S\backslash N)):(\mathbf{C_*}\ Sophie)]$	(<t), m3<="" th=""></t),>
7	[CONJD: et]-[S\(S\(S\N)):(B (\mathbb{C}_* Paul)(\mathbb{C}_* Sophie)]	(>Bx)
8	$[S:((B(C_*Jean)(C_*Marie)) aime)]-[CONJD:et]$	(h)
9	$[(S\N)/N:aime]-[S\(S/(S\N)):(\mathbf{B}\ (\mathbf{C_*}\ Jean)\ (\mathbf{C_*}\ Marie))]-[CONJD:et]$	(<dec)< th=""></dec)<>
10	$[-[S\setminus (S/(S\setminus N)): (\Phi\ et\ (B(C_*\ Jean)(C_*\ Marie))(B(C_*\ Paul)(C_*\ Sophie))]$	(<conjd></conjd>)
11	$[S:((\Phi\ et\ (\mathbf{B}\ (\mathbf{C_*}\ Jean)\ (\mathbf{C_*}\ Marie))(\mathbf{B}\ (\mathbf{C_*}\ Paul)(\mathbf{C_*}\ Sophie)))\ aime)]$	(<)

Génotype

```
[S: ((\Phi \ et \ (B \ (C_{\downarrow} \ Jean) \ (C_{\downarrow} \ Marie))(B \ (C_{\downarrow} \ Paul)(C_{\downarrow} \ Sophie))) \ aime)]
13 [S: (et ((\mathbf{B} (\mathbf{C}_{*} Jean) (\mathbf{C}_{*} Marie)) aime) ((\mathbf{B} (\mathbf{C}_{*} Paul)(\mathbf{C}_{*} Sophie)) aime))]
                                                                                                                            (Φ)
14 [S: (et((C_*Jean)((C_*Marie)aime))((B(C_*Paul)(C_*Sophie))aime))]
                                                                                                                            (B)
15 [S: (et(((\mathbf{C}_{\downarrow}Marie)\ aime)\ Jean)\ ((\mathbf{B}(\mathbf{C}_{\downarrow}Paul)(\mathbf{C}_{\downarrow}Sophie))\ aime))]
                                                                                                                            (\mathbf{C}_{\bullet})
16 [S: (et ((aime Marie) Jean) ((B (C, Paul)(C, Sophie)) aime))]
                                                                                                                            (\mathbf{C}_{\bullet})
17 [S: (et ((aime Marie) Jean) ((\mathbb{C}_* Paul)((\mathbb{C}_* Sophie) aime)))]
                                                                                                                            (B)
18 [S: (et ((aime Marie) Jean) (((C, Sophie) aime) Paul))]
                                                                                                                            (\mathbf{C}_{\bullet})
     [S: (et ((aime Marie) Jean) ((aime Sophie) Paul))]
19
                                                                                                                            (\mathbf{C}_{\bullet})
```

E3: le drapeau est blanc et rouge

≠> le drapeau est blanc et le drapeau est rouge

Phénotype

1	$[N/N:le]-[N:drapeau]-[(S\N)/(N\N):est]-[N\N:blanc]-[CONJN:et]-[N\N:roug]-[NN:roug]-[$	e]
2 3	$ [N:(le\ drapeau)]-[(S\N)/(N\N):est]-[N\N:blanc]-[CONJN:et]-[N\N:rouge] \\ [S/(S\N):(\mathbf{C}_*\ (le\ drapeau))]-[(S\N)/(N\N):est]-[N\N:blanc]-[CONJN:et] $	(>T), M1
4	$[S/(N/N):(\textbf{B}\ (\textbf{C}_{*}\ (le\ drapeau))\ est)]-[N\N:blanc]-[CONJN:et]-[N\N:rouge]$	(>B)
5	$[S:((\textbf{B} (\textbf{C}_{*}(\textit{le drapeau})) \textit{est}) \textit{blanc})]-[\texttt{CONJN}:\textit{et}]-[\texttt{N}\backslash \texttt{N}:\textit{rouge}]$	(>)
6	$[S/(N/N):(\textbf{B}\ (\textbf{C}_{*}\ (le\ drapeau))\ est)]-[N\N:blanc]-[CONJN:et]-[N\N:rouge]$	(>dec)
7	$[S/(N/N):(\textbf{B} (\textbf{C}_{*} (le \ drapeau)) \ est)]-[N/N:(et \ blanc \ rouge)] \tag{$<$}$	CONJN>)
8	$[S:((\mathbf{B}(\mathbf{C}_{*}(le\ drapeau))\ est)\ (et\ blanc\ rouge))]$	(>)

Génotype

9	[S: ((B (C* (le drapeau)) est) (et blanc rouge))]	
	$[S:((C_*(le\ drapeau))\ (est\ (et\ blanc\ rouge)))]$	(B)
11	[S: ((est (et blanc rouge)) (le drapeau))]	(C _*)
	<u> </u>	

D'autres exemples et plus de détails sont donnés dans (Biskri 1995). Les analyses sont implémentées. Nous ne donnons pas ici les détails de l'algorithme.

6. CONCLUSION

Avec ce travail, nous sommes arrivés à la conviction que les Grammaires Catégorielles associées à un formalisme applicatif peuvent rendre compte correctement des langues naturelles. Même si, beaucoup de choses restent à faire pour établir un système prêt à répondre à des besoins industriels, nous pensons que ce que nous présentons est

assez solide pour être considéré comme un modèle de traitement des langues naturelles relativement fiable.

Ce modèle, d'ailleurs, présente des qualités que nous jugeons importantes. En premier lieu, nous trouvons qu'il est élégant dans le sens où il permet un suivi de traitement très aisé. Deuxièmement, l'utilisation combinée des types syntaxiques et des combinateurs permet de véhiculer un maximum d'informations avec un minimum de ressources (variables). Cette propriété, même si nous sommes en face uniquement d'une étude théorique, a l'avantage de réduire (voire éliminer) des traitements que nous dirons superflus, comme celui fondé sur l'unification pour retrouver les interprétations sémantiques dans le modèle de Steedman et Pareschi (1987).

Par ailleurs, le modèle de la GCCA prend la forme d'une compilation. On ne fait pas un traitement sémantique parallèle à un traitement syntaxique. La construction de l'interprétation sémantique fonctionnelle se fait dans la continuité du calcul syntaxique avec le passage du phénotype au génotype.

Comme nous l'avons signalé précédemment, ce travail s'inscrit dans un cadre de travail beaucoup plus général. D'autres traitements dans le génotype suivent celui que nous avons présenté. Nous pensons en particulier au traitement des paraphrases, de la passivation et de la reflexivation etc. Ainsi, nous pouvons conseiller au lecteur de consulter (Desclés 1990) pour comprendre comment peuvent se faire ces genres de traitements.

Pour résumer, ce modèle est capable de réaliser pour un noyau du français, les objectifs suivants :

- (i) produire une analyse qui vérifie la correction syntaxique des énoncés;
- (ii) engendrer automatiquement des structures prédicatives qui rendent compte de l'interprétation sémantique fonctionnelle des énoncés.

Ce modèle a en outre les propriétés suivantes :

- 1- Un premier calcul vérifie la correction syntaxique, ce calcul se poursuit par une construction de l'interprétation sémantique fonctionnelle. Cela est rendu possible par l'introduction des combinateurs à des positions spécifiques de la chaîne syntagmatique.
- 2- Nous introduisons des éléments de nature sémantique fonctionnelle par des outils syntaxiques applicatifs (les combinateurs).
- 3- Nous calculons l'interprétation sémantique fonctionnelle par des méthodes syntaxiques applicatifs (la réduction des combinateurs).

Pour résumer, nous interprétons à l'aide de techniques purement syntaxiques. La distinction syntaxe/sémantique devrait alors être repensée dans une autre perspective.

RÉFÉRENCES

- ADES, A., STEEDMAN, M. (1982): «On the order of words», *Linguistics and Philosophy*, 4, pp. 517-558.
- BARRY, G., PICKERING, M. (1990): «Dependancy and constituency in categorial grammar», dans Lecomte 1992, L'ordre des mots dans les grammaires catégorielles, pp. 38-57.
- BISKRI, I. (1995): La Grammaire Catégorielle Combinatoire Applicative dans le cadre de la Grammaire Applicative et Cognitive, Thèse de Doctorat, EHESS, Paris.
- BISKRI, I., MEUNIER, J.G., NAULT, G. (1997): «Extraction des connaissances terminologiques au moyen des Grammaires Catégorielles: un modèle hybride», Actes du colloque JST97, Avignon, France.
- CURRY, B.H., FEYS, R. (1958): Combinatory logic, Vol. I, North-Holland.
- DESCLÉS, J.P. (1990): Langages applicatifs, langues naturelles et cognition, Paris, Hermes.
- DESCLÉS, J.P., SEGOND, F. (1990): «Topicalization: categorial analysis and applicative grammar», dans Lecomte 1992, L'ordre des mots dans les grammaires catégorielles, pp. 13-37.
- DESCLÉS, J.P., BISKRI, I. (1996): «Logique combinatoire et linguistique: Grammaire Catégorielle Combinatoire Applicative», Revue Mathématiques, Informatiques et Sciences Humaines, Paris.
- HADDOCK, N. (1987): «Incremental interpretation and Combinatory Categorial Grammar», dans N. Haddock, E. Klein, et G. Morill, Working papers in cognitive science volume I: Categorial Grammar, Unification Grammar and Parsing, Edinburgh Univ., pp. 71-84.
- LECOMTE, A. (1994): Modèles Logiques en Théorie Linguistique: éléments pour une théorie informationnelle du langage, Synthèse de travaux présentés en vue de l'habilitation à diriger des recherches.
- MOORTGAT, M. (1988): Categorial Investigation, Logical and Linguistic Aspects of the Lambek Calculus, Foris Publications.
- OEHRLE, R.T., BACH, E., WHEELER, D. (1988): Categorial Grammars and Natural Languages Structures, D. Reidel.
- PARESCHI, R., STEEDMAN, M. (1987): «A lazy Way to chart parse with categorial grammars», Acte du colloque ACL, Stanford.
- SHAUMYAN, S.K. (1987): A Semiotic Theory of Natural Language, Bloomington, Indiana Univ. Press.
- STEEDMAN, M. (1989): Work in progress: Combinators and grammars in natural language understanding, Summer Institute of Linguistic, Tucson University.
- SZABOLCSI, A. (1987): «On combinatory categorial grammar», Actes du Symposium on logic and languages, Budapest, Debrecen, Akademiai Kiado, pp. 151-162.

LA FORME DES MOTS NOUS RENSEIGNE-T-ELLE SUR LEUR NATURE ?

Sylviane CARDEY⁽¹⁾, Zahra EL HAROUCHY⁽¹⁾, Peter GREENFIELD⁽²⁾

(1) Centre de recherche Lucien Tesnière; (2) Laboratoire d'Informatique de Besançon, Besançon, France

Une des étapes du traitement automatique des langues (souvent la première) consiste à déterminer à quelle catégorie grammaticale appartient chacune des unités ou chacun des «mots» du texte traité, en d'autres termes, elle consiste à déterminer la **nature des «mots»**. Aussi, nous voudrions montrer ici comment un «mot», par ses caractéristiques morphologiques, peut nous renseigner sur sa nature. Nous présentons un dictionnaire dit en *intention* par opposition aux dictionnaires dits en *extension* souvent utilisés pour essayer de réaliser cette tâche en traitement automatique du français. Ce dictionnaire permet de gagner en temps de traitement et en espace de mémoire pour la machine. Au lieu d'avoir comme entrées, ainsi que nous les trouvons dans les dictionnaires électroniques, chacune des formes canoniques ou fléchies (mot-forme) de la langue française associée à une classe, notre dictionnaire électronique (de très petite dimension) présente des entrées sous forme de caractéristiques morphologiques classées en ensembles et sous-ensembles et ordonnées selon certaines contraintes; chacune des caractéristiques définies étant associée à une des parties du discours.

Toute analyse des langues implique, entre autres tâches, la classification des éléments de la langue traitée. Les définitions des classes de mots appelées aussi, depuis longtemps, parties du discours ou catégories grammaticales, n'ont cessé d'être discutées; nous utilisons le découpage en classes traditionnel : le substantif, l'adjectif, le pronom, le verbe, l'article, l'adverbe, la préposition, la conjonction et l'interjection.

LES UNITÉS TRAITÉES

Nous ne traitons, avec notre dictionnaire, que les mots simples du français. Ce qu'il faut signaler est qu'un mot simple pour une machine (puisque notre travail est réalisé pour être confié à un ordinateur) n'est pas la même chose qu'un mot simple pour un humain. Le mot simple en Traitement Automatique des Langues est une chaîne de caractères comprise entre blanc(s) et/ou signe(s) de ponctuation. Dans les exemples qui suivent : porte, portefeuille, porte-monnaie, porte-à-faux, porte à porte, œil d'une porte, porte la main à, seul 'porte' est un mot simple pour nous alors que pour la machine 'portefeuille' le sera également, voire même dans certains cas 'porte-monnaie' et 'porte-à-faux'. Nous traitons de ce fait les composés soudés.

SUFFIXES ET TERMINAISONS

«On peut classer les suffixes selon leur étymologie mais il est préférable, dans une grammaire décrivant la langue d'aujourd'hui, de les inventorier plutôt en rapport avec leur fonctionnement — c'est-à-dire d'après la nature des mots qu'ils servent à former; suffixes formant des noms et/ou des adjectifs (dérivation appelée souvent nominale), suffixes formant des verbes (dérivation verbale)...» (Grevisse 1988 : 246). La plupart des ouvrages de grammaire mentionnent déjà le fait que certains suffixes donnent une indication sur la catégorie à laquelle un mot appartient; certains suffixes servent à former plutôt des noms, d'autres plutôt des adverbes. Pour n'en retenir que quelques-uns, (Dubois 1995 : 10-15) :

suffixes s	ervant à former des noms	suffixes servant à f	ormer des adjectifs
-ace	populace	-able, -ble	aimable, audible
-ade	orangeade	-aire	solaire, polaire
-age	balayage	-ique	chimique, ironique

suffixe servant à former des adverbes : -ment gentiment suffixes servant à former des verbes : -ailler tournailler -asser rêvasser

Il est intéressant de voir que la grammaire normative classe aussi les suffixes d'après la nature des mots qu'ils servent à former.

Aussi, nous avons voulu tenter l'expérience d'un classement non pas uniquement des suffixes au sens traditionnel mais de ce que nous appelons terminaisons des mots; chacune de ces terminaisons devant par la suite déterminer à quelle(s) catégorie(s) grammaticale(s) les mots qui la possèdent appartiennent. Nos terminaisons ne sont qu'un certain nombre de lettres allant de 1 à 8 qui se trouvent à la fin du mot. On peut se demander s'il y a quelque rapport entre la terminaison ainsi définie et la classe, étymologiquement ou historiquement, parfois la réponse est oui mais le plus souvent, à notre connaissance, il n'y a aucun rapport.

NOTRE DICTIONNAIRE AUTOMATIQUE

La linguistique informatique a demandé et permis la création de dictionnaires à usages multiples. Il existe aujourd'hui différents dictionnaires électroniques; des dictionnaires élaborés spécialement pour le traitement automatique de la langue qu'il ne faut pas confondre avec les dictionnaires d'usage qui ne sont que des variantes des versions sur papier présentées sur support magnétique.

L'architecture d'un dictionnaire, quelle qu'elle soit, est toujours symbolisée par une entrée et une sortie. Cette architecture reflète non seulement la façon dont est conçu le dictionnaire mais également son but. Le produit final, le dictionnaire, ayant une ou plusieurs applications. La construction des dictionnaires électroniques varie : soit le dictionnaire contient l'ensemble des formes canoniques et fléchies soit il génère les différentes formes que peuvent prendre les mots à partir d'une forme canonique donnée et de règles flexionnelles. Il y a d'autres types de dictionnaires mais nous ne citons que ces derniers qui cherchent, entre autres, à atteindre le même but que nous, c'est-à-dire déterminer la ou les catégorie(s) grammaticale(s) à laquelle ou auxquelles un mot peut appartenir.

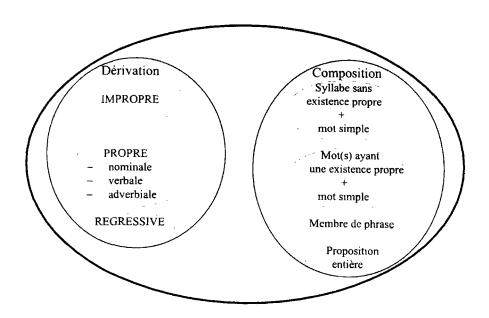
Le dictionnaire présenté ici, que nous appelons dictionnaire automatique, met en œuvre une reconnaissance de la catégorie grammaticale des mots par l'analyse de leur

terminaison. Pour construire ce dictionnaire, nous avons recensé toutes les terminaisons susceptibles d'être retenues, gardé certaines d'entre elles et rejeté les autres en fonction du nombre de «mots» concernés puis nous les avons classées. «La linguistique historique n'explique un mot b par un mot a précédent que si le passage de a à b est le cas particulier d'une règle générale valable pour bien d'autres mots, et fait comprendre ainsi que a' soit devenu b', a" devenu b", etc. Cette régularité implique que la différence entre a et b tient à tel ou tel de leurs constituants et que dans tous les autres mots où ce constituant apparaît, il soit affecté par le même changement» (Ducrot et Schaeffer 1995 : 21). Notre dictionnaire a été créé sur ce principe. Une terminaison, pour être retenue, doit concerner un groupe homogène (mots appartenant à la même catégorie grammaticale). Nous avons relevé, lorsqu'il y en avait, les exceptions, soit les mots appartenant à des catégories grammaticales différentes de la catégorie dominante. Dans le cas où les exceptions étaient trop nombreuses, autrement dit, si les différents mots du groupe, ayant cette terminaison, appartenaient à des catégories disparates, nous n'avons pas tenu compte de la terminaison telle quelle mais avons élargi la terminaison en question à une voyelle et/ou une consonne la précédant afin de voir si ces dernières n'avaient pas une incidence sur la nature des mots.

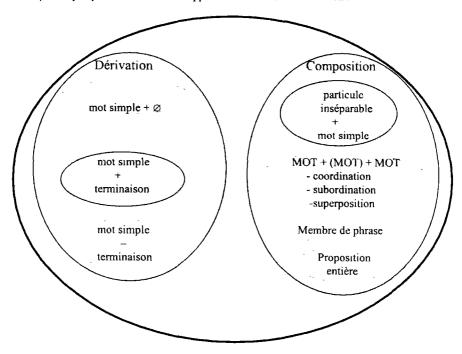
Par exemple nous avons pour la terminaison **aire** : les mots terminés par **aire** sont des *noms* sauf les mots terminés par **baire** qui sont des *adjectifs*, ceux terminés par **faire** qui sont des *verbes à l'infinitif*, ceux en **iaire** qui sont des *adjectifs* et ceux en **laire** qui sont également des *adjectifs*.

LE SYSTÈME DE LA MORPHOLOGIE LEXICALE ET LE DICTIONNAIRE AUTOMATIQUE

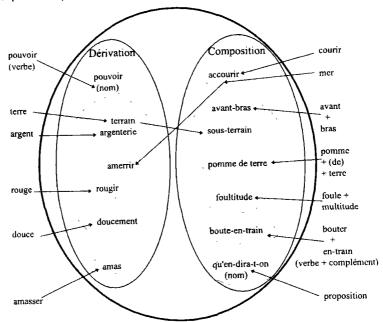
Revoyons rapidement comment fonctionne le système de la morphologie lexicale. Nous parlons en général de dérivation et de composition que l'on peut représenter de la façon suivante (Cardey 1987 : 51-52) :



ou d'une façon un peu plus détaillée faisant apparaître les modes de formation :



Donnons quelques exemples sur sa façon de fonctionner :



En ce qui concerne notre dictionnaire, nous retrouvons certaines formes de dérivations connues mais tout est organisé de façon complètement différente. Prenons un exemple (rappelons que souvent nos terminaisons n'ont rien à voir avec le système dont nous avons l'habitude)

exemple:

nous avons dans notre dictionnaire la règle générale suivante : «les mots terminés par **able** sont des *adjectifs*». Comme exceptions à cette règle générale, nous mettons des *noms* («cartable,...»), des *verbes conjugués* («accable,...»), des *ambiguïtés morphologiques de type nom ou verbe conjugué* («sable, table,...») et des *ambiguïtés de type nom ou adjectif* («comptable,...»).

Un autre exemple, dont la terminaison ne nous est pas familière : (nous ne donnons pas la signification des symboles qui sont bien connus), est donné ci-dessous :

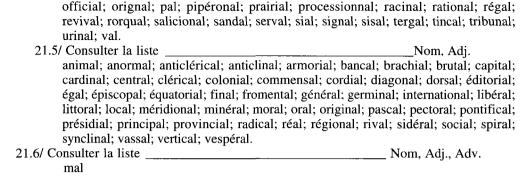
mot se terminant par ague	Nom
liste d'exceptions	Nom, Verbe conj.
bague; blague; dague; drague	-
liste d'exceptions	Nom, Adj., Verbe conj
vague	•
liste d'exceptions	Verbe conj.
rague; zigzague	

Lorsqu'une terminaison a été adoptée, nous lui avons, puisqu'il est important de reconnaître le plus grand nombre de mots possibles, ajouté les flexions (féminin, pluriel, féminin-pluriel,...).

Ainsi que nous le voyons avec la règle des mots terminés par **ague**, le dictionnaire est constitué d'un ensemble de règles. Chaque règle comprend une condition (la terminaison) et un opérateur (la ou les catégorie(s) grammaticale(s)). Chaque règle générale peut être accompagnée de sous-règles qui peuvent être composées d'une terminaison dépendant de la règle générale (**baire** qui dépend de **aire**) et/ou de listes de mots ayant cette même terminaison («blague, bague» qui dépend de **ague**) associées à un ou des opérateur(s) différent(s) de celui de la règle générale. Cette façon de procéder a donné lieu à une hiérarchisation des sous-règles pour une même règle générale. Les règles générales, elles, pouvant se lire dans n'importe quel ordre.

Soit la règle suivante :

21/ Mot se terminant par al	_Adj.
21.1/ Mot se terminant par aal	Nom
21.2/ Mot se terminant par hal	Nom
21.2.1/ Consulter la liste	Adj.
catarrhal; nymphal; triomphal; zénithal	
21.3/ Mot se terminant par oal	Nom
21.4/ Consulter la liste	Nom
amiral; ammonal; arsenal; aval; bacchanal; bal; bocal; cal	l; canal; caporal; caracal;
carnaval; cérémonial; chacal; chenal; cheval; chloral; confe	
dispersal; diurnal; fanal; fécial; festival; foiral; gal; galg	
grémial; journal; madrigal; majoral; mémorial; mistral;	; monial; narval; nopal;



Prenons quelques exemples en vue d'expliquer le fonctionnement et surtout l'ordre de lecture des données.

Le mot «cheval» qui se termine par **al** doit pouvoir être traité par notre règle 21. Nous ne devons pas appliquer l'opérateur Adj., il faut d'abord lire les sous-règles. La catégorie grammaticale de «cheval» ne peut pas être définie par les sous-règles 21.1, 21.2, 21.3. La sous-règle 21.4 est une liste à consulter. Le mot s'y trouve, l'opérateur de cette sous-règle, *Nom*, correspond donc à la catégorie grammaticale à laquelle «cheval» appartient.

Pour le mot «matinal», tout comme pour le mot précédent, nous devons parcourir les sous-règles; aucune d'elles ne peut déterminer la nature de ce mot; aussi l'opérateur Adj. de la règle générale 21 correspondra à la catégorie à laquelle «matinal» appartient.

Prenons maintenant «triomphal», nous avons une sous-règle qui traite des mots en **hal**, cette dernière comprend elle-même une sous-règle où «triomphal» apparaît; «triomphal» sera donc un *adjectif*.

La formalisation adoptée pour ces règles évite toute confusion possible entre les terminaisons. Cette hiérarchisation de lecture au sein d'une règle est valable tant pour l'humain que pour la machine; ce type de lecture a déjà été utilisé dans certains autres de nos travaux (Cardey et Greenfield 1992).

AVANTAGES D'UN TEL DICTIONNAIRE

Nous avons ici un ensemble de 579 terminaisons accompagnées de sousterminaisons (sous-règles) et/ou de listes d'exceptions.

Le nombre d'entrées peut paraître important mais essayons de comparer notre dictionnaire avec les autres dictionnaires. Prenons la règle 160 qui reconnaît les formes en **er** (El Harouchy 1997 : 250, tab. 8).

Terminaison er et ses exceptions	Opérateur	Dictionnaire Automatique	Robert
er	Verbe inf.	0	4932
ier	Nom	0	1025
fier	Verbe inf.	0	125
Liste	Nom	3	3
Liste	Adj.	2	2
Liste	Nom, Adj.	1	1
gier	Verbe inf.	0	5
Liste _	Nom, Adj.	1	1
Liste	Adj.	42	42
Liste	Nom, Adj.	144	144
Liste	Verbe inf.	127	127
Liste	Nom, Verbe inf.	5	5
Liste	Adv.	3	3
Liste	Nom	184	184
Liste	Adj.	9	9
Liste	Nom, Verbe inf.	41	41
Liste	Nom, Adj.	13	13
Liste	Verbe inf., Adj.	1	1
Liste	Adv.	1	1
Liste	Nom, Adj., Adv.	1	1
Liste	Nom, Adj., Verbe inf.	1	1
Total		579	6666

Cette comparaison s'effectue d'une manière quelque peu spéciale puisqu'une terminaison traite tous les mots qui, en contexte, ont cette terminaison. Nous ne comptabilisons donc pas ces mots qui sont représentés par une unique terminaison et attribuons le chiffre zéro (nous aurions pu mettre le chiffre 1 puisque la terminaison représente une entrée) à cette terminaison. Les seuls chiffres supérieurs à zéro pour le dictionnaire automatique concernent les listes d'exceptions qui sont donc des entrées à part entière. Pour le dictionnaire automatique, un chiffre proche de zéro indique que la terminaison est bien choisie. Il en va autrement pour les autres dictionnaires où plus le chiffre est élevé, plus le nombre d'entrées du dictionnaire est important; l'exhaustivité dépendant de ce nombre d'entrées.

Le dictionnaire Robert sur support électronique recense 6666 entrées pour les mots en **er**, le dictionnaire automatique comptabilise un total de 579 «entrées». Ne parlionsnous pas au départ de gain de temps et d'espace mémoire pour le traitement ?

Un autre avantage de ce dictionnaire est qu'il permet de lister les ambiguïtés morphologiques du français.

Voyons l'exemple suivant	
6/ Mot se terminant par ace	Nom

6.1/ Consulter la liste	Nom, Verbe
conj.	
agace; dédicace; espace; glace; grimace; menace; place; pr	éface; surface; trace
6.2/ Consulter la liste	Nom, Adj.
biplace; boniface; contumace; dace; monoplace; rapace	
6.3/ Consulter la liste	Adj.
coriace; efficace; fugace; inefficace; loquace; perspicac vorace	e; sagace; salace; tenace;
6.4/ Consulter la listevivace	Adj., Adv.
6.5/ Consulter la liste	Verbe conj.
déglace; délace; déplace; désenlace; efface; enlace; ereplace; retrace; surglace; verglace	ntrelace; lace; remplace;

Cette règle générale (règle 6) comprend des exceptions. Certaines des sous-règles ont pour opérateur plusieurs catégories grammaticales. Les sous-règles 6.1, 6.2 et 6.4 vont concerner tous les mots en ace qui sont des mots polycatégoriels. Ainsi, nous obtenons tous les mots polycatégoriels de notre langue.

Ce dictionnaire permet donc de définir les catégories grammaticales auxquelles une forme ambiguë est susceptible d'appartenir. Ainsi, nous avons établi des règles et de ce fait des constantes. Ce qui fait de ce dictionnaire un outil.

Le dictionnaire permet une reconnaissance des mots non ambigus et des mots polycatégoriels. Il permet de mettre en relief 69 types d'ambiguïtés morphologiques, parmi lesquelles à titre d'exemple, nous trouvons :

```
Nom - Participe passé - Verbe Conjugué (reçus par exemple)
Nom - Adverbe - Verbe conjugué (soit)
Nom - Adverbe - Adjectif - Verbe conjugué (ferme, double, trouble)
Nom - Adjectif - Verbe conjugué - Interjection (fixe)
```

Pour terminer, nous donnerons l'exemple d'une phrase traitée par ce dictionnaire

```
la petite joue au ballon
la sous-règle 2.12 (Art., Nom, Pro. Pers)
petite sous-règle 281.4 (Nom, Adj.)
joue sous-règle 398.1 (Nom, Verbe conj.)
au sous-règle 475.1.5 (Art.)
ballon règle 368 (Nom)
```

CONCLUSION

Nous ne pouvons pas dire ici si le hasard fait parfois bien les choses ou s'il y a une logique sous-jacente dans la formation des mots et plus particulièrement dans le cas de nos terminaisons qui ont pu être mises en correspondance avec les catégories grammaticales. Nous pencherions plutôt, dans le cas de notre dictionnaire, vers la première hypothèse sauf pour les cas bien connus des linguistes. Mais une chose assez troublante est que nous avons réalisé, au sein de notre équipe de recherche, le même type de dictionnaire capable

également de déterminer par la terminaison des noms de la langue française si un nom est féminin ou masculin. Il existait quelques règles (les noms terminés par le suffixe -ier, -age, -illon,... sont masculins et ceux terminés par le suffixe -ade, -aie, -ée, sont en principe féminins) mais peu nombreuses et incomplètes; aussi sur le même principe que celui qui a fait l'objet de cet article, un autre dictionnaire a été créé qui définit le genre de tous les noms de la langue française.

RÉFÉRENCES

- CARDEY, Sylviane (1987): Traitement algorithmique de la grammaire normative du français pour une utilisation automatique et didactique, Thèse d'État de Lettres, Besançon, 745 p.
- CARDEY, Sylviane et Peter GREENFIELD (1992): «A Platform for Teaching Natural Language Morphology», *The Ninth International Conference on Technology and Education*, Paris, vol. 3, pp. 1529-1531.
- DUBOIS, Jean et René LAGANE (1995) : *Grammaire*, Livres de Bord, Larousse, coll. «Bordas», Paris, 514 p.
- DUCROT, Oswald et Jean-Marie SCHAEFFER (1995): Nouveau dictionnaire encyclopédique des sciences du langage, Paris, Seuil, 668 p.
- EL HAROUCHY, Zahra (1992): Traitement automatique du genre en français, DEA Sciences du langage, Besançon, 173 p.
- EL HAROUCHY, Zahra (1997): Dictionnaire et Grammaire pour le traitement automatique des ambiguïtés morphologiques des mots simples en français, Thèse de Doctorat en Sciences du langage, Linguistique et Informatique, Université de Franche-Comté, 486 p.
- GREVISSE, Maurice (1988): Le bon usage, 12e éd., Paris, Duculot, 1768 p.

UN CAS DE MÉTAPHORE : CRÉATIVITÉ LINGUISTIQUE ET RÔLE EN INNOVATION BIOTECHNOLOGIQUE

Myriam BOUVERET

Universite Paul Valery-Montpellier III, France

1. INTRODUCTION

La néologie en terminologie fait l'objet de règles de formation précises. Or, si l'on observe le fonctionnement sémantique de la dénomination spécialisée, on constate que celle-ci n'est pas qu'une unité de nomenclature, mais qu'elle se comporte comme un signe de langue générale. La créativité lexicale en terminologie est donc loin de se réduire à une néologie selon les règles du système, voire à une néologie de formation gréco-latine. La métaphore est un cas de fonctionnement néologique très riche reposant sur la néologie sémantique et plus seulement sur la néologie de langue.

Nous nous proposons d'analyser la création d'une dénomination spécialisée en biotechnologies, fondée sur une métaphore. Nous montrerons comment une innovation technologique sur laquelle repose la création d'une société est née à partir d'une analogie avec le vocabulaire de la panification. Cette étude repose sur le suivi d'une recherche biotechnologique pendant trois ans, retraçant l'évolution du procédé du laboratoire à l'industrie. Deux hypothèses seront examinées. La première envisage le transfert de vocabulaire non comme un emprunt (néologie dénominative), mais comme une métaphore (néologie sémantique). On montrera qu'il fonctionne en subduction tant au plan du lexique qu'au plan cognitif. La seconde est que la nature du contexte dans le fonctionnement sémantique de cette métaphore est triple : intra-discursive (mot), discursive et extra-linguistique. Outre les connaissances conceptuelles liées à un domaine, les connaissances sont pratiques mais aussi discursives, révélées et activées dans la mise en discours à travers un ensemble de cooccurrents typiques

2. L'ENJEU DES BIOTECHNOLOGIES EN AROMATISATION

Notre objet est d'analyser la naissance d'une dénomination spécialisée, les motivations et le processus de sa création depuis ses débuts. Le laboratoire de biotechnologie au sein duquel nous avons enquêté¹ est spécialisé dans la recherche d'arômes par voie biotechnologique. Depuis 1985, ses activités de recherches fondamentales et appliquées concernent les substances d'arômes. C'est dans ce cadre que s'inscrit la recherche que nous avons étudiée ici sur la molécule beta-ionone à partir de beta-carotène par voie enzymatique. Ces travaux ont donné naissance à une innovation : un réacteur nommé

¹ Anonyme pour des raisons de confidentialité.

pétrin, qui fit l'objet d'un dépôt de brevet d'invention. Ce pétrin doit améliorer la production d'arômes naturels par voie biotechnologique. Deux enjeux majeurs ont été le moteur de cette innovation : le label naturel et la réduction des coûts pour une production en quantité industrielle. Les arômes élaborés aujourd'hui peuvent comporter jusqu'à quatrevingts composants différents. Mais la difficulté pour les arômes de l'industrie agroalimentaire, destinés à être incorporés dans des denrées alimentaires, est d'obtenir un label d'arômes naturels, les plus intéressants sur le marché étant donné l'engouement du consommateur pour le «naturel», mais aussi les plus coûteux en fabrication. À l'heure actuelle, les législations sur les labels aromatiques sont très strictes et visent à réglementer la qualité des matières premières, leur origine et les procédés de fabrication. Selon la législation, cinq types d'arômes sont autorisés : les arômes naturels, les arômes renforcés, les substances identiques aux naturelles, les substances artificielles et les arômes de transformation. Mais dans les catégories de produits comme les beurres, crèmes, margarines, crèmes glacées, sorbets et boissons non alcoolisées (saufs boissons aux arômes naturels et artificiels), seuls les arômes naturels sont désormais autorisés.

Dans ce laboratoire, la recherche sur la molécule beta-ionone commence dès 1985. dans le cadre d'une thèse de doctorat. Cette molécule obtenue par voie naturelle est très recherchée. Elle est présente dans les produits de consommation comme le thé, les pruneaux, les jus d'abricot et de mangue, certains jus de fruits chauffés, les algues séchées et le tabac. Cette substance est caractéristique de l'arôme de framboise. Produire la molécule beta-ionone par voie naturelle à des rendements intéressants pour l'industrie constitue toujours un défi. Afin de produire 300 g d'ionone, il faut 1 kilo de carotène. On peut obtenir le beta-carotène en concentrant du jus de carotte, mais il faudrait utiliser 558 kg de jus de carotte ou l'acheter directement à environ 2 000 francs le kilo. Le coût de production serait trop élevé. Une première étude a donc été conduite, situant la réaction en milieu liquide principalement composé de farine, d'huile, de concentré de jus de carotte, d'autres ingrédients et de beaucoup d'eau. Ce travail dura trois ans jusqu'en 1988. Mais un milieu trop liquide ne permettait pas d'arriver aux résultats escomptés, et la recherche s'est poursuivie pendant une année, au cours de laquelle des études comparatives ont été menées entre un milieu liquide, c'est-à-dire un mélange d'ingrédients donnant une consistance liquide, et un nouveau milieu où le liquide était évacué: un milieu qui prenait la consistance d'une pâte.

Cette recherche d'un nouveau milieu réactionnel a conduit à changer de réacteur, il a fallu modifier le système traditionnel, connu pour un milieu liquide. Dès 1989 le nouveau système fonctionne. Il permet de produire plusieurs autres molécules naturelles qui sont alors commercialisées comme l'hexanal naturel à odeur/goût de pomme verte donnant aux aliments un «goût de frais». On n'arrive toujours pas aux résultats satisfaisants pour la molécule beta-ionone. Néanmoins, le système élaboré, un réacteur désigné sous le nom de pétrin, et le nouveau milieu réactionnel, un milieu pâteux, sont une véritable innovation pour la recherche en aromatisation agro-alimentaire et donneront lieu à un dépôt de brevet en 1993. Dès 1989, le système est pressenti comme étant une innovation technologique sur un marché très porteur, celui des arômes naturels. Et à partir de ce procédé une société de production d'arômes naturels est créée. La société produit avec ce système un ensemble de molécules aromatiques naturelles et poursuit les travaux sur le projet du beta-carotène. En 1995, la recherche sur ce thème est toujours en cours et les dénominations autour du procédé et du milieu évoluent en même temps que la réflexion : le milieu n'est plus complètement pâteux, mais d'aspect mousseux, et le pétrin est désormais dénommé

batteur. Parallèlement, des travaux sont conduits dans le laboratoire de biotechnologie afin d'améliorer les performances de l'un des réactants : l'oxygène.

3. COMPARAISON ENTRE LES CHAMPS LEXICAUX DE L'AROMATISATION ET DE LA PANIFICATION

Les recherches sur la molécule beta-ionone à partir de beta-carotène ont commencé en milieu liquide dans le laboratoire de biotechnologies. Afin d'obtenir la production de beta-ionone, l'ensemble des ingrédients sont mélangés et travaillés à l'intérieur d'un réacteur. Il existe plusieurs sortes de réacteurs, différents selon la nature et la quantité du milieu réactionnel. Une fois le mélange incorporé, la réaction est obtenue par une coupure des chaînes de carbones grâce à une oxydation des lipides, instables, qui réagissent en présence de l'oxygène, c'est la co-oxydation. Les molécules instables qui vont permettre cette coupure sont des radicaux libres. L'oxygène est donc un facteur déterminant, de même que sa circulation. L'autre facteur important est le maintien d'une température constante dans le réacteur. Le milieu est à la base composé de farine, d'huile, d'une lipase : la lipoxygenase, de concentré de jus de carotte : le carotène, et de beaucoup d'eau. Mais d'après le directeur du laboratoire, le milieu était noyé dans trop d'eau, ce qui ne convenait pas. Deux raisons ont donc motivé la concentration du milieu : d'une part la réaction en milieu liquide occasionnait une dégradation à terme de l'un des ingrédients. La plupart des travaux connus sur la lipoxygénase ont en effet été réalisés en milieu très hydraté et l'on a pu démontrer que la réaction en milieu liquide est vingt fois supérieure à celle qui est réalisée en milieu concentré. Mais paradoxalement, ce n'est pas un avantage, car le mélange de l'ensemble des réactants est réalisé avec une mobilité accrue, ce qui entraîne la décomposition de certains éléments. Une réaction en milieu concentré en revanche démarre moins vite mais donne de meilleurs résultats pour obtenir les molécules recherchées. Et d'autre part, la seconde raison est qu'un milieu moins liquide permet d'obtenir une plus grande quantité, ce qui rend le procédé financièrement exploitable à échelle industrielle. Le passage d'un milieu liquide à un milieu concentré s'est donc fait par le recours à un ingrédient particulier : la farine de soja, et par diminution d'un autre : l'eau. Le milieu ainsi produit prenait une consistance toute différente de celui de départ. Il devenait visqueux et prenait l'allure d'une pâte. Et c'est dans ce changement de milieu et le changement de réaction que cela a entraîné que se situe l'originalité du procédé, originalité qui l'est encore à l'heure actuelle où ce type de milieu n'existe pas en aromatisation. On connaît le milieu liquide, le milieu semi-liquide, le milieu solide, mais pas le milieu pâteux. Le cheminement que l'on observe donc est le passage d'un milieu liquide à un milieu plus visqueux. Or ce milieu plus visqueux a l'aspect et la consistance d'une pâte et il se travaille comme une pâte. Cette analogie décrite par les auteurs, dans leur récit oral de l'invention du procédé, se retrouve dans le brevet. Afin de comprendre comment est réalisée le transfert de termes du vocabulaire de la panification à ces travaux, on relèvera pour commencer dans le texte du brevet le vocabulaire utilisé par analogie avec celui de la panification avant d'établir une comparaison avec un article sur la panification.

1- Selon le procédé, on effectue la mise en contact d'une source ou préparation de *lipoxygénases* et d'hydroperoxyde-lyase, d'une source naturelle d'acides gras polyinsaturés et d'une source naturelle de carotène, en milieu concentré à *viscosité* élevée, par *malaxage*, une alimentation en oxygène du mélange réactionnel étant effectuée au cours du *malaxage* dans un réacteur où règne une pression au moins égale à la pression atmosphérique.

- 2- Le procédé de préparation de substances d'arômes par voie d'oxydation enzymatique est essentiellement original en ce que la teneur en eau est relativement faible, de sorte que la réaction de biosynthèse s'effectue en *milieu à viscosité élevée*, donc concentré en composants bioréactionnels. De ce fait, tout facteur qui augmentera le contact enzyme-substrats accélérera le développement des réactions enzymatiques. On favorisera donc les réactions enzymatiques en assurant un *pétrissage* intensifié des divers composants.
- 3- La réaction s'effectue dans un *malaxeur*, en *milieu concentré à viscosité élevée*, l'alimentation en oxygène du mélange réactionnel étant effectuée au cours du *malaxage* par un apport d'air ou d'oxygène ou d'un mélange air/oxygène dans un *malaxeur-réacteur* où règne une pression au moins égale à la pression atmosphérique.
- 4- La réaction est conduite en *milieu concentré*, à viscosité élevée, sous agitation, en présence d'air et/ou d'oxygène, sous pression [...] Si le débit est trop important il en résultera des pertes en arômes élevées par entraînement. On injectera par exemple de 1 à 2 l/m. d'air et/ou d'oxygène pour 2 kg de *mélange pâteux*. La pression d'air, d'air et/ou d'oxygène au sein du *réacteur* (*mélangeur-malaxeur*) pourra aller jusqu'à 8 bars.
- 5- En ce qui concerne la *viscosité* du milieu *pâteux*, elle a été caractérisée à l'aide de deux types de mesures.
- 6- La vitesse d'agitation du *mélange pâteux* sera généralement comprise entre 60 et 200 tours.min.
- 7- Sur cette figure, on a représenté de façon schématique un réacteur au sein duquel s'effectuent de façon avantageuse les opérations d'homogénéisation et de *malaxage* du *mélange pâteux*, d'oxygénation du *mélange visqueux*, ainsi que la bioréaction qui donne lieu à la formation des arômes par voie d'oxydation enzymatique.
- 8- Un outil tournant 6, assurant les fonctions de *mélangeage* et de *malaxage*, par mouvement planétaire, est entraîné en rotation par un axe secondaire tournant 7. Cet outil tournant 6, qui constitue le dispositif mécanique d'agitation du *milieu visqueux* 8, en assure simultanément l'homogénéisation et l'oxygénation au sein de l'enceinte étanche 1.
- 9- En atmosphère libre, qui surmonte le *milieu visqueux*, composé d'air, d'un mélange air/oxygène ou d'oxygène pur, sous pression de l'ordre de 1 à 8 bars, est au contact du *milieu visqueux* 8 brassé par l'outil tournant 6.
- 10- Les effets conjugués de la pression et du renouvellement de l'atmosphère ont pour conséquence de parfaire les conditions d'oxygénation du *milieu visqueux*.
- 11- On a obtenu un $milieu\ p \hat{a}teux$, qui a été $malax \hat{e}$ à raison de 90 t.min -1, à 25°C.
- 12- Les meilleures concentrations en ionones, obtenues après une durée d'essai de 26 h, sont de 95mg.kg-1 de *pâte* pour l'alpha-ionone et de 300 mg.kg-1 de *pâte* pour la beta-ionone.
 - 13- La réaction s'effectue dans un malaxeur, en milieu concentré à viscosité élevée.

14- Procédé de préparation [...] caractérisé en outre par le fait que le milieu bioréactionnel a une *viscosité* telle que la force d'adhérence du milieu pâteux peut être comprise entre 5N et 8N.

Les contextes 1 à 4 décrivent un milieu à viscosité élevée ou un milieu concentré à viscosité élevée et le mélange des composants est nommé mélange pâteux. Le terme pétrissage est alors employé : «On favorisera donc les réactions enzymatiques en assurant un pétrissage intensifié des divers composants». L'analogie avec le vocabulaire de la panification est très forte. Progressivement, on arrive à la dénomination de milieu pâteux (5), tout d'abord en cooccurrence avec viscosité : la viscosité du milieu pâteux et à celle de mélange pâteux (4 et 6). Puis, tous les exemples de réaction sont décrits avec le syntagme milieu pâteux comme en (11) : «on a obtenu un milieu pâteux». Le terme pâte est ensuite utilisé comme en (12) pour dénommer le produit obtenu dans le réacteur dans chacun des onze essais de réaction. Ainsi, de la page 8 à 16 du brevet décrivant les onze exemples de réaction, c'est-à-dire dans la plus grande partie du brevet, les seuls termes utilisés sont milieu pâteux, malaxé, pâte, mélangeur-malaxeur. Ni visqueux, ni concentré ne sont employés alors que dans le début du brevet, le syntagme milieu concentré à viscosité élevé avait été actualisé plusieurs fois, comme synonyme de pâteux, évité. Les onze essais de réactions préparées avec des milieux de proportions différentes dans le réacteur-malaxeur sont tous décrits de la manière suivante : «On a obtenu un milieu pâteux». Ainsi, dans un moment du récit de brevet qui décrit les résultats d'expérience, c'est le terme pâteux qui s'affirme comme seul valable. Les quatre dernières pages du brevet, intitulées «Revendications» et établissant les conclusions, posent comme synonymes les deux dénominations : milieu concentré à viscosité élevée, syntagme définitionnel et milieu pâteux. À l'écrit, les termes utilisés par analogie avec le vocabulaire de la panification sont donc les suivants : viscosité, visqueux, pétrissage, pâteux, malaxage, malaxé.

Le terme *pétrin* n'est pas actualisé à l'écrit, mais il est utilisé à l'oral dans les entretiens :

«Pour notre procédé on a été cherché des termes dans un autre domaine qui était celui un peu de la boulangerie, panification. [...] Notre réacteur, c'est un *pétrin* maintenant pour faire des réactions enzymatiques» (Entretien avec le directeur, 27 janvier 1994).

Dans le brevet, ce n'est pas pétrin qui est employé mais un terme plus générique malaxeur, dérivé directement de l'opération de malaxage décrite à l'aide du verbe malaxer. Dans le brevet, c'est donc le paradigme pâte, pâteux, pétrir, malaxer, malaxeur, plutôt que pâte, pâteux, pétrir, pétrin qui est actualisé. Pour décrire le procédé, de nombreux synonymes sont employés : mélangeur-malaxeur ou malaxeur, réacteur ou réacteur-malaxeur. Le terme mélangeur-malaxeur est un composé dénommant les fonctions du pétrin : il mélange et malaxe. Réacteur est le terme hyperonyme, dénomination d'appareil plus générale et réacteur-malaxeur décrit la spécificité d'un réacteur, c'est un hyponyme. Le néologisme sémantique pétrin est donc évité au profit d'une néologie de langue créée à partir d'une dérivation morphologique : malaxeur, sur le modèle de réacteur, mélangeur.

Quelques extraits d'un article sur l'histoire du pain (Godon 1995) nous permettront d'établir la comparaison entre le champ lexical de départ de la métaphore (A) : la panification et celui du milieu d'accueil (B) : l'aromatisation.

- 1- On préparait ensuite une *pâte* peu consistante par *malaxage* dans de grandes jarres en terre et on la cuisait dans des moules en terre.
- 2- Plusieurs opérations se succèdent pendant 30 minutes environ, jusqu'à ce que la *pâte* n'*adhère* plus aux mains de l'ouvrier : délayage du levain, *mélange* et brassage avec la farine, parfois incorporation du sel, découpage de la *pâte*, pliage et battage jusqu'à obtention d'une *pâte* bien liée et aérée.
- 3- L'eau est indispensable pour préparer la *pâte*. Elle mouille la farine, donne la cohésion à la pâte et fournit une mobilité nécessaire aux constituants de la farine pour la réalisation des réactions chimiques.
- 4- L'eau joue un rôle primordial dans la formation de la pâte. Au cours du pétrissage, par la multiplication des contacts, les molécules d'eau se lient aux divers groupes électriquement chargés des constituants, puis, par son intermédiaire, des liaisons se forment entre les constituants eux-mêmes. La présence d'eau favorise aussi l'établissement, entre les lipides et les protéines, d'interactions hydrophobes dues aux forces de répulsion des chaînes carbonées vis-à-vis des molécules d'eau.
- 5- La fabrication du pain dure un peu plus de cinq heures. Pendant ce temps, le boulanger agit sur la *pâte* à certains moments bien précis; entre ces opérations, il laisse reposer ou fermenter la *pâte*, selon le stade de la fabrication. On dit qu'il laisse la *pâte* «travailler».
- 6- La première opération de la panification est le *pétrissage* : celui-ci permet de mélanger les divers constituants, puis de leur donner, par l'établissement de liaisons chimiques, la structure indispensable pour conférer au pain son aspect. La *pâte* devient *élastique* par formation d'un réseau, le gluten, constitué par des protéines de la farine. L'air indispensable à l'action de la levure est emprisonné par l'action mécanique du *pétrin*.
- 7- Ce dernier se forme au cours du *pétrissage* par l'établissement de diverses liaisons entre les gliadines et les gluténines, pour donner naissance à la masse *viscoélastique* que l'on obtient quand on passe la *pâte* sous un filet d'*eau*. Les gliadines constituées d'unités de masse moléculaires de 40 000 donnent, seules, une masse extensible, molle et de faible *élasticité* tandis que les gluténines forment un réseau tenace et peu extensible.
- 8- L'ensemble des *lipides* ne représentent que deux pour cent environ de la farine [...]. L'ensemble des lipides peut être oxydé par la *lipoxygénase*, l'enzyme qui libère des hydropéroxydes favorables aux propriétés mécaniques des *pâtes*, à partir des acides gras insaturés.

On retrouve ici des termes semblables à ceux du brevet d'aromatisation : pâte, malaxage, adhère, mélange, viscosité, pétrissage. Cette association de termes est équivalente à celle du milieu d'aromatisation. On pourrait conclure que ces termes appartiennent et constituent l'isotopie de la panification. Mais en fait, ce sont seulement ceux qui sont en commun avec l'ensemble identifié pour l'aromatisation. Ce que l'on fait apparaître ici, c'est donc uniquement ce qu'il y a de commun entre le milieu de panification et le milieu d'aromatisation. Il semblerait alors que cet ensemble cohérent forme les cooccurrents typiques du pétrissage, plus spécifiquement que ceux de la panification.

4. PRODUCTION DE LA MÉTAPHORE

Le premier niveau de réalisation de cette métaphore est donc le mot. Si tous les autres termes peuvent décrire une activité différente, celui de *pétrissage* est bien spécifique de la panification et est lié à l'appareil : le *pétrin*. La métaphore est réalisée dans le discours du brevet par ce terme clé. *Pétrissage* est en effet le terme pivot, «focus» (cf. Black 1962) ou «foyer» (cf. Le Guern 1973) de la métaphore qui permet le passage du champ lexical A à celui du champ lexical B. L'unité lexicale est donc le premier niveau de réalisation de cette métaphore et le passage du milieu A au milieu B se fait grâce à un partage de caractéristiques communes. Mais il ne s'agit pas d'une simple substitution d'un mot à un autre. On peut aller plus loin et constater qu'il s'agit d'un véritable ensemble de termes en commun. La métaphore ne se situe pas seulement dans le cadre du mot, ni dans celui de la phrase. Elle est réalisée dans le cadre du discours et c'est cette entité qui permet d'identifier le processus métaphorique.

Le deuxième niveau de réalisation de cette métaphore est donc le discours. Celui-ci tout entier est porteur de son développement. Elle déborde largement le cadre du mot et l'analogie est sous-jacente à tout le discours du brevet. Les termes tels qu'ils sont employés en cooccurrence dans le texte : visqueux, viscosité, pâteux, pâte, malaxer, malaxage, malaxeur sont issus du vocabulaire de la panification. Mais la viscosité dans un autre contexte peut être appliquée à un tout autre domaine de l'agro-alimentaire : au chocolat par exemple. Le terme viscosité associé à une pâte peut s'appliquer aux fromages : on parle de la viscoélasticité des pâtes fromagères. En revanche, la consistance trop liquide du chocolat empêche de le malaxer, on le mélange. De même que la consistance trop solide du fromage empêche de le mélanger ou de le malaxer. Cette pâte spécifique a donc toutes les caractéristiques d'une pâte à pain. Et c'est la cohérence du discours à travers l'association de cooccurrents typiques de la pâte à pain qui crée cette isotopie de la panification.

Le troisième niveau de réalisation de la métaphore est extra-linguistique. Elle est déterminée par un partage de savoirs communs, c'est un savoir expérientiel. Au delà du discours, on retrouve aussi des ingrédients majeurs semblables entre les deux milieux : farine, eau et lipides. La particularité se situe donc dans une association d'ingrédients : farine, huile, poudre (carotène) et eau, qui ont la consistance d'une pâte. Et plus encore, certaines applications sont identiques : la présence d'une quantité d'eau justement dosée pour une meilleure circulation des réactants; la farine assure la cohésion; l'air est indispensable dans la réaction. Ces trois éléments sont ceux qui sont déterminants et novateurs dans le procédé d'aromatisation : une réaction aérobie sur un milieu constitué de farine, de carotène, d'huile et de très peu d'eau. Par ailleurs, nous avons vu plus haut que pour les expériences conduites sur le milieu liquide en aromatisation, on avait pu constater que la rapidité de la réaction en début de phase donnait une concentration moindre de la molécule par rapport à un milieu pâteux où le mélange des réactants était plus lent. Ici, une réaction analogue d'amoindrissement du goût et de l'odeur (odorante, goût) a été constatée lors d'une vitesse de pétrissage augmentée. Un autre point commun existe : dans les deux milieux pâteux, de panification et d'aromatisation, on constate la présence de l'enzyme lypoxygénase. Et le constat établi pour le milieu de panification renforce le parallèle avec la réaction du milieu d'aromatisation :

Si l'origine des substances aromatiques du pain est encore imparfaitement connue, il est cependant certain que les enzymes jouent un rôle essentiel dans la formation de divers

produits précurseurs des arômes au cours du pétrissage et de la fermentation : leur excès modifie en effet beaucoup les qualités organoleptiques du pain (Godon 1995 : 39).

Il semblerait donc qu'à travers l'analogie avec la panification, il ne s'agisse pas que d'une question d'emprunt de vocabulaire : celui-ci est la trace linguistique d'une analogie qui tend à se situer bien au-delà. En l'occurrence, l'analogie est aussi celle de la réaction chimique, des ingrédients. Si l'on a pu constater que la métaphore débordait le cadre du mot pour s'étendre au cadre d'un contexte discursif, on doit admettre ici, que le contexte de la métaphore est aussi d'ordre pratique. La métaphore se réalise donc à trois niveaux : le mot, le discours et la pratique. Ce sont les trois niveaux de contexte que l'on peut identifier dans le fonctionnement sémantique de ce processus néologique.

L'analogie établie ici a un rôle de support de réflexion. On peut donc parler aussi de métaphore cognitive. Ainsi, la métaphore du *pétrin* repose sur un processus d'analogie scientifique et celle-ci a un rôle de support de réflexion dans la recherche. La métaphore est utilisée comme un moteur d'innovation. Et en effet, c'est ce qui apparaît clairement dans un chapitre de l'étude de 1995 dans le cadre d'une recherche du laboratoire de biotechnologies visant à améliorer la circulation d'oxygène dans le procédé.

Cette structure a été établie à partir de la structure biochimique de la *pâte à pain*. Une extrapolation nous a semblé possible dans la mesure où les deux milieux ont une composition très semblable : nous y trouvons lipides, protéines, eau (Document interne, 1995).

Non seulement l'analogie apparaît dans le terme *pâte à pain*, mais c'est la réflexion en termes de panification qui a permis d'établir la structure du milieu réactionnel. L'article cité en bibliographie du document confirme d'ailleurs un mécanisme de recherche par analogie avec le milieu réactionnel de la pâte à pain en panification : «Exemple du rôle de la lipoxygénase en panification [...]» (Nicolas et Potus, 1994, *Science des aliments*, pp. 627-642).

Nous avons donc affaire ici à une métaphore qui se réalise dans le cadre du mot et qui s'étend au discours et plus largement aux pratiques, à l'expérience scientifique. On a pu localiser un foyer de la métaphore : le *pétrissage*, établissant le passage discursif à la panification. Mais, au-delà de ce terme pivot, c'est d'une part la cohérence du discours avec l'association des termes *pâte*, *malaxage*, *viscosité*, *mélanger*, identifiés comme cooccurrents typiques du *pétrissage*, et d'autre part l'analogie des ingrédients et des réactions/applications qui sont le support de cette métaphore. Ainsi, ce procédé d'analogie s'apparente à celui d'une métaphore. Il en a toutes les caractéristiques de partage de deux univers permettant le glissement, mais il en a aussi toutes les caractéristiques de dissociation. Comme pour le processus métaphorique, l'analogie n'est pas totale, mais c'est le partage de certains points communs² entre la panification et l'aromatisation qui permet l'activation de celle-ci.

² Identifiés comme «typiques» par Black (1962) et Lakoff (1980) ou «inhérents» par Le Guern (1973) et Rastier (1987).

CONCLUSION

Du point de vue linguistique, concernant la diffusion, l'impact de cette néologie est pour l'instant moindre. En revanche, sur le plan cognitif, le processus néologique de la dénomination métaphorique constitue un moteur d'innovation très fort. On ne peut pas se contenter d'analyser cette néologie comme un emprunt de vocabulaire avec simple déplacement de sens. La métaphore a entraîné un transfert de termes et a motivé un transfert de technologie. Dans la métaphore du pétrin, on peut dire que bien qu'il s'agisse du même domaine, l'agro-alimentaire, les deux activités de panification et d'aromatisation sont bien distinctes. De plus le référent est différent dans les deux cas. En aromatisation, le pétrin est fermé, possède une double enveloppe et maintient une température constante, caractéristiques qu'il ne possède pas en panification. Comme l'affirme M. Slodzian, le transfert de connaissances n'est pas qu'un simple transfert de vocabulaire :

«L'analyse des seuls procédés de dénomination ne permet pas d'accéder aux modèles de représentation des concepts. Sans un travail de reconstitution du cadre discursif et sémiotique global, il est difficile d'accéder au chaînage des concepts qui structurent un micro-domaine de savoir. Cette approche de la fabrique linguistique des concepts présuppose un recentrage de l'objectif scientifique, le transfert des connaissances ne pouvant être confondu avec un simple transfert de vocabulaires» (Slodzian 1995 : 243).

Parler de deux acceptions différentes du terme *pâte* ou de l'appareil *pétrin* en milieu A et B ne suffit pas pour expliquer le recours au vocabulaire de la panification. Selon nous, la métaphore dégagée ici n'est ni un emprunt, ni une néologie dénominative, mais elle est une néologie sémantique à part entière. Cette néologie s'apparente à un processus de polysémie avec mécanisme de subduction (cf. Picoche 1984). Elle est réalisée tant au niveau lexical que discursif et pratique, le contexte d'activation de la métaphore se situe donc à trois niveaux, autant linguistiques qu'extra-linguistiques. La métaphore est bien une néologie sémantique. Même si elle n'a pas une importance quantitative dans le renouvellement de vocabulaires spécialisés, son rôle est fondamental dans la construction et l'avancée des connaissances. Elle déplace les champs de recherche et les champs lexicaux. Dans cette perspective, l'étude de la polysémie peut permettre de dégager des pistes concernant le fonctionnement sémantique des langues spécialisées.

RÉFÉRENCES

- BLACK, M. (1962): Models and metaphors. Studies in language and philosophy, Cornell University Press, 267 p.
- BOULANGER, J.-C. (1984): «Quelques observations sur l'innovation lexicale spontanée et l'innovation lexicale planifiée», La banque des mots, n° 27, Paris, CILF, pp. 3-29.
- BOUVERET, M. (1996): Néologie et terminologie: production de sens du terme, Thèse de doctorat, Université. Paul Valéry-Montpellier III.
- BRES, J. (1987): «Une enquête sociolinguistique en milieu minier: corpus et problèmes méthodologiques», Lengas, n° 16, éd. Université Paul Valéry-Montpellier, pp. 74-98.
- CORBIN, D. (1987): Morphologie dérivationnelle et structuration du lexique, 2 vol., Tübingen, Niemeyer, 937 p.

M. Bouveret

- DARMESTETER, A. (1979), (1° éd. 1887, Delagrave): La vie des mots étudiés dans leurs significations, Champs libres.
- FALL, K., LEARD, J.-M. et P. SIBLOT (dir) (1996): Polysémie et construction du sens, Praxiling, Universite Paul-Valery-Montpellier, 193 p.
- FERNEZ-WALCH, S. (1991): L'innovation de produit au quotidien en entreprise industrielle, Thèse de doctorat en Économie industrielle, École des Mines de Paris.
- GARDIN, B. (1974): «La néologie, aspects socio-linguistiques», in Guilbert (dir), pp. 67-73.
- GAUDIN, F. (1991): «Terminologie et travail scientifique: mouvement des signes, mouvement des connaissances», in Gaudin et Assal (dir), pp. 111-132.
- GAUDIN, F. et A. ASSAL (dir) (1991): «Terminologie et sociolinguistique», Cahiers de linguistique sociale, n° 18, CNRS-Université de Rouen, 213 p.
- GAUDIN, F. et M. BOUVERET (sous presse): «Partage des noms, partage des notions? Approche sociolinguistique de difficultés terminologiques en situation interdisciplinaire», Thoiron Ph. (dir), La dénomination.
- GODON, B. (1995): «Le pain», La science, mars, Dossier hors série, pp. 38-47.
- GUESPIN, L. (1974): «Néologie et énonciation: problèmes théoriques et méthodologiques», in Guilbert (dir), pp. 74-82.
- GUILBERT, L. (dir.) (1974): «La néologie lexicale», Langages, n° 36, Paris, Larousse, 128 p.
- GUILBERT, L. (1975): La créativité lexicale, Paris, Larousse, 285 p.
- KITTAY, E. et A. LEHRER (1981): «Semantic fields and the structure of metaphor», *Studies in language*, 5-1, pp. 31-63.
- KLEIBER, G. (1990): La sémantique du prototype. Catégories et sens lexical, Paris, PUF, coll. «Linguistique nouvelle», 199 p.
- LAFONT, R. (1978): Le travail et la langue, Flammarion, 301 p.
- LAKOFF, G. et M. JOHNSON (1980): *Metaphors we live by*, The University of Chicago Press, 241 p.
- LATOUR, B. et S. WOOLGAR (1988): La vie de laboratoire, la production des faits scientifiques, La découverte, 299 p. (traduit de l'anglais, Laboratory life, the construction of scientific facts, Sage publications Inc, 1979).
- LECOINTRE, S. et D. LEEMAN (dir) (1994): «Le lexique: construire l'interprétation», *Langue française*, n° 103, Paris, Larousse, 125 p.
- LE GUERN, M. (1973): Sémantique de la métaphore et de la métonymie, Larousse Université, 125 p.

- LERAT, P. (1995): Les langues spécialisées, Paris, PUF, coll. «Linguistique nouvelle», 198 p.
- LERAT, P. (1996): «Terminologie et connaissances sémantiques», Actes des Journées de sémantiques lexicales brestoises, à paraître.
- MOLINO, J. (1979): «Métaphores, modèles et analogies dans les sciences», *Langages*, n° 54, Paris, Didier Larousse, pp. 83-101.
- PICOCHE, J. (1984): «L'utilisation des notions de signifié de puissance et de subduction en lexicologie», *Cahiers de lexicologie*, vol. 1, n° 44, pp. 41-49.
- PICOCHE, J. (1986): Structures sémantiques du lexique français, Nathan, 144 p.
- RASTIER, F. (1987): Sémantique interprétative, Paris, PUF.
- RASTIER, F. (1996): «Le défigement des expressions figées et leur interprétation», in Fall et alii (dir), pp. 17-24.
- SIBLOT, P. (1993): «De la prototypicalité lexicale à la stéréotypie discursive: la casbah des textes français», in *Lieux communs; topoï, stéréotypes, clichés*, Kimé, Plantin C.(éd), pp. 342-354.
- SIBLOT, P. (1995): «Comme son nom l'indique....» Nomination et production de sens, Thèse d'État, Université Paul Valéry-Montpellier III.
- SLODZIAN, M. (1995): «Transfert de connaissances nouvelles et aménagement terminologique», *Meta*, Montréal, Presses de l'Université de Montréal, pp. 239-243.

LES NOMENCLATURES ONT-ELLES UNE MÉMOIRE... MÉTAPHORIQUE ? LES SUBSTRATS DE LA NOMINATION

Christine PORTELANCE

Université du Québec à Rimouski, Canada

Ce titre annonce une double interrogation concernant, d'une part, l'existence d'un phénomène sémantique diachronique rattaché à un «ensemble de termes employés dans une science ou une technique», phénomène apparenté à une mémoire et, d'autre part, la nature d'une telle mémoire, c'est-à-dire les propriétés qui la définissent. En d'autres termes, nous cherchons à savoir si les nomenclatures conservent de l'expérience passée une sorte de dépôt pouvant agir sur nos représentations linguistiques à venir et si cette reviviscence est de nature métaphorique. La question de la nomination a été abondamment traitée, surtout par le biais d'études sur la référence¹, cette dernière toutefois ne fait pas l'enjeu de notre démarche, comme nous le verrons plus loin. La nomination est ici traitée comme un jeu de langage², comme une interaction entre «expérience-du-monde» et «expérience-du-langage», et c'est donc en tant que phénomène d'usage plutôt qu'en fonction de sa valeur ontologique qu'elle sera examinée.

L'étude de Louis Guilbert sur la formation du vocabulaire de l'aviation nous sert de champ d'observation pour différentes raisons : premièrement, le domaine est plus technique que scientifique, il annonce ainsi la technoscience; deuxièmement, la période étudiée précède la réalisation d'un vol véritable, on peut donc déceler une convergence entre le développement conceptuel, la formation terminologique et la réalisation technique. Nous nous intéressons principalement aux transferts terminologiques d'une sphère d'expérience à une autre et, à cet égard, l'étude de Louis Guilbert s'inscrit dans la foulée de celle de Wexler³, qui déjà soulignait l'importance de ces transferts terminologiques dans la formation d'un vocabulaire. Or c'est précisément le mouvement même de la migration terminologique qui nous semble porteur d'un nouvel éclairage sur la nomination et qui, par conséquent, constitue l'objet réel de notre réflexion.

¹ On pense ici aux travaux de G. Frege, de B. Russell, de W.O. Quine et de P.F. Strawson. Pour un résumé de ces travaux, cf. Hottois (1989).

² Le point du vue est wittgensteinien même si Wittgenstein dans les *Investigations* conteste la place centrale qu'occupe la nomination dans les études sur le langage. Sa démarche s'attache à identifier et à décrire tous les autres jeux de langage. Nous intégrons la réflexion sur les autres jeux pour revenir à la nomination et tenter ainsi de montrer qu'il y a plus qu'une simple opération de désignation de nature indicielle.

³ Peter J. Wexler, La Formation du vocabulaire des chemins de fer en France (1778-1842) publiée en 1955.

VERUM IPSUM FACTUM

«Le vrai est le même que le fait⁴». On ne connaît que ce qu'on a fait. Cette maxime du XVIII^e siècle, attribuée à Giambattista Vico⁵, justifie à elle seule le choix d'examiner un produit de la technologie plutôt qu'une théorie scientifique. En choisissant d'étudier une réalisation concrète, nous pouvons alors sans remords éviter de questionner l'antériorité de la réalité sur l'observation, une perspective qui, de surcroît, a été fortement ébranlée depuis les développements de la physique quantique⁶ dans la première moitié du siècle. Évacuation élégante, pourrait-on dire, de la question de la référence et de la valeur de vérité. Mais un tel choix n'a pas pour seul mérite de se soustraire à la querelle subjectivisme *contra* objectivisme, car il correspond à une posture qui s'inscrit en faux contre une lecture manichéenne du monde. À l'instar de Putnam (1984 : 9) pour qui «l'esprit et le monde construisent conjointement l'esprit et le monde», nous avons la conviction que la réalité consiste en un va-et-vient entre le sujet et l'objet et que cette réflexivité est incontournable⁷.

Qui plus est, la terminologie étudiée par Guilbert s'applique alors à un objet encore virtuel. Cette situation, un obstacle dans le cas d'une étude synchronique simultanée, sert plutôt une démarche comme la nôtre puisque non seulement l'objet n'est pas une chimère, on le sait aujourd'hui, mais la concomitance des constructions langagière et aéronautique nous permet un regard sur ce que le sociologue Bruno Latour (1995) appelle la «science (ou la technoscience) en action». Dès lors, il devient possible de conjuguer l'étude du «fait» à celle du «faire»⁸.

LE TRAITÉ DE FORMATION DU VOCABULAIRE DE L'AVIATION DE LOUIS GUILBERT (1965)⁹

Les documents dépouillés par Louis Guilbert, plus de trois cents, proviennent de sources diverses : brevets, correspondances, revues, journaux ainsi que différents ouvrages dont certains de fiction, notamment ceux de Jules Vernes. Les énoncés d'où les unités terminologiques ont été tirées sont réunis dans un recueil et forment le deuxième tome de l'ouvrage. La période étudiée s'étend sur trois décennies, de 1861 à 1891, Guilbert justifie ce choix :

«Cette délimitation repose sur des motifs d'ordre linguistique et d'ordre historique. 1861, c'est la date d'apparition du mot *hélicoptère*, 1891 suit l'année du brevet d'Ader où est apparu pour la première fois le mot *avion*. Sur le plan du mouvement de la société, la période choisie offre un contexte historique, social et économique caractérisé en premier lieu par un puissant essor des techniques et par leur application à la grande industrie, désormais en plein épanouissement.» (9)

⁴ De factum, facere «faire».

⁵ Citée et commentée par Glasersfeld dans Watzlawick (1988 : 30).

⁶ Cf. Heisenberg, W. (1971) Physique et philosophie, Albin Michel.

⁷ Voir à ce propos Portelance (1997).

⁸ Nous utilisons les termes les plus neutres possibles (construit et construction, fabriqué et fabrication sont par ailleurs trop connotés) dans le sens où, en anglais, les termes made et making auraient été tout à fait appropriés.

⁹ Cet ouvrage étant le seul de L. Guilbert auquel nous ferons référence, seuls les numéros de pages figureront entre parenthèses.

Il n'est pas facile de faire la synthèse des analyses de Guilbert — un grand nombre d'observations sur un grand nombre de cas —, on examinera les hypothèses posées au départ, les grandes lignes de ses observations¹⁰ et les conclusions qu'il en tire.

Comme point de départ, Louis Guilbert choisit de s'appuyer sur les propos d'Antoine Meillet et de Ferdinand Brunot. Du premier, il retient que l'influence des causes sociales sur les faits linguistiques se fait particulièrement sentir dans «l'innovation apportée au sens des mots»; du second, qu'il n'existe pas de concordance régulière et constante entre l'évolution parallèle du lexique et des «réalités de la vie et de la pensée» (8). Il pose ensuite la question à laquelle toute sa thèse s'efforcera de répondre :

«Quand la causalité historique suscite une création lexicale, celle-ci vient-elle simplement s'intégrer dans une collection de termes ou provoque-t-elle une redistribution des éléments de signification d'un ensemble lié par un réseau de relation ? [...] En bref, peut-on considérer les notions de diachronie et de structure conciliables ?» (8-9)

C'est bien la question de la structuration du lexique, posée à partir du concept saussurien de valeur, qui sert de cadre à la thèse. Au premier chapitre, il tente d'englober toutes les unités relevées dans un seul champ sémantique autour d'un noyau formé par *navigation aérienne*, «carrefour à partir duquel rayonnent toutes les bifurcations du champ sémantique englobant, [...], les concepts et les signes qui ont trait au déplacement des hommes dans l'air à l'aide d'une machine» (31). Il observe qu'un grand nombre de lexèmes à l'intérieur du champ existent déjà dans d'autres champs, mais que l'opération de transfert faite à partir d'un plan de signification stable ne crée pas un autre plan aussi stable. Il termine par une interrogation, cruciale pour le point qui nous intéresse : «Navigation est-il le même signe quand il signifie "navigation maritime" et "navigation aérienne" (71). Le chapitre deux examine le champ morpho-sémantique d'aviation, un néologisme de forme et, partant, non pertinent dans la perspective de notre étude. Le chapitre suivant est entièrement consacré aux vocabulaires alimentant l'opération de transfert terminologique. Voici les domaines d'activité par ordre d'apparition dans le texte :

Les sciences connexes de l'aéronautique : physique, mathématiques, astronomie, météorologie, mécanique.

Les sciences naturelles : ornithologie, entomologie, anatomie, ichtyologie.

La marine.

Les transports : maritime, ferroviaire.

Les principales observations concernent les moyens linguistiques de transfert (196-202). Les néologismes d'emploi sont classés en deux catégories : les lexèmes simples et les syntagmes nominaux complexes. Des premiers, Guilbert observe que l'intégration au nouveau champ se fait grâce à un réseau d'oppositions distinctives et il ajoute :

«C'est la constance des rapports contextuels ainsi noués qui les distingue des emplois métaphoriques accidentels du discours. C'est aussi leur appartenance à un ensemble lexical d'origine qui leur confère un statut différent de la métaphore épisodique.» (197)

¹⁰ En privilégiant, il va de soi, les données pertinentes à notre étude.

Des arguments sur lesquels nous aurons l'occasion de revenir. Il note de plus qu'une opposition de type *navigation maritimelnavigation aérienne* provoque souvent le transfert de lexèmes en série comme *hublot* et *hauban*, qui font ainsi double emploi. Les syntagmes forment des unités de significations complexes autonomes qui apparaissent sous différentes formes. Dans le cas d'un emprunt à la physique ou aux mathématiques, le syntagme se compose d'un premier élément emprunté (une notion théorique, par exemple) auquel s'ajoute une expansion nominale, comme dans *surface de sustentation* ou *angle d'attaque*. Dans d'autres cas, le deuxième élément, de par son contenu sémantique, agit comme marqueur de champ et opère ainsi le transfert; c'est ce deuxième procédé, le plus fréquent, qui est présenté par Guilbert comme caractéristique de la formation d'un vocabulaire. Nous reproduisons l'analyse de l'adjectif de transfert *aérien* en ne rapportant, pour des raisons d'intelligibilité, qu'une partie des exemples.

1. Le premier élément est représenté par un terme d'un autre domaine

wagon aérien
navire aérien
dériveur aérien
mer aérienne
touriste aérien
nageoire aérienne
combat naval aérien
cloche à plongeur aérienne
chevaux de halage aériens

2. Le premier élément désigne un concept de vaste compréhension

équilibre aérien mouvement aérien direction aérienne stabilité aérienne +

3. Le premier élément appartient à la langue générale

exercice aérien scène aérienne bond aérien montagne russe aérienne

Dans ce dernier cas, les syntagmes sont intégrés selon le hasard du discours par la fantaisie de l'auteur, Guilbert les associe à la métaphore.

Le dernier chapitre traite du vol et des différentes parties de l'avion, il ne présente qu'un intérêt secondaire dans le cadre de notre étude. Quant aux conclusions, elles s'organisent autour de six thèmes, nous les résumons à partir de deux points du vue : extralinguistique et linguistique.

Selon Guilbert, le plan extralinguistique représente le passage de la théorie à la pratique. À partir d'une accumulation de données issues des sciences (physique, ornithologie, etc.) et des techniques plus anciennes (transport maritime ou ferroviaire), il y a formulation de nouvelles données théoriques vérifiées par l'expérimentation de modèles élémentaires. La recherche technique, dont la progression s'organise selon les lois de

l'histoire, ne régit pas le processus linguistique puisque l'«évolution des plans du signifiant et du signifié¹¹» n'est pas nécessairement concomitante, mais cette recherche sert plutôt d'impulsion au processus. Le terme *aéroplane*, par exemple, désigne tout d'abord un aérostat de forme plane, puis un appareil de type moderne. L'expérience peut donc évoluer sans susciter de nouvelles dénominations et Guilbert de conclure à une relative autonomie du plan notionnel par rapport au plan linguistique. Par ailleurs, les deux types de relations présentant un caractère systémique observable dans un ensemble lexical donné, estime-t-il, sont la structure du contenu sémantique et la structure morpholexicale.

Malgré la nouveauté du domaine d'activité, le nombre de néologismes d'emploi dépasse de beaucoup celui des néologismes de formes, mais néanmoins les mutations linguistiques à l'intérieur des différents lexiques particuliers s'épaulent mutuellement et retentissent sur la langue. Il existe, de plus, une forte tendance à la description, comme le montrent les multiples syntagmes nominaux complexes à valeur analytique. Ces syntagmes sont la plupart du temps formés à partir d'un lexème, emprunté à un autre lexique, auquel s'ajoute une ou plusieurs déterminations à valeur descriptive permettant l'intégration au nouveau vocabulaire. Il y a également utilisation de la composition synthétique à partir d'éléments grecs ou latins, comme si le développement de chaque science ou de chaque technique, note-t-il, contribuait à généraliser certains éléments susceptibles de générer des séries analogiques et permettait au mécanisme de composition synthétique de s'installer ainsi graduellement dans la langue. On retrouve également bon nombre de mots de la langue commune, comme oiseau, aile, navire, appareil, introduits dans le nouveau domaine par une sorte de métaphore, mais dont la permanence dans le lexique interdirait, selon lui, d'y voir une «création poétique».

En outre, une remarque de C. Bally sur les similitudes entre l'expression littéraire et l'expression technique, en ce qui a trait à la recherche d'objectivité, d'exactitude et au rejet de la banalité de la langue commune, se vérifie, surtout dans la langue des brevets où les inventeurs semblent vouloir souligner un acte de création par un acte de dénomination, comme si la propriété juridique pouvait être mieux «garantie» par une sorte de «propriété linguistique».

Pour Louis Guilbert, le lexique, relais direct entre l'expérience du monde et la langue, présente bien une organisation linguistique, mais cette organisation ne correspond pas directement à la structuration de l'expérience.

LE TRANSFERT TERMINOLOGIQUE

Notre intention n'est pas de faire ici la critique de la méthode ou des analyses de Louis Guilbert, mais bien de discuter et de relier certaines de ses observations, comme le statut sémantique d'une tête de syntagme (produisant une série syntagmatique). Au chapitre un, Guilbert s'interroge certes sur l'identité de navigation dans navigation aérienne et navigation maritime, mais il oublie d'y revenir lorsqu'il décrit, au chapitre trois, de nombreux cas de transfert sans toutefois en dégager toutes les implications. Il observe, en effet, qu'à un élément emprunté s'ajoute une détermination souvent adjectivale. Cependant,

¹¹ Cette «séparation» par Guilbert des deux plans n'apparaît pas très orthodoxe vu le caractère indissociable du sã et sé dans le signe, illustré par le célèbre exemple du recto verso d'une feuille. Mais peut-être ne faut-il y voir qu'un glissement sémantique involontaire...

en ajoutant aérien à navigation, on ne fait pas qu'un simple transfert de lexème, on opère une transformation du noyau sémique du lexème emprunté puisque aérien fait bel et bien disparaître le sème «eau» de navigation, si bien qu'on voit apparaître l'expression navire marin (Guilbert 1965 : 199). La navigation sert de notion de référence, de prototype dirait Lakoff (1987), mais ce faisant le lexème acquiert alors un statut de générique et oblige à la création d'un terme spécifique pour désigner ce qu'il désignait avant le transfert. Si l'expression navire marin laisse une impression pléonastique, c'est qu'elle n'est pas passée dans la langue au même titre que la série navigation maritime, navigation aérienne, navigation spatiale. Une transformation à répétition du noyau sémique de navigation permet de nos jours la navigation sur des réseaux informatiques. Il semble bien que ces faits illustrent, on ne peut mieux, une certaine forme de structuration du lexique en terme de valeur. Ce phénomène, non exclusif au domaine technoscientifique, révèle en outre un trait de la nomination.

Vient ensuite la question de la relative autonomie des plans notionnel et linguistique. L'exemple de Guilbert est celui d'aéroplane, il lui sert à montrer la non coïncidence de l'expérience linguistique et de l'expérience technique. Cette induction est effectivement possible. Or, en posant le problème différemment, on peut utiliser le même exemple pour illustrer exactement la situation inverse. Et pour cause, tous les signes utilisés dans la période étudiée sont de nature instable : plusieurs disparaîtront, certains passeront dans la langue et d'autres, comme aéroplane, connaîtront des fluctuations avant de se fixer. Le signe aéroplane a bien survécu à l'objet qui l'avait fait naître — reste à savoir si l'objet désigné était réel ou virtuel — et on peut penser que l'existence de la série aéro y est pour quelque chose. Comme le fait remarquer Eco (1992 : 111-112), si les êtres humains «instituent et réorganisent les codes, c'est seulement parce que [les codes] existent déjà». Eco voit dans l'invention d'un signe, un signe flou, imprécis, puisque, ditil, «[l]es inventions ne s'organisent pas selon un système d'oppositions nettes, mais le long d'un continuum gradué [...]». Ainsi, à sa suite, on peut constater que la notion d'atome a bien évoluée depuis Démocrite : description conceptuelle après description conceptuelle, le mot atome est en quelque sorte marqué par l'histoire, marqué par des étapes successives de nomination et, en ce sens, il arrive que l'étymologie puisse parfois se superposer à l'histoire des sciences telle une généalogie de la nomination.

MÉTAPHORE, ENCORE ET TOUJOURS LA MÉTAPHORE

D'emblée, Guilbert semble écarter la métaphore du processus de nomination. En effet, lorsqu'il mentionne la métaphore dans les questions de transfert, c'est pour la rejeter en arguant, rappelons-le, que «c'est leur appartenance à un ensemble lexical d'origine qui leur confère un statut différent de la métaphore épisodique.» (197). De même, dans les cas d'utilisation de mots courants, comme on l'a vu plus haut, c'est la permanence dans le lexique qui sert d'empêchement. Associant création poétique et métaphore, rien de surprenant qu'il procède de telle sorte. Ces arguments ne reposent sur rien de véritablement étoffé, mis à part l'argument d'autorité¹². Soulignons tout de même l'existence d'une certaine intuition.

¹² Il faut dire que la métaphore était pratiquement évacuée d'office du champ d'investigation de la linguistique structurale.

Il existe toute une kyrielle¹³ d'ouvrages et d'articles portant sur la métaphore. Mode pléthorique ou passion, ce trope, mille fois ramené, mis à toutes les sauces, se dispute avec la métonymie et la synecdoque la place centrale de la rhétorique. Rares pourtant sont les analyses qui offrent plus qu'un développement des idées exposées par Aristote. Certains, comme Nietzsche, ¹⁴ adoptent une position extrême en avançant que le langage est par essence métaphorique et nient l'autre pôle, celui de l'écart ou de la figure ornementale. Les études sur la métaphore privilégiées sont ici celles qui s'intéressent à la dimension cognitive de la métaphore.

Notre démarche consiste à voir si les transferts terminologiques présents dans la formation de vocabulaires techniques sont de nature métaphorique en les soumettant à l'épreuve de quelques conceptions de la métaphore.

LA PROPOSITION ARISTOTÉLICIENNE

«La métaphore consiste à donner à une chose un nom qui lui est étranger en glissant du genre à l'espèce ou de l'espèce au genre, ou de l'espèce à l'espèce, ou bien par analogie.» (*Poétique* 1997 : 49)

À première vue, on trouve dans cette définition quatre types de métaphores, mais on a dit souvent qu'Aristote donnait un sens générique au terme *métaphore*, car les deux premiers types correspondent en fait à la synecdoque¹⁵. Eco (1988) propose de voir le passage de l'espèce à l'espèce non pas comme une métaphore à trois termes, à intersection sémique, mais bien comme un rapport analogique, ce qui laisse une seule véritable catégorie de métaphore. Eco (1988 : 158) fait remarquer que le dernier type, comportant quatre termes (A est à B ce que C est à D¹⁶), correspond à la métaphore cognitive puisque, dans ce cas, Aristote expose ce que la métaphore fait connaître alors que dans les autres cas, il explique comment se produit et se comprend la métaphore. L'acquisition de connaissances se fait par une sorte de conflit conceptuel¹⁷ — la coupe de Dionysos ne ressemble pas en effet au bouclier d'Ares — et, ce faisant, on y apprend quelque chose à la fois sur le «métaphorisé» et sur le «métaphorisant». Étant donné l'existence d'un conflit ou incompatibilité, on ne peut affirmer, contrairement à certaines croyances bien ancrées, que l'analogie se fonde sur la ressemblance, elle fait plutôt «surgir» la ressemblance¹⁹, sinon la métaphore par analogie ne pourrait nous dire que ce que l'on sait déjà.

¹³ Plus de 4000 titres relevés en 1971, cf. Danesi (1988 : 9).

¹⁴ Cf. Kremer-Marietti (1993).

¹⁵ Quoique la traduction utilisée par Le Guern (1973 : 32) n'en présente que trois, l'analogie étant présentée comme le procédé de passage...

¹⁶ L'exemple d'Aristote : la coupe de Dionysos et le bouclier d'Ares.

¹⁷ Pour une analyse tropologique à partir de la notion de «conflit conceptuel», cf. Prandi (1992).

¹⁸ Le mot *surgir* remplace l'expression *crée* utilisée lors de la communication pour tenir compte de la discussion lancée par G. Kleiber sur la création à partir de «rien» et la préexistence de la «ressemblance». Ce changement ne modifie pas notre réflexion puisque notre posture épistémologique, bien établie au début du texte, rejette un constructivisme pur et dur. Notre point de vue est plutôt «interactionniste» point de vue que l'on peut rapprocher de celui de Robert (1993).

¹⁹ Pour l'analogie appliquée à la création de concepts scientifiques, cf. Schlanger (1991) et Halleyn (1987).

Cette proposition appliquée à navigation donne l'analyse suivante : navigation (espèce) passe à navigation aérienne (espèce) et, par rétroaction, navigation (espèce) passe à navigation (genre), crée au passage navigation maritime (espèce) et permet la série navigation spatiale, sur internet, etc. Le procédé est métaphorique. Le lexème navigation subit la transposition; il semble toutefois superficiel de n'y voir que la perte d'un sème /eau/. Nous y reviendrons.

LA PHÉNOMÉNOLOGIE OU PHANÉROSCOPIE DE C.S. PEIRCE

Il n'est pas facile d'extirper la métaphore de la sémiotique fondée par Peirce puisque cette théorie présente un système conceptuel complet dont les éléments une fois isolés perdent de leur sens. Par conséquent, la présentation qui suit possède inévitablement un caractère réducteur.

La sémiotique peircéenne ou phanéroscopie veut rendre compte de l'acquisition de connaissances par un sujet vu comme étant à l'intérieur de la pensée et comme utilisant, pour cette acquisition, des signes qui ne forment pas des unités fixes, mais qui correspondent à un moment de la sémiosis. Le mouvement de la sémiosis va de la priméité (la pensée en terme de pure possibilité) à la secondéité (la pensée événementielle actualisant le premier) vers la tercéité (la connaissance régissant le second, la pensée codifiée) dans un mouvement ad infinitum. À ces trois paliers, trois ordres logiques : l'abduction, l'induction, et la déduction. L'abduction «suppose quelque chose de différent de ce qui a été observé et fréquemment quelque chose qu'il serait impossible d'observer directement.» (1978 : 223). Autrement dit, devant un nombre élevé de solutions impossibles à vérifier une par une, un choix s'opérera, intuitivement, par plausibilité. L'abduction créatrice en sciences se produit lorsque la vraisemblance se réalise. Dans cet univers sémiotique, un signe se divise en icône, indices et symbole. Une icône (deuxième) ne peut représenter une idée (troisième) qu'en termes de possibilité (premier), mais un signe peut être iconique, il portera alors le nom d'hypoicône. Pour Peirce, une métaphore est une hypoicône «qui représent[e] le caractère représentatif d'un represantamen en représentant un parallélisme dans quelque chose d'autre [...]» (1978 : 149). La particularité de cette définition réside dans l'inclusion d'une capacité de représentation dans la représentation, une telle capacité naît dans un parallélisme avec autre chose, c'est donc une capacité latente d'un signe. Une telle métaphore peut s'appliquer au transfert de nomination, mais non à la dénomination, néanmoins le pouvoir explicatif de cette définition est plutôt faible. Par ailleurs, on reconnaît la filiation du continuum proposé par Eco dans le mouvement infini de la sémiosis et, dans cette logique temporelle, l'opération de transfert pourrait correspondre à un saut fait de l'abduction (premier) à la déduction (troisième), c'est ce que nous retiendrons de l'apport de Peirce.

LAKOFF ET JOHNSON

Dans leur ouvrage conjoint (1980), Lakoff et Johnson accordent à la métaphore une place prépondérante dans la compréhension du monde et de soi et postulent qu'elle nous est aussi essentielle que le sens du toucher. Leur champ d'observation est le langage de la vie quotidienne, ce qui peut sembler loin de nos préoccupations. Les métaphores structurent en

partie la vie de tous les jours et se retrouvent abondamment dans le langage littéral²⁰. Le nombre élevé d'occurrences de ces métaphores quotidiennes leur sert d'argument de base pour critiquer ce qu'ils appellent le mythe de l'objectivisme. Cette critique, qui ramène à l'avant-plan l'hypothèse Sapir/Worf, sera développée par Lakoff (1987). À la suite de cette réflexion en collaboration, les deux auteurs publieront en 1987, chacun de leur côté, des ouvrages pertinents pour notre réflexion.

Chez Lakoff (1987), nous examinerons la notion de prototype et la notion de catégorie radiale. Le prototype, dans sa première version, est une notion opérante dans l'analyse des séries syntagmatiques terminologiques. En effet, comme nous l'avons déjà montré (Portelance 1997), dans une série syntagmatique les différentes expansions de la tête de syntagme constituent des distinctions par rapport à un élément central de la série, élément que l'on peut voir comme le prototype d'une catégorie d'objets et il arrive même qu'avec une innovation technique un élément central soit détrôné au profit d'un autre, ce qui a pour effet de modifier en conséquence les syntagmes appartenant à la série. Dans la catégorisation, le recours à un prototype n'est qu'indirectement lié à la métaphore, aussi, cette notion, en permettant une certaine description du transfert terminologique, n'autorise pas pour autant d'y apposer l'étiquette de métaphore.

La catégorie radiale, selon Lakoff, forme un ensemble dont les membres ne présentent pas un trait commun, mais sont liés par une ressemblance de famille, concept emprunté à Wittgenstein (1986 : 148). Rappelons que ce concept avait été initialement proposé pour décrire un réseau complexe d'analogies, analogie d'ensemble comme de détails, et son exemple de référence, tout sauf anodin, était l'ensemble de tous les jeux (jeux de cartes, jeux de dames et... jeux de langage). Dès lors, le prototype n'est plus un élément central de la catégorie, comme dans la première mouture de la théorie du prototype, il est formé de propriétés définissant la ressemblance. Kleiber (1990) discutant de l'évolution de cette théorie qualifie, non sans une pointe d'ironie, cette deuxième définition du prototype de «prototypique». Lakoff se sert des catégories linguistiques du dyirbal comme illustration : dans cette langue, les noms appartenant à une catégorie sont en quelque sorte chaînés les uns aux autres par leur appartenance à un système de croyances mythologiques plutôt que par des traits sémantiques communs.

En examinant, à partir de ce point de vue, la série qui nous concerne, il semble difficile de déterminer un invariant sémantique entre navigation maritime, aérienne, spatiale et sur internet, puisque le sème /déplacement/ paraît impossible dans le cas de navigation sur internet. On pourrait certes avancer que cette dernière expression est métaphorique et que les autres ne le sont pas. À l'examen des transpositions antérieures de navigation, ce raisonnement tient du cercle vicieux; pour en sortir et trouver l'analogie, il faut identifier l'élément qui permet la ressemblance. Nous croyons que ce rapport analogique pourrait être l'exploration, en tant que «passage d'un site à un autre», parce que navigation par internet survient après navigation spatiale qui fait surgir, avant /exploration/, les sèmes /voyage/ et /découverte/. Ainsi, à chaque ajout, la ressemblance de famille se développe et on peut en effet y voir une sorte de lien radial; par conséquent, en vertu du concept saussurien de valeur, l'analyse sémique doit alors être repensée au fur et à mesure que la famille s'agrandit.

²⁰ Leur approche pose un problème de la définition du littéral, sur cette question cf. Nuessel dans Danesi (1988).

Appliquée au développement d'une théorie scientifique, la ressemblance de famille permet de voir une rupture épistémologique comme un éclatement du réseau analogique de ressemblances et justifie le choix par Kuhn (1970 : 74) de l'expression *paradigme* pour désigner les courants scientifiques.

Avec The Body in the Mind (1987), Johnson propose de redonner à l'expérience du corps la place qu'il estime lui revenir dans l'élaboration des projections métaphoriques et des schémas²¹ élaborés mentalement. La station du corps, explique-t-il, permet par exemple de projeter la quantité en terme de verticalité : plus est haut; moins est bas. Se dégagent alors des catégories, telles devant/derrière, intérieur/extérieur, contenant/contenu, etc., qui deviennent, dans une approche aux accents kantiens, les données sensibles de base servant à toute conceptualisation et donnent un point de départ à une théorie de l'imagination. Cette posture séduit dans ce qu'elle contient de promesses eu égard à une compréhension de l'être parlant. De plus, cette représentation «incarnée» a dégagé autour de la représentation de l'espace un champ d'étude fécond actuellement fort prisé en sémantique cognitive. En revanche, ce retour à la phénoménologie, on pense à Merleau-Ponty, se radicalise dans les travaux de Varela et al. (1991), lui qui, 10 ans auparavant, avançait que notre expérience n'a pas d'autre fondement que l'existence d'une histoire biologique et sociale commune conduisant à une succession d'interprétations (dans Watzlawick 1988: 344). Appliquée aux nomenclatures, pareille vision leur accorderait une mémoire des mots absolument vertigineuse..., mais impossible à reconstituer.

EN GUISE DE CONCLUSION

L'invention tient en quelque sorte du mystère, de nombreux témoignages parlent même d'illumination²². Ce qui est fascinant dans le mécanisme d'abduction, c'est comment, en brassant du connu, on arrive à faire surgir du «nouveau». À l'instar d'Halleyn (1987) nous croyons que l'imagination scientifique présente une structure rhétorique : la pensée, en passant de l'abduction à la déduction, est métaphorique; en procédant par bond, elle créé une analogie, une idée qui apparaît au sujet vraisemblable. L'étape de vérification valide ou invalide l'hypothèse. Mais il faut se garder d'y voir une analogie simple; il serait plus réaliste, il nous semble, de concevoir un réseau de relations analogiques. Quant au mécanisme déclencheur de l'idée, nous proposons²³ de l'associer à la notion de «voir comme» utilisée par Wittgenstein pour traiter des figures ambiguës, comme celle du lapin-canard (1986 : 326) dont il est dit :«[...] nous l'interprétons et la voyons comme nous l'interprétons». En transposant «voir comme» à la métaphore, non seulement nous installons la pensée abductrice dans une dimension spatio-temporelle, mais nous affirmons que la métaphore cognitive permet simultanément d'affirmer et d'infirmer : nommer ou créer un objet par métaphore permet en effet de dire ce qu'il est et ce qu'il n'est pas. En ce sens, l'analogie créée par la métaphore manifeste également un caractère binaire. Dans les cas de transferts terminologiques, c'est donc la nomination plutôt que la dénomination qui est métaphorique; c'est pourquoi, à notre avis, les terminologies spécialisées forment un réservoir de possibilités de «voir comme» et sont

²¹ Ce terme (schemata dans le texte anglais) est repris à Kant. On peut voir dans cette proposition une démarche conceptuelle inverse de celle qui a produit le concept de «schéma corporel».

²² Cf. l'ouvrage de A. Koestler (1964), *The Act of Creation*, New York, MacMillan cité par Schlanger (1991).

²³ Une proposition similaire est discutée par Ricoeur (1975).

elles-mêmes, dans certains cas, issues de l'opération «voir comme». Un des meilleurs exemples, à cet égard, se trouve en physique quantique où les caractères attribués aux quarks sont la couleur, la vérité, l'étrangeté, le charme et la beauté... Le nouveau n'est créé qu'à partir d'une idée, et pour que cette idée se transforme en théorie, elle doit passer par l'étape de la diction. En outre, cette mise en langage, étape incontournable, installe une rétroaction sur l'appareil linguistique servant à la nomination.

Par ailleurs, comme le fait remarquer Schlanger (1991), une œuvre théorique devenue culturelle, comme celle de Copernic ou de Kepler, joue quelquefois un rôle paradigmatique et peut, en tant qu'oeuvre, devenir un objet d'étude («poétique» dirait Halleyn 1987). Entre 1861 et 1891, l'aviation n'est encore qu'une fiction, un vieux rêve en voie de se réaliser certes, mais un rêve que l'on peut faire remonter au mythe d'Icare. Cassirer (1946 : 88), traitant du pouvoir de la métaphore, affirme qu'il y a réciprocité entre la forme du langage et celle du mythe, qui se conditionnent et se déterminent l'une l'autre. Par ailleurs, Peirce, en fondant le pragmatisme, pose qu'il n'y a d'utilité ou de rationalité dans les croyances que lorsque ces dernières permettent l'action (1993 : 333). Par la suite, l'aviation non seulement s'actualisera, mais ses développements subséquents font de la «navigation» plus qu'une technique; la navigation devient structure imaginante car, à travers le prisme de la métaphore spatiale, elle permet aujourd'hui de voyager dans un espace virtuel qui se veut l'icône d'un village global.

RÉFÉRENCES

ARISTOTE (1997): Poétique, trad. du grec par O. Bellevue et S. Auffret, Mille et une nuits, 93p.

ARISTOTE (1973): Rhétorique, livre III, Paris, Les Belles Lettres, 173 p.

CASSIRER, E. (1953): Language and Myth, trad. de l'all. par S.K. Langer (1946 pour la 1^{re} éd.) New York, Dover, 103 p.

DANESI, M. (dir) (1988): *Metaphor, Communication & Cognition*, Monography series of the Toronto Semiotic Circle, no 2, University of Toronto, 117 p.

ECO, U. (1992): La production des signes, Paris, Le Livre de poche, 124 p.

ECO, U. (1988): Sémiotique et philosophie du langage, [1984], trad. de l'ital. par M. Bouzaher, Paris, PUF, 285 p.

GUILBERT, L. (1965): La formation du vocabulaire de l'aviation, Paris, Larousse, 709 p.

HALLEYN, F. (1987): La structure poétique du monde: Copernic, Kepler, Paris, Seuil, 275 p.

HOTTOIS, G. (1989): Penser la logique, Bruxelles, De Boeck Université, 273 p.

JOHNSON, M. (1987): The Body in the Mind: the Bodily Basis of Meaning, The University of Chicago Press, 233 p.

KLEIBER, G. (1990): La sémantique du prototype, Paris, PUF, 199 p.

KREMER-MARIETTI, A. (1992): Nietzsche et la rhétorique, Paris, PUF, 269 p.

KUHN, T.S. (1970): La structure des révolutions scientifiques, trad. de l'anglais par Laure Meyer, Paris, Champs, Flammarion, 284 p.

C. Portelance

- LAKOFF, G. (1987): Women, Fire and Dangerous Things. What Categories Reveal about the Mind, 614 p.
- LAKOFF, G. et M. JOHNSON (1980): *Metaphors we live by*, The University of Chicago Press, 242 p.
- LATOUR, B. (1995): La science en action [1989], trad. de l'angl. par M. Biezunski, Paris, Folio Gallimard, 663 p.
- LE GUERN, M. (1973): Sémantique de la métaphore et de la métonymie, Paris, Larousse, 126p.
- PEIRCE, C.S. (1993): À la recherche d'une méthode, [inédit] trad. de l'angl. et édité par J. Deledalle et M. Balat sous la dir. de G. Deledalle, Presses universitaire de Perpignan, 375p.
- PEIRCE, C.S. (1978): Écrits sur le signe, rassemblés, traduits et commentés par G. Deledalle, Paris, Seuil, 263 p.
- PORTELANCE, C. (1997): «De la nomination. Catégorisation et syntagmatique», Toronto, Sémiotique appliquée/Applied Semiotics, http://www.epas.utoronto.ca:8080/french/as-sa/
- PRANDI, M. (1992): Grammaire philosophique des tropes, Paris, Minuit, 253 p.
- PUTNAM, H. (1984): Raison, vérité et Histoire [1981], trad. de l'angl. A. Gerschenfeld, Paris, Minuit, 242 p.
- RICOEUR, P. (1975): La métaphore vive, Paris, Seuil, 412 p.
- ROBERT, S. (1993): Les mécanismes de la découverte scientifique, Presses de l'Université d'Ottawa, 262 p.
- STENGERS, I. et J. SCHLANGER (1991): Les concepts scientifiques [1988], Folio Gallimard, 190 p.
- VANDELOISE, C. (1986): L'espace en français: sémantique des prépositions spatiales, Paris, Seuil.
- VARELA, F.J., THOMPSON, E. et E. ROSCH (1991): *The Embodied Mind*, Cambridge, The MIT Press, 308 p.
- WATZLAWICK, P. (dir) (1988): L'invention de la réalité [1981], trad. de l'all. par A.-L. Hacker, Paris, Seuil, 374 p.
- WITTGENSTEIN, L. (1986): Tractatus logico-philosophicus [1908] suivi de Investigations philosophiques, trad. de l'all. par P. Klossowski, Paris, TEL Gallimard, 364 p.

CLASSES D'OBJETS ET BASE DE CONNAISSANCES¹

Man-Ghyu PAK

Université de Kwandong, Corée du Sud

1. PRÉSENTATION DU PROBLÈME

Dans le cas particulier de la compréhension de la langue par une machine, ce qui nous intéresse ici, un des points fondamentaux est bien sûr le traitement des informations sémantiques. Or, la détermination du sens des phrases est le plus souvent difficile à cause de l'ambiguïté des mots de la langue (polysémie). Comme la sémantique nous pousse à aborder la relation entre les mots et la réalité, ce ne sont plus des connaissances linguistiques qui nous permettent de comprendre ces phrases, mais nos connaissances extralinguistiques, c'est-à-dire celles qui touchent le monde. Ces connaissances sur la réalité, sur les entités du monde, permettent une appréciation différente sur le sens des phrases.

Considérons les phrases :

- (1) J'ai envie de porter quelque chose d'à la mode
- (2) Elle porte quelque chose d'affriolant
- (3) Elle porte celui qui est à rayures
- (4) Il porte celle qui est (fripée + élimée)
- (5) Elle ne veut porter que ce qui lui va le mieux
- (6) Elle va porter quelque chose d'un peu plus (extravagant + original)

Il semble très difficile de faire deviner au système que le complément d'objet de *porter* signifie un <vêtement>: en effet d'une part ce complément est pronominalisé, d'autre part le verbe *porter* est polysémique. La méthode dont on se sert généralement pour résoudre ce type de problème consiste à construire une base de connaissances de grande envergure qui comprenne des connaissances extralinguistiques, c'est-à-dire des connaissances générales sur le monde, et à la faire utiliser par le système. Mais cette méthode pose bien des problèmes : par exemple, 1) elle ne donne pas toujours des résultats précis parce qu'elle se fonde sur la théorie des probabilités et 2) elle est très coûteuse.

¹ Cette recherche a été financée par la bourse de recherches de l'Université de Kwandong. Je tiens à remercier Gaston Gross et Denis Le Pesant d'avoir lu le manuscrit et de m'avoir donné de précieux conseils.

Dans cet exposé, nous allons montrer que la notion d'opérateurs appropriés permet de résoudre ce genre de problème. Les opérateurs appropriés sont des opérateurs (verbes, adjectifs, noms) qui définissent une classe d'objets. Cette notion a été introduite par G. Gross pour décrire les traits sémantiques des noms des langues naturelles. La classe d'objets <vêtements>, par exemple, est définie, en position de N0, par les opérateurs appropriés tels que «être à la mode», «aller bien à» et «être à rayures». Notre méthode consiste à recourir au dictionnaire des opérateurs appropriés pour lever les ambiguïtés résultant de la pronominalisation : nous croyons que l'on peut considérer les classes d'objets, non seulement comme un système de connaissances linguistiques, mais aussi comme un système de connaissances sur le monde.

Notre travail a pour but de montrer que les opérateurs appropriés permettent de lever les ambiguïtés du pronom et qu'ils peuvent remplacer, au moins en partie, une base de connaissances parce qu'ils contiennent en fait des connaissances sur le monde.

2. LIMITES DES BASES DE CONNAISSANCES

Dans cette section, nous allons faire une évaluation critique des bases de connaissances.

Notons, d'abord, l'inconvénient consistant à accumuler et à gérer des données de taille énorme qui occupent beaucoup de place et qui sont donc très coûteuses. De plus, en réalité, il est impossible de représenter et d'enregistrer toutes les connaissances sur le monde : il est illusoire de développer une machine dotée des mêmes compétences que le cerveau humain.

Cette méthode ne peut s'appliquer qu'à un domaine de spécialité. C'est pourquoi, aujourd'hui, elle n'est utilisée, généralement, que pour la constitution de <systèmes experts>². Mais les phrases que nous donnons en exemples appartiennent non à un domaine de spécialité, mais à la langue de tous les jours, c'est-à-dire qu'elles sont des expressions de base. Ceci nous conduit à dire que les bases de connaissances ne permettent pas de reconnaître le sens des phrases, même fondamentales.

Généralement, une base de connaissances est représentée par des expressions métalinguistiques : par exemple, la règle de production, les langages de primitives, les réseaux sémantiques, etc. La règle de production est du type <si - alors>. Pour la compréhension de la phrase (3), on formule la règle de production suivante :

Si quelque chose est à rayures, alors c'est un vêtement.

Dans un langage de primitives (Schank 1975), les notions sont décomposées en atomes de sens, les primitives sémantiques, dont la combinatoire engendre les sens à représenter. Les réseaux sémantiques ($semantic\ network^3$) sont des graphes étiquetés dont les nœuds

² Un système expert se compose de la base de connaissances et du moteur d'inférence: la première à son tour est constituée de la base de règles, ensemble des connaissances construites par les experts, et de la base de faits, construite par les usagers pour résoudre des problèmes données.

³ On pourra citer comme travail représentatif *Propositional Semantic Network* de Catelfranchi.

représentent des notions (appelées communément *concepts*) et les arcs des relations. À l'origine, les réseaux sémantiques ont été introduits pour tenter de représenter la mémoire associative humaine. Ces méthodes formalistes ont toutes recours à une forme de calcul pour réaliser des inférences et reconnaître des proximités sémantiques ou des relations de paraphrase. Mais le problème est que ces formalismes sont très compliqués à construire et à gérer parce que la structure des nœuds est trop simple. Par conséquent, la recherche d'une expression complexe prend beaucoup de temps.

Le problème le plus grave posé par les bases de connaissances est que celles-ci ne donnent pas toujours de résultats précis du fait qu'elles se fondent sur la théorie des probabilités. Puisqu'elles s'appuient sur des informations non structurées, elles ne peuvent nous donner que l'ordre d'application et la probabilité statistique. Elles sont donc peu fiables.

3. NOTION DE CLASSE D'OBJETS ET D'OPÉRATEURS APPROPRIÉS

Le traitement automatique traditionnel se faisait sans composante sémantique, à savoir sans analyse sémantique. Mais il est aujourd'hui clair que la traduction ne peut pas fonctionner sans traitement des informations sémantiques. La plupart des systèmes se servent de <traits syntactico-sémantiques> tels que humain, concret, abstrait, etc. pour lever les ambiguïtés des mots polysémiques. Les traits sont certes utiles: concret et abstrait permettent de donner une bonne traduction à deux emplois différents du verbe français porter: porter par to bear en cas de complément abstrait et par to carry si l'objet est concret, comme on peut le voir:

avec un objet abstrait dans:
 Luc porte un nom français
 Luc bears a French name

- avec un nom concret dans : Luc porte une valise Luc carries a case

Cependant ces traits sont insuffisants. Ils n'apportent pas les informations indispensables à la reconnaissance automatique des différents emplois de *porter*, dans la mesure où ces traductions ne valent pas pour tous les abstraits et tous les concrets. Par exemple, un *vêtement* réfère à un concret. Le verbe *porter* qui lui est approprié ne se traduit pourtant pas par *to carry* mais par *to wear*:

Il porte des vêtements de tous les jours He wears clothes for every day

Si le substantif *vêtement* est codé «concret» comme les termes *valise* ou *sac*, on voit que cette seule spécification ne permet pas de sélectionner la bonne traduction en anglais. Pour prédire la traduction *to wear*, il est nécessaire d'ajouter une information sémantique plus précise telle que «vêtement». G. Gross appelle ces nouvelles informations sémantiques des *classes d'objets*. Cette information doit être notée pour chaque substantif

désignant un vêtement. Le code, noté pour chaque élément de la classe, permet à un analyseur de prévoir la reconnaissance (ou la génération) de l'emploi et la bonne traduction.

Le problème est de déterminer la manière de définir les classes d'objets. Leur originalité réside dans le fait que celles-ci ne sont pas seulement définies de façon sémantique mais par les relations syntaxiques qu'elles entretiennent avec une ou plusieurs classes de verbes et d'adjectifs, appelées opérateurs appropriés⁴. Les opérateurs appropriés verbaux de la classe <vêtements> seraient : porter, enlever, mettre, changer de, cintrer, enfiler, passer, reprendre, retoucher (en position objet) ou aller bien, boulocher, tomber bien (en position sujet), etc., ce qui veut dire que la classe d'objets <vêtements> est définie par le fait d'être sélectionnée à la fois en position N1 en (7) ou N0 en (8) :

- (7) Paul porte N1
 Paul enlève N1
 Paul mettre N1
 Paul change de N1
 Paul cintre N1
 Paul enfile N1
 Paul passe N1
 Paul reprendre N1
 Paul retouche N1
- (8) No va bien à qqnNo boulocheNo fait des plisNo tombe bien

De la même façon, on peut définir la classe <vêtements> par les opérateurs adjectivaux, (9) ou (10) :

- (9) Paul est accoutré de N1Paul est boudiné dans N1Paul est vêtu de N1
- (10) **N0** est à la mode **N0** est à rayures

N0 est affriolant

N0 est ajusté

N0 est élimé

No est seyant

Voici un extrait de la liste des opérateurs qui définissent la classe d'objets <vêtements>5:

⁴ Pour plus d'informations sur le processus de formation d'une classe d'objets et sur la notion d'opérateurs appropriés, voir Le Pesant (1994).

⁵inc : (trait) inanimé concret, <vêtem> : classe d'objets <vêtements>, E : traduction anglaise.

1. Opérateurs verbaux :

a) En position d'objet :

mettre/N0:hum/N1:inc<vêtem>/E:to put on porter/N0:hum/N1:inc<vêtem>/E:to wear enlever/N0:hum/N1:inc<vêtem>/E:to take off changer/N0:hum/N1:inc<vêtem>/E:to take off changer/N0:hum/N1:inc<vêtem>/E:to take in at the waist enfiler/N0:hum/N1:inc<vêtem>/E:to slip on;to put on;to slip over passer (usure)/N0:hum/N1:inc<vêtem>/E: o slip in reprendre/N0:hum/N1:inc<vêtem>/E:to alter;to take in;to let out retoucher/N0:hum/N1:inc<vêtem>/E:to alter nettoyer à sec/N0:hum/N1:inc<vêtem>/E:to dry-clean repasser/N0:hum/N1:inc<vêtem>/E:to iron

b) En position de sujet :

aller bien à /N0:inc<vêtem>/N1:hum/E:to fit hum (taille) boulocher/N0:inc<vêtem> tomber bien/N0:inc<vêtem>/E:to fall to the feet faire des plis/N0:inc<vêtem>/E:to have a crease

2. Opérateurs adjectivaux :

a) En position de sujet :

à la mode/N0:inc<vêtem>/E:fashionable à plis/N0:inc<vêtem>/E:pleated à rayures/N0:inc<vêtem>/E:striped affriolant/N0:inc<vêtem:déshabillé>/E:titillating ajusté/N0:inc<vêtem>/E:fitted amidonné/N0:inc<vêtem>/E:starched élimé/N0:inc<vêtem>/E:worn thin fripé/N0:inc<vêtem>/E:crumpled seyant/N0:inc<vêtem>/E:becoming;flattering

b) En position d'objet :

accoutré/N0:hum/de N1:inc<vêtem>/E:got up in attifé/N0:hum/de N1:inc<vêtem>/E:got up in botté/N0:hum/de N1:inc<vêtem>/E:got up in botté/N0:hum/de N1:inc<vêtem>/E:strapped into sanglé/N0:hum/dans N1:inc<vêtem>/E:strapped up tight in vêtu/N0:hum/de N1:inc<vêtem>/E:dressed in

Nous présentons aussi un extrait de la la liste de la classe d'objets <vêtements> ainsi définie⁶ :

 $^{^6}$ G: genre, C: classe d'objets, inc: (trait) inanimé concret, D: domaine R: registre, S: spécifications, E: traduction anglaise, vêt du b: vêtement du bas, vêt du h: vêtement du haut, vêt de c: vêtement de corps, vêt coiff: coiffure, vêt chauss: chaussures.

```
bas/G:nm/C:vêtem du b/D:habill./R:/S:/E:stocking
béret/G:nm/C:vêtem coiff/D:habill./R:/S:/E:beret
bikini/G:nm/C:vêtem de c/D:habill./R:/S:/E:bikini
blouson de cuir/G:nm/C:vêtem du h/D:habill./R:/S:/E:leather jacket
blue jean/G:nm/C:vêtem du b/D:habill./R:/S:/E:jeans
boots/G:nmp/C:vêtem chauss/D:habill./R:/S:/E:boots
botte/G:nf/C:vêtem chauss/D:habill./R:/S:/E:boot
bottine/G:nf/C:vêtem chauss/D:habill./R:/S:/E:ankle-boot
cale on/G:nm/C:vêtem du b/D:habill./R:/S:/E:underpants;US:shorts
casque/G:nm/C:vêtem coiff/D:mil./R:/S:/E:helmet
casquette/G:nf/C:vêtem coiff/D:habill./R:/S:/E:cap
chandail/G:nm/C:vêtem du h/D:habill./R:/S:/E:pullover,sweater
chapeau/G:nm/C:vêtem coiff/D:habill./R:/S:/E:hat
chaussette/G:nf/C:vêtem du b/D:habill./R:/S:/E:sock
chaussure/G:nf/C:vêtem chauss/D:habill./R:/S:/E:shoe
chemise/G:nf/C:vêtem du h/D:habill./R:/S:/E:shirt
corset/G:nm/C:vêtem de c/D:habill./R:/S:/E:corset
costume/G:nm/C:vêtem (masc)/D:habill/R:/S:complet/E:suit
cravate/G:nf/C:vêtem du h/D:habill./R:/S:/E:tie
culotte/G:nf/C:vêtem du b/D:habill./R:/S:enfant/E:pants
gabardine/G:nf/C:vêtem du h/D:habill./R:/S:/E:gabardine
jean/G:nm/C:vêtem du b/D:habill./R:/S:/E:jeans
jupe/G:nf/C:vêtem du b/D:habill./R:/S:/E:skirt
jupon/G:nm/C:vêtem du b/D:habill./R:/S:/E:(waist) petticoat;US:half slip
manteau/G:nm/C:vêtem du h/D:habill./R:/S:/E:coat;overcoat
pantalon/G:nm/C:vêtem du b/D:habill./R:/S:/E:trousers;US:pants
pardessus/G:nm/C:vêtem du h/D:habill./R:/S:/E:overcoat,topcoat
pyjama/G:nm/C:vêtem/D:habill./R:/S:/E:pyjamas
```

L'existence de la classe des <vêtements> est justifiée par le fait qu'elle permet la définition des substantifs exigés par beaucoup de verbes et d'adjectifs.

Si on dresse la liste des classes d'objets de cette façon, on pourra faire identifier et reconnaître par le système les différents emplois d'un même verbe polysémique en les décrivant comme suit :

```
porter <vêtements>: porter une veste
porter <marque>: porter une cicatrice
porter <arme>: porter un coutelas
porter <signe>: porter un slogan (sur la poitrine)
porter <enfant>: porter son enfant en elle (depuis six mois)
porter <nom>: porter un drôle de nom
```

4. RÉSOLUTION SUR LA BASE D'OPÉRATEURS APPROPRIÉS

Nous allons maintenant montrer comment notre méthode résout le problème posé par les phrases (1)-(6). La résolution se fera par les étapes suivantes.

D'abord, le système fait l'analyse morphologique, puis syntaxique du texte donné. Ensuite, il tente de lever l'ambiguïté des mots en se référant au dictionnaire des classes d'objets. Ce travail se fait en deux étapes : D'abord, on étiquette le verbe par ses classes d'objets sur la base du dictionnaire. Ensuite, on reconnaît l'identité de l'acception en combinant le verbe et le substantif (c'est-à-dire le complément d'objet) qui ont en commun la même classe d'objets.

Cependant, lorsque le complément d'objet est un pronom, donc extérieur aux classes d'objets, une reconnaissance de ce type ne marchera pas. Dans ce cas, le système tient compte du type de pronoms. Classiquement, on classe les pronoms en trois catégories :

- les pronoms à usage introductif qui permettent d'introduire un objet ou un ensemble d'objets dans le discours ou dans la conversation (e.g. *quelque chose* dans les phrases (1), (2) et (6);
- les pronoms à usage anaphorique qui renvoient à une entité déjà apparue dans le discours (e.g. celui dans la phrase (3), celle dans la phrase (4));
- les pronoms à usage déictique qui permettent de désigner directement des entités du monde (ce dans la phrase (5); celui dans la phrase (3), celle dans la phrase selon les cas).

(La distinction entre anaphore et déictique est aujourd'hui très contestée. cf. Kleiber 1990 et 1991)

Pour les pronoms à usage introductif, ils n'ont pas beaucoup de sens en tant que tel. Mais, dans nos exemples, ils obtiennent du sens grâce à leur complément comme d'à la mode, d'affriolant et d'un peu plus (extravagant + original). Il faut donc faire reconnaître ce sens par la machine. Dans ce cas, le système recherche toutes les classes d'objets attribuées au verbe en consultant cette fois le dictionnaire des opérateurs appropriés. Si la séquence régie par le pronom en question est reconnue comme un opérateur approprié dans le dictionnaire, le système considère la classe d'objets définie par cet opérateur approprié comme étant le bon référent de ce pronom. S'il s'agit de traduction automatique, le système traduit le verbe par son homologue de la langue cible.

Au cas où le pronom en question est du type anaphorique, il est bien entendu que le système recherche d'abord dans le contexte linguistique des candidats à l'antécédent. Il sélectionne ensuite le candidat approprié, en appliquant la même technique que celle utilisée pour le pronom à usage introductif. (Pour les problèmes généraux de la reconnaissance des anaphores, voir Le Pesant (à paraître)).

Lorsque le pronom en question est du type déictique, le problème est encore plus délicat à traiter : en effet il faut examiner de façon exacte et exhaustive la situation

extralinguistique pour trouver le référent. Mais il existe des cas où les classes d'objets aident à reconnaître le référent. Nous y reviendrons.

Considérons, par exemple, le cas de la reconnaissance de *porter une jupe*. Si l'on fait passer cette séquence dans le dictionnaire des classes d'objets, le substantif *jupe* recevra la marque correspondant à sa classe d'objets de la façon suivante :

```
(11) jupe : <vêtements>
```

et le verbe porter recevra la marque suivante :

```
(12) porter1: N1 = <vêtements>
porter2: N1 = <marque>
porter3: N1 = <arme>
porter4: N1 = <signe>
porter5: N1 = <enfant>
porter6: N1 = <nom>
.....
```

La machine lèvera l'ambiguïté du verbe et reconnaîtra l'identité de l'acception en combinant le substantif et le verbe *porter1*, qui ont en commun le même code de classe d'objets <vêtements>.

Mais lorsque le complément d'objet est un pronom comme (1)-(5), la reconnaissance de ce type ne marche pas. Au cas où le pronom est du type introductif, à la suite de la programmation, le système recherche toutes les classes d'objets attribuées au verbe, ici <vêtements>, <marque>, <arme>, <signe>, <enfant>, <nom>, etc. en consultant cette fois le dictionnaire des opérateurs appropriés. Considérons d'abord le cas des phrases (1) et (2). Rappelons-les pour raisons de commodité:

- (1) J'ai envie de porter quelque chose d'à la mode
- (2) Elle porte toujours quelque chose d'affriolant

Si la séquence régie par le pronom en question est trouvée en tant qu'opérateur approprié dans le dictionnaire, ce qui est le cas de (1) et (2), le système nous le montre de la façon suivante :

- (1') à la mode/N0:inc<vêtements>
- (2') affriolant/N0:inc<vêtement>

Il en tire correctement la conclusion que le pronom *quelque chose* réfère à un vêtement, parce que le sujet (N0) des séquences trouvées, i.e. à la mode et affriolant est codé de <vêtements>. En traduction automatique, il traduira le verbe par son homologue de la langue cible de la façon suivante :

```
porter <vêtement>: to wear
```

Considérons maintenant le cas où le pronom est du type anaphorique, comme dans les phrases (3) et (4). Nous les rappelons ici.

- (3) Elle porte celui qui est à rayures
- (4) Il porte celle qui est (fripée + élimée)

Dans ce cas, le système cherche des candidats pour l'antécédent dans le contexte linguistique. Imaginons qu'il trouve 4 noms masculins singuliers pour candidats à *cellui* dans la phrase (3) et 4 noms féminins singuliers pour candidats à *celle* dans les phrases (4). Il procède de la même façon que celle utilisée pour le cas de pronom à usage introductif, c'est-à-dire qu'il recherche maintenant toutes les classes d'objets attribuées au verbe, en consultant le dictionnaire des opérateurs appropriés. Il trouve enfin la séquence régie par *celui* et *celle* en tant qu'opérateur approprié et nous montre le résultat de la façon suivante :

- (3') à rayures/N0:inc<vêtements>
- (4') fripé/N0:inc<vêtements> élimé/N0:inc<vêtements>

Il arrive à reconnaître le sens de chaque pronom en choisissant parmi les 4 noms candidats celui qui est codé par la classe <vêtements>.

Nous avons dit plus haut que, pour les expressions déictiques, le traitement est plus délicat parce que le travail destiné à trouver leurs référents ne peut se réaliser que par la recherche exacte et exhaustive de la situation extralinguistique. Cependant, dans ce cas aussi, les classes d'objets peuvent aider. Considérons, par exemple, la phrase (5) que nous rappelons:

(5) Elle ne veut porter que ce qui lui va le mieux

Ici, on interprétera le pronom *ce* comme une sorte de vêtements grâce à la partie régie par lui. On a en fait

(5') aller bien à /N0:inc<vêtements>

dans le dictionnaire des opérateurs appropriés de la classe d'objets <vêtements>. Considérons maintenant la phrase:

Celle-ci fait des plis

qui peut être interpétée le plus souvent comme 'La jupe est froissée'. Il n'est pas si difficile de reconnaître cette interprétation, parce qu'on trouve

faire des plis/N0:inc<vêtements>

dans le dictionnaire.

Il peut arriver que si le système ne trouve pas la séquence régie par le pronom, ce qui est le cas de (6) dont *extravagant* (ou *original*) ne figure pas dans le dictionnaire des opérateurs appropriés : en effet, dans cette phrase il y a moins de probabilité que le pronom complément d'objet soit interprété comme *vêtement* que dans les autres phrases. Dans ce cas, le système procède à la méthode de type *préférence* qui permet d'opérer un

aroitrage sur l'ensemble des candidats possibles. Par exemple, ici, on peut avoir pour interprétations du complément d'objet <vêtements>, <marque>, <arme>, <signe>, <enfant>, <nom>, etc., i.e. toutes les classes d'objets avec le verbe *porter*. Si le système a dans l'article de *porter* une liste hiérarchique qui détermine l'ordre d'application de ses classes d'objets, il peut arriver à choisir une interprétation appropriée.

5. CONCLUSION

Nous avons montré que les classes d'objets et les opérateurs appropriés permettent de lever bien des ambiguïtés concernant les pronoms. Pour finir, nous allons souligner les avantages de notre méthode sur celle des bases de connaissances.

- Notons d'abord qu'il n'y a pas à accumuler et à gérer des données de taille énorme, mis à part les données linguistiques. La méthode n'est donc pas très coûteuse.
- Celle-ci ne s'applique pas seulement à un domaine de spécialité, mais à la langue générale, alors que les bases de connaissances ne sont utilisées que pour la constitution de systèmes experts.
- À la différence de bases de connaissances qui s'appuient sur des expressions métalinguistiques, très compliquées à construire, notre méthode se fonde directement sur des expressions de la langue naturelle, ce qui facilite la tâche parce qu'elle n'exige pas un processus compliqué pour l'analyse et la génération du sens.
- Finalement, les bases de connaissances ne donnent pas toujours des résultats précis puisqu'elles se fondent sur la théorie des probabilités. Notre méthode, elle, est basée sur une analyse linguistique stricte.

Nous ne nions pas la nécessité des bases de connaissances, mais seulement nous remarquons que celles-ci doivent se limiter aux domaines de spécialité et que la méthode à base de classes d'objets a pour effet de remplacer une assez grande partie de bases de connaissances.

RÉFÉRENCES

- GROSS, Gaston (1992a): «Forme d'un dictionnaire électronique», L'environnement traductionnel. La station de travail du traducteur de l'an 2001, Sillery (Québec), Presses de l'Université du Québec et AUPELF-UREF.
- GROSS, Gaston (1992b): «Classes d'objets et enseignement», Actes du colloque sur le traitement automatique, Séoul.
- GROSS, Gaston (1994a): «Classes d'objets et synonymie», Annales littéraires de l'Université de Besançon, Série linguistique et sémiotique, vol. 23, Besançon.
- GROSS, Gaston (1994b): «Un outil pour le FLE: les classes d'objets», Actes du colloque du FLE, Lille, Presses Universitaires de Lille.

- GROSS, Gaston (1994c): «Classes d'objets et description des verbes», Langages, n° 115, Paris, Larousse.
- GROSS. Gaston (1995): «Les classes d'objets», Les concepts opératoires de l'INaLF, rapport de l'INaLF, pp. 83-88.
- GROSS, Gaston (1996): Une sémantique nouvelle pour la traduction automatique: les classes d'objets, manuscrit.
- KLEIBER, Georges (1990): «Anaphore-Deixis: deux approches concurrentes», *La Deixis*, Colloque en Sorbonne, 8-9 juin 1990, par M.-A. Morel, L. Danon-Boileau, Paris, PUF, Linguistique Nouvelle.
- KLEIBER, Georges (1991): «Anaphore-Deixis: où en sommes-nous?», L'information grammaticale, num. 51, octobre 1991, pp. 3-16.
- LE PESANT, Denis (1994): «Les compléments nominaux du verbe lire, une illustration de la notion de «classe d'objets». Langages, n° 115, Paris, Larousse.
- LE PESANT, Denis (à paraître): «Anaphores et classes d'objets».
- LEHRBERGER and BOURBEAU (1988): Machine Translation: Linguistic Characteristics of MT Systems and General Morphology of Education, Amsterdam, John Benjamins Publishing Co.
- PAK, Man-ghyu (1993): Restructuration de la complétive des verbes de cognition et des verbes de communication en français et en coréen: lexique-grammaire comparé. Thèse de doctorat, Université Nationale de Séoul.
- PAK, Man-ghyu (1997): «Traduction automatique et classes d'objets: le problème de porter un vêtement en français et en coréen», *META*, vol. 42, n° 1, mars, Montréal, Presses de l'Université de Montréal, pp. 155-167.
- PUSTEJOVSKY, J. (1993): «Type Coercion and Lexical Selection», Semantics and the Lexicon, Dordrecht, Kluwer Academic Publishers.
- SOWA, J. F. (1993): «Lexical Structures and Conceptual Structures», Semantics and the Lexicon, Dordrecht, Kluwer Academic Publishers.
- SCHANK, R. C. (1975): Conceptual Information Processing, American Elsevier.

ANALYSE DES NOMINALISATIONS DANS UN CORPUS SPÉCIALISÉ: COMPARAISON AVEC LE FONCTIONNEMENT EN CORPUS «GÉNÉRAL»

Anne CONDAMINES

Équipe de Recherche en Syntaxe et Sémantique, UMR 5610 du CNRS, Toulouse, France

INTRODUCTION

En linguistique, un changement important est en cours dans la méthodologie : l'analyse de corpus est en train d'être reconnue comme une approche digne d'intérêt. Les phénomènes langagiers ne sont plus étudiés seulement par introspection mais à travers leur manifestation réelle dans des corpus. Ce mouvement devrait permettre d'établir des liens très étroits entre la lexicologie et la terminologie qui, depuis toujours, met en oeuvre des corpus spécialisés pour construire des listes de termes et même des définitions. Cette convergence de méthodes pourrait avoir des conséquences fructueuses sur l'analyse des corpus spécialisés en terminologie, pour au moins deux raisons. D'une part, l'analyse de corpus spécialisés pourrait bénéficier des développements du point de vue de la définition de méthodes et de la systématisation qui vont avoir lieu en analyse de corpus; deux caractéristiques (méthodologie et systématicité) qui trop souvent font défaut à la terminologie, les données étant recueillies de façon intuitive. D'autre part, pourrait commencer à se mettre en oeuvre une caractérisation des corpus spécialisés, par rapport aux corpus généraux. C'est dans cet objectif d'analyse «différentielle» que se situe l'article. Le fonctionnement des nominalisations dans un corpus spécialisé sur le génie logiciel est décrit par rapport au fonctionnement de ces mêmes nominalisations, dans deux corpus de référence considérés comme généraux. Nous utilisons essentiellement une approche statistique, la plus facile à mettre en oeuvre dans un premier temps, qui nous permet de faire émerger des tendances concernant le fonctionnement des nominalisations.

1. CONTEXTE DE L'ÉTUDE

1.1 Systématisation de l'analyse de corpus en terminologie

Contrairement à la lexicologie, où l'utilisation de corpus est assez récente, surtout en France, la terminologie a toujours fait appel, dans sa pratique, à des textes sur le domaine à l'étude. Sur un domaine spécialisé en effet, il est impossible de faire confiance à son intuition linguistique pour mettre au jour des régularités de fonctionnement permettant d'élaborer un système. Au moment où la linguistique commence à accorder ses lettres de noblesse à l'analyse de corpus, on aurait pu espérer que ce nouveau type d'approche aurait pu bénéficier des acquis de la terminologie en matière d'analyse de corpus. Malheureusement, il n'en est rien; trop souvent en effet, les corpus en terminologie ne

sont utilisés que pour attester d'occurrences, plus rarement pour justifier du choix des termes et encore plus rarement encore pour justifier du choix des relations conceptuelles; les critères qui président à ces choix relèvent encore bien trop souvent de l'intuition.

L'analyse de corpus (entendue au sens d'analyse à l'aide d'outils) pourrait être à l'origine d'une mutation majeure en linguistique ainsi que le pense C. Blanche-Benveniste dans un récent article :

«les concordanciers établis sur des corpus informatisés, en permettant de totaliser les emplois des mots sur des millions de cas, changent totalement la nature de l'analyse distributionnelle. Pour M. Hallyday (1991), les changements sont si considérables qu'on ne peut plus maintenir les grandes oppositions méthodologiques hérités de Saussure» (1997 : 32).

La prise de conscience de l'importance de l'analyse de corpus en linguistique accentue encore la nécessité d'inclure la terminologie dans la linguistique; si les corpus sont considérés comme matériau de base pour l'étude de la «langue générale» alors, il n'y a plus de différence majeure entre lexicologie et terminologie et il faut s'employer à identifier ces convergences méthodologiques. Qui dit corpus dit situation d'énonciation et cette situation peut fort bien correspondre à une activité mettant en oeuvre des connaissances expertes; on est alors exactement dans le type de contexte dans lequel sont réalisées les études en terminologie et en langue spécialisée. Il suffit alors de caractériser le lexique propre au corpus à l'étude, lui-même constitué de façon qu'il soit représentatif d'un discours spécialisé¹, en relation avec une connaissance particulière (voire, une activité particulière); c'est bien ce que soutient Biber (1996 : 171-197), un des chefs de file de l'analyse de corpus.

«In descriptive lexicography, which is concerned with the actual use of words, new meanings are discovered only by examining the use of a word in actual discourse contexts [... In addition, a use perspective is required to investigate the stylistic preferences of individuals, the differing linguistic preferences of groups of speakers ² and the ways in which "registers" (or "genres") favor some words and structures over others.»

Notre travail se situe dans ce courant qui vise à développer l'analyse de corpus pour la terminologie.

1.2 Mise en place d'une analyse différentielle

Pour mener à bien cette analyse de corpus spécialisés, nous mettons en place une analyse différentielle afin de systématiser l'intuition qui nous semble guider le terminologue, celle qu'un terme est un mot ou un groupe de mots qui fonctionne de manière inattendue, en tout cas d'une manière qui est inconnue du terminologue, non spécialiste du domaine et de la langue spécialisée qui y est à l'oeuvre. Dans ce cas d'ailleurs, plutôt que de «mémoire des mots», qui est évoqué dans le titre du colloque, on

¹ Dans l'article intitulé «Langue de spécialité ou discours spécialisé ?» (1997), nous discutons des différences de contenu entre ces deux appellations.

² C'est nous qui soulignons.

pourrait parler de mémoire du terminologue! Tout comme le linguiste, le terminologue n'est autre qu'un locuteur qui, par introspection, convoque la mémoire qu'il a des phénomènes langagiers et essaie de retrouver, soit des régularités, soit des irrégularités, pour systématiser les phénomènes langagiers.

On retrouve l'idée de déviance dans différents types de travaux sur la langue. Par exemple, l'étude de la métaphore se fait souvent à partir de cette idée de dysfonctionnement :

«Il est généralement admis que la métaphore repose crucialement sur une 'incorrection' ou un 'délit', mais ce trait, une fois dénommé et intégré, donne rarement lieu à une caractérisation satisfaisante» (Kleiber 1994 : 177-206).

Cette même idée de «déviance» a été exploitée dans le projet Eurotra, pour représenter la terminologie par rapport au lexique général, déjà traité. Les sous-langages sont caractérisés de la façon suivante :

- «- the restrictive mode : by excluding certain features of GL (general language), SL (Sublanguage) can be described as a restricted form of language;
- the deviant mode: SL can show specific features which are not found in GL and, therefore can be considered a deviant form of GL;
- the preferential mode: this approach of SL phenomena is complementary to the restrictive and deviant modes, and is expressed in terms of preferences» (Mc Naught *et al.* 1991: 1-7).

On trouve enfin cette idée de déviance dans les études sur la néologie.

«Si un changement intervient dans une règle, il se produit au niveau de la performance, sous la forme d'une déviation, d'une 'faute', et sa transformation en règle nouvelle implique un usage répété, une longue évolution [...] le changement des règles grammaticales échappe à la création consciente. Aucun locuteur, en effet, n'a un comportement linguistique naturel qui le conduit à faire volontairement des fautes [...] Les déviations qui, accumulées, constituent l'usage nouveau échappent à sa volonté, mais créent la règle nouvelle.» (Guilbert 1975 : 29)

L'étude que nous présentons ici nous permet de mettre en oeuvre sur une grande échelle, cette idée de déviance. En effet, elle est faite à partir d'un corpus spécialisé, étudié par rapport à un corpus «général».

1.3 Analyse des nominalisations

Nous avons choisi de faire porter notre attention sur les nominalisations. En effet, d'une part, en tant que noms, elles sont, selon l'approche «traditionnelle» qui privilégie cette forme, considérées comme pouvant être des termes et d'autre part, comme elles sont sémantiquement apparentées à un verbe, leur étude permet une ouverture sur l'étude des actions et événements qui sont trop souvent méconsidérés au profit des seuls objets du domaine. Remarquons, avec plusieurs auteurs, que le nom n'est pas la forme la plus

naturelle d'expression d'une action (on pourrait, à nouveau ici utiliser la notion de «déviance»):

«Il existe aussi des cas de discordance : la nominalisation en est l'illustration. Dans le cas du nom déverbal par exemple, l'action ne trouve pas dans le verbe son expression grammaticale privilégiée mais rencontre une catégorie qui lui est 'étrangère', le nom, et va lui imposer sa combinatoire morphosyntaxique spécifique...» (Rémi-Giraud 1996 : 109).

«La nominalisation — exemplairement le nom d'action — en est l'illustration : par cette métaphore grammaticale, l'action, dégagée de ses attaches référentielles et de l'emprise du temps se trouve mise sur le même plan que les substances autonomes et 'éternelles'» (*ibidem* : 115).

«Comme le nom est réputé représenter un objet, la nominalisation est fort utilisée pour créer un effet d'objectivation : c'est pourquoi elle est massivement attestée dans les textes scientifiques (notamment positivistes) et dans les discours qui les imitent (langue de bois)» (Rastier 1996 : 51).

2. ÉTUDE STATISTIQUE

Les résultats que nous présentons concernent essentiellement des données statistiques. Les études syntaxique et sémantique viendront dans un second temps; leur mise en place sur une grande échelle demande des préparations beaucoup plus lourdes, en particulier, il faudrait au moins disposer d'un corpus automatiquement étiqueté et avec un étiquetage désambiguisé. Précisons cependant que l'analyse des données statistiques donne une tendance qui permet d'orienter l'analyse sémantique et syntaxique.

2.1 Mise en place de l'étude

2.1.1 Contexte pratique

Le corpus spécialisé qui est utilisé sert de support à un travail mené dans le cadre du GIS (Groupement d'Intérêt Scientifique) Sciences de la Cognition en collaboration avec EDF (Électricité de France), l'IRIT (Institut de Recherche en Informatique de Toulouse) et le LLI (Laboratoire de Linguistique et Informatique). Ce projet vise à affiner un modèle de base de connaissances terminologique (BCT), à définir une méthode assistée par ordinateur de constitution de données terminologiques et à évaluer les possibilités d'utilisation de cette BCT pour constituer un système de consultation de textes. Bien qu'elle ait un lien direct avec ce projet, l'étude qui est présentée ici vise plutôt une réflexion théorique sur le fonctionnement des langues spécialisées. Le corpus, le guide MOUGLIS (Méthodes et OUtils de Génie Logiciel pour l'Informatique Scientifique), est constitué par un ensemble de documents méthodologiques sur l'organisation de projets et les techniques de génie logiciel appliquées au développement de logiciels scientifiques à la Direction des Études et Recherches d'EDF. Mouglis est à la fois un guide qui précise comment réaliser les différentes phases de la conception d'un logiciel, et un modèle de rédaction des documents qui accompagnent chacune de ces phases. Il comporte environ 58 000 mots.

2.1.2 Les données étudiées

La première difficulté vient de la sélection des nominalisations déverbales, qui sont celles qui nous intéressent plus particulièrement. Il est très difficile de définir des critères pour choisir les nominalisations parmi la liste des noms repérés dans le corpus par NOMINO³. Il faudrait combiner à la fois des critères morphologique et des critères sémantiques. Nous avons privilégié, dans un premier temps, une approche assez intuitive qui correspond à la définition donnée par Defranc et Willems : «sous le terme de nom déverbal ou 'nominalisation', nous entendons les formes morphologiquement apparentées aux verbes et ayant en commun avec lui le sème de 'procès'» (1996 : 221).

Parmi les 1151 formes nominales distinctes proposées par Nomino, nous avons retenu 292 nominalisations, soit à peu près le quart. Parmi ces 292 nominalisations, nous n'avons conservé que les 185 formes qui apparaissent au moins trois fois dans le corpus.

2.1.3 Deux corpus de référence

Deux corpus, que nous avons considérés comme «généraux» ont été utilisés comme références : le corpus du TLF (XIX^e et XX^e siècle) et le corpus «Le Monde» sur CD-Rom. Bien qu'on puisse toujours mettre en question l'existence d'un corpus «général», c'est-à-dire neutre du point de vue du domaine et du point de vue du style, on peut penser d'une part que la quantité des données permet d'aboutir à une certaine stabilisation des variations et, d'autre part, que le fait d'utiliser deux corpus permet de neutraliser des variations de style. Le corpus du TLF a été mis en oeuvre pour la première caractérisation, le second pour les deux suivantes.

Pour ce qui concerne le TLF, nous l'avons mis en oeuvre via le logiciel HYPERBASE (logiciel construit par É. Brunet (Brunet 1995)) qui propose une fonctionnalité rendant compte de la spécificité du vocabulaire du corpus à l'étude par rapport au vocabulaire du corpus du TLF.

2.1.4 Trois caractéristiques examinées

Trois caractéristiques concernant le fonctionnement statistique des nominalisations sont examinées.

- Dénombrement d'occurrences

Il s'agit d'étudier le nombre d'occurrences de chaque nominalisation du corpus spécialisé par rapport au corpus de référence.

Hyperbase fournit une caractérisation statistique, directement utilisable. Trois classes de nominalisations ont été ainsi définies à partir des trois caractérisations proposées par Hyperbase :

³ Outil d'analyse de corpus particulièrement destiné à la terminologie, réalisé par le centre ATO, à la demande du RINT.

- lexèmes en excédent,
- lexèmes en déficit.
- lexèmes absents du modèle.

Les deux premières sont obtenues à partir de «la valeur absolue de l'écart réduit afin de mettre en relief ce qui est le plus significatif»; le troisième concerne les lexèmes absents du du fichier REFER qui contient les 10 000 formes les plus fréquentes du TLF (XIX^e et XX^e siècles) c'est-à-dire celles dont la fréquence dépasse 500.

Nous avons considéré une quatrième classe composée des nominalisations qui n'ont aucune spécificité, c'est-à-dire qui fonctionnent comme dans le corpus général.

- Proportion de formes nominales par rapport aux formes verbales

Le corpus du Monde nous a permis d'examiner des critères plus complexes.

Le premier critère concerne la proportion de formes verbales par rapport à la forme nominalisée (par exemple, formes verbales de *gérer* par rapport à *gestion*) et ce dans une perspective de comparaison du fonctionnement dans notre corpus par rapport au corpus de référence. Ce premier élément a pour but de vérifier l'hypothèse d'une utilisation de la nominalisation plus élevée dans les corpus spécialisés que dans les corpus généraux.

Pour chaque nominalisation, nous avons calculé le chi2 qui met en oeuvre quatre chiffres : nombre de nominalisations dans le corpus, nombre de nominalisations dans *Le Monde*, nombre de formes verbales dans le corpus, nombre de formes verbales dans *Le Monde*. Le chi2 est «signé», c'est-à-dire précédé d'un (-) lorsqu'apparaît un déficit de formes nominales par rapport aux formes verbales, dans le corpus.

Une des difficultés que nous avons rencontrées dans le dénombrement est celle des ambiguïtés de formes. En effet, beaucoup de formes de verbes conjugués peuvent être identiques à des formes d'autres catégories (par exemple, acquis peut être soit une forme verbale (participe passé ou présent à la deuxième personne ou une forme nominale au pluriel (les acquis)). Dans le corpus spécialisé, le nombre d'occurrences est suffisamment faible pour que la vérification systématique des catégories soit possible. En revanche, dans le corpus du Monde où il y a parfois des milliers d'occurrences, la vérification est impossible et nous avons fait le choix de ne pas examiner cette caractéristique lorsque des ambiguïtés subsistaient ce qui est encore plus fréquent que pour le corpus spécialisé, dans la mesure où l'interface de recherche du CD-Rom «Le Monde» ne tient pas compte des accents (ainsi, la forme rencontres correspond soit au nom au pluriel, soit au participe passé du verbe rencontrer soit au présent 2^e personne).

Pour le repérage de ces formes verbales, seules les formes correspondant aux temps et aux personnes utilisés dans le corpus spécialisé ont été retenues, soit les 3^{es} personnes du singulier et du pluriel à l'infinitif, au présent et au futur de l'indicatif, le participe passé et le participe présent. En effet, nous avons considéré que le style même du document (un

manuel) plutôt que le domaine imposait ces temps et personnes et qu'il était préférable de comparer les formes verbales sur des bases similaires⁴.

- Proportion de formes nominales au singulier par rapport aux formes nominales au pluriel

Ce critère a pour but de mettre au jour un nombre éventuellement «anormal» de formes nominales au pluriel ou au singulier. Nous voulions tester en particulier l'hypotèse que l'utilisation accrue des nominalisations en corpus spécialisé soit due au fait que l'une des valeurs sémantiques de la nominalisation soit plus fréquente.

On considère généralement qu'une nominalisation peut avoir (au moins) deux valeurs sémantiques : elle peut renvoyer soit au procès, soit au résultat (état ou objet tangible) du procès. Une nominalisation au pluriel renvoie le plus souvent à la deuxième valeur (résultat, objet tangible); ainsi, *les achats, les constructions* s'interprètent avec cette deuxième valeur et s'intègrent très mal dans des constructions qui contraignent l'interprétation «processive» :

Ainsi un nombre significativement élevé de nominalisations au pluriel pouvait correspondre à un nombre important de concepts «objets créés» et contrevenir l'idée que l'utilisation des nominalisations viendrait en lieu et place de l'utilisation des verbes.

2.2 Résultats de l'étude

- Fréquence de nominalisations par rapport au corpus de référence

Les résultats obtenus grâce à Hyperbase sont les suivants :

80 nominalisations hors modèle (non comprises dans la liste des formes les plus fréquentes);

74 nominalisations en excédent (par rapport au corpus TLF);

31 nominalisations conformes au modèle;

0 nominalisation en déficit (par rapport au corpus du TLF).

Ces premiers résultats mettent en évidence, d'emblée, une utilisation importante de la nominalisation dans le corpus spécialisé. En effet, 154 nominalisations (80 + 74) apparaissent comme plus fréquentes que dans le corpus de référence, soit plus de 83 %. En revanche, aucune nominalisation n'apparaît en déficit. Cette première évaluation statistique confirme donc, de façon très nette, l'hypothèse d'une utilisation plus fréquente des nominalisations en corpus spécialisé.

Une remarque importante doit cependant être faite. Elle consiste en une mise en garde d'Étienne Brunet lui-même : ces résultats sont à prendre avec des réserves. On peut

^{*?}pendant les achats,

^{*}pendant les constructions.

⁴ L'étude sur toute les formes verbales, que nous avons aussi faite n'entraîne que très peu de variations.

s'étonner par exemple que ni *listage*, ni *interfaçage* ne soient caractérisés comme plus fréquents dans le corpus spécialisé.

- Nominalisations par rapport aux formes verbales

L'examen des résultats obtenus (avec un autre corpus de référence, je le rappelle) met en évidence une convergence très nette entre les résultats obtenus ici et ceux obtenus avec le critère précédent.

Si l'on considère les résultats correspondant à un chi2 avec un seuil de 0.01 (la probabilité de l'hypothèse nulle est inférieure à 6,63), on constate les éléments suivants :

nominalisations du premier groupe («hors modèle») :

- 27 résultats sur 47 sont supérieurs à 6,63, c'est-à-dire que dans 27 cas sur 47 examinés, le nombre de formes nominales est, de façon significative, supérieur au nombre de formes verbales, par rapport au corpus de référence
- le chi2 moyen est de 80,93

nominalisations du deuxième groupe :

- 20 résultats sur 35 sont supérieurs à 6,63
- le chi2 moyen est de 35,3

nominalisations du troisième groupe :

- 5 résultats sur 18 sont supérieurs à 6,63
- le chi2 moyen est de 4,86 (soit inférieur à celui que nous avons considéré comme significatif : 6,63).

On peut donc dire que, dans l'ensemble, les nominalisations qui n'ont pas une fréquence spécifique dans le corpus ne sont pas utilisées de façon plus élevée que les formes verbales correspondantes. Inversement, pour les deux groupes de nominalisations qui ont une fréquence élévée par rapport au corpus de référence, on constate une utilisation de ces nominalisations bien plus grande que les formes verbales correspondantes. Il y a bien convergence de résultats pour ces deux premières caractérisations : une forme nominale qui est utilisée plus fréquemment dans le corpus spécialisé que dans le corpus général est aussi utilisée plus fréquemment que la forme verbale, toujours par rapport au corpus de référence.

- Nominalisations au pluriel par rapport aux nominalisations au singulier

Dans les trois groupes de verbes, peu de résultats sont supérieurs à 6,63 ou inférieurs à (-)6,63, c'est-à-dire que peu de nominalisations ont une utilisation au pluriel beaucoup plus ou beaucoup moins élevée que dans le corpus de référence :

- 16/60 dans le premier groupe,
- 13/57 dans le deuxième groupe,
- 8/25 dans le troisième groupe.

Soit, au total 37 résultats «anormaux» sur 142 (un peu plus du quart).

Cela signifie que peu de nominalisations ont un comportement déviant, du point de vue du pluriel, par rapport au corpus de référence. L'examen des moyennes de chi2 confirme ces résultats. En effet, dans aucun des trois groupes la moyenne des chi2 est inférieure à (-) 6,63 ou supérieure à 6,63, c'est-à-dire une moyenne qui mettrait en évidence un fonctionnement spécifique :

- groupe 1 : moyenne des chi2 : (-)3,43 - groupe 2 : 0,33 - groupe 3 : 0,67

Notons cependant que, lorsqu'il y a fonctionnement déviant, il va plutôt dans le sens d'une augmentation des formes plurielles (résultat inférieur à (-)6,63) (21 cas sur 37).

Au terme de l'analyse des résultats des trois critères, nous pouvons faire les constatations suivantes. Le corpus spécialisé que nous avons étudié (et sans doute peut-on élargir cette constatation à l'ensemble des corpus spécialisés mais cela reste à vérifier) privilégie les nominalisations par rapport aux formes verbales mais ces nominalisations ne sont pas plus souvent au pluriel que dans les corpus généraux, ce qui tend à montrer que les nominalisations ne sont pas utilisées pour évoquer des objets mais bien en lieu et place des formes verbales.

2.3 Déviance statistique et mise en évidence de termes

La question qui se pose, à l'issue de cette analyse statistique est celle de savoir si on pourrait utiliser, de façon systématique, ce type d'approche (la mise en oeuvre des trois critères) pour repérer les nominalisations qui sont des termes.

Plusieurs raisons nous amènent à penser que cela ne serait pas suffisant pour repérer l'ensemble des termes comportant une nominalisation.

Tout d'abord, il est bien connu que les termes sont, le plus souvent, des syntagmes et non des mots isolés. En conséquence, ce qui pourrait être mis en évidence avec les trois critères, c'est que telle ou telle nominalisation fait partie d'un terme (ou de plusieurs); reste à savoir lequel (ou lesquels) d'autant qu'une nominalisation peut se trouver soit comme tête (ex : développement logiciel en version définitive), soit comme expansion (ex : cycle de développement).

Ensuite, la mise en oeuvre systématique de ces tests supposent que toutes les données soient disponibles. Or, nous avons vu que pour un certain nombre de cas, les formes verbales peuvent être confondues avec d'autres formes, ce qui empêche la réalisation du test. Pour remédier à cet inconvénient, il faudrait pouvoir disposer de corpus étiquetés sans erreur, ce qui, si on en croit l'article de Habbert et Salem (1995), est loin d'être le cas.

Enfin, nous y reviendrons, l'analyse statistique, seule, ne suffit pas à épuiser les possibilités de repérage de déviance.

Nous ne pensons pas cependant que ces réserves, importantes, ruinent la possibilité d'utiliser les résultats statistiques pour repérer les termes. Il nous faut tout de même mettre en évidence la pertinence des résultats que nous avons obtenus.

Nous avons considéré les nominalisations qui obéissent aux trois critères étudiés : fréquence anormalement élevée dans le corpus, formes nominales significativement plus élevées dans le corpus que dans le corpus de référence, formes nominales au pluriel significativement plus élevées dans le corpus que dans le corpus de référence. Ce dernier critère qui n'est pas caractéristique des corpus spécialisés en général devient pertinent lorsqu'on considère les nominalisations une par une. Rappelons que beaucoup de nominalisations n'ont pas pu subir tous les tests pour des problèmes de catégorisation.

Quinze nominalisations possèdent ces trois caractéristiques de fonctionnement. Treize de ces nominalisations apparaissent dans certains des termes qui ont été retenus à la fin de l'analyse terminologique; ce sont : configuration, évolution, fourniture, spécification, test, vérification, conception, réalisation, organisation, installation, réunion, approbation, demande, certains étant particulièrement productifs : gestion apparaît dans 43 termes, test dans 55, conception dans 39, spécification dans 18...

Les deux nominalisations qui n'apparaissent dans aucun terme retenu sont consolidation et décision. Consolidation n'a que trois occurrences, ce qui rend les tests statistiques assez peu fiables. Reste décision dont rien, dans l'examen des occurrences, laisse penser qu'on pourrait l'intégrer dans un terme.

Ces résultats, somme toute assez pertinents, tendent à prouver que l'approche statistique, à défaut de donner des résultats fermes et définitifs, permet de mettre sur la voie d'éléments terminologiques.

3. AUTRES CRITÈRES DE DÉVIANCE : ÉLÉMENTS SYNTAXIQUES ET SÉMANTIQUES

De nombreux phénomènes, syntaxiques ou sémantiques, pourraient être examinés afin de caractériser le fonctionnement des nominalisations dans un corpus spécialisé. Nous en proposons quelques-uns.

- Une valeur sémantique est préférée

Nous avons déjà rappelé la possibilité, pour une nominalisation, de présenter une polysémie (processus, résultat statif, résultat sous forme d'objet). Il arrive que la langue spécialisée privilégie une de ces valeurs, différente de celle de la langue générale. Par exemple *spécification*, qui, dans la langue générale, renvoie au fait de spécifier, renvoie couramment, dans notre corpus, au contenu d'un document, voire, à ce document même comme le montrent les extraits suivants :

«...non conformité d'un produit ou d'une documentation avec les spécifications du logiciel...»

«...Les spécifications du logiciel définissent en particulier :...»

- Un type de complément est préféré

On trouve souvent évoquée l'ambiguïté de syntagmes comme «le dessin de Paul» qui suppose soit un Paul dessinateur, soit un Paul modèle pour un dessin. Il se peut qu'un

corpus spécialisé, voire la langue spécialisée en général, privilégie certains types de compléments plutôt que d'autres. Ainsi, notre corpus spécialisé fait apparaître très peu de compléments humains pour des nominalisations qui, en théorie du moins, devraient accepter des compléments humains (la spécification de Paul, le test de Paul...). Là encore, il faut se garder de se contenter de son intuition sur les potentialités de tel ou tel lexème et privilégier la comparaison avec un corpus de référence afin d'évaluer les différences du corpus spécialisé : analyse très difficile à réaliser automatiquement.

- Le complément peut être élidé

Nous avons déjà évoqué (Condamines 1995 : 219-238) la possibilité pour un complément d'objet direct, «normalement» obligatoire, de disparaître dans des contextes d'énonciation bien déterminés, ce qui manifestait, pour le verbe, un fonctionnement terminologique. La possibilité d'emploi absolu du verbe serait ainsi l'indice d'une référence implicite car suffisamment partagée par l'ensemble des locuteurs. Ainsi l'énoncé suivant, dans le contexte bancaire : «vous pouvez déposer librement sur votre compte» permet de penser que «déposer» est un terme. Bien qu'elle ne travaille pas sur des langues spécialisées, Michèle Noailly fait à peu de choses près le même genre de constat :

«Par 'emploi absolu' d'un verbe, on entend des emplois où le complément du verbe transitif, direct ou indirect, est absent, sans que cela implique que le verbe en question ait globalement changé de sens [...] cet objet, s'il est nécessairement existant dans l'univers de référence, est linguistiquement considéré comme sans pertinence». (Noailly 1996 : 73-90).

Pour les nominalisations, qui sont beaucoup plus tolérantes aux emplois sans compléments, cette étude est assez difficile à mettre en oeuvre, d'autant qu'elle doit tenir compte des valeurs sémantiques. En tout cas, elle nécessite une comparaison avec le fonctionnement en corpus général avec toujours une difficulté pour mettre en place l'expérimentation.

Même si ces tests ne peuvent être effectués dans un premier temps, leur pertinence peut du moins être évaluée à la main sur quelques exemples. Cette évaluation sera notre prochaine étape, le choix des nominalisations à étudier prendra bien sûr en compte les résultats de l'analyse statistique.

CONCLUSION

L'étude statistique et différentielle des nominalisations d'un corpus spécialisé nous a permis de montrer, grâce à deux tests, la validité d'une hypothèse souvent évoquée pour les corpus spécialisés, celle d'une utilisation plus importante de ce type de formes dans ce type de corpus. Un troisième test nous a permis de montrer que les nominalisations étaient probablement utilisées à la place des formes verbales. Nous avons pu ainsi mettre en oeuvre sur une grande échelle une approche différentielle puisque tous les tests sur le corpus ont été effectués par rapport à des corpus généraux.

En pousuivant dans cette voie de l'analyse différentielle (par rapport à un corpus général, servant de référence), nous espérons arriver à mettre au jour un ensemble de fonctionnements statistiques mais aussi syntaxiques et sémantiques, (malgré la difficulté

pour mettre au point les expérimentations pour ces deux dernières caractérisations) propres aux corpus spécialisés. On peut sans doute espérer que cette caractérisation permettra d'aider au repérage de termes, en rendant un peu plus systématique l'approche intuitive qui sous-tend la construction de terminologies.

RÉFÉRENCES

- BESZTERDA, J. et J. SYPNICKI (1986): «Les noms d'action en tant que termes», *META*, vol. 32, n° 3, pp. 255-259.
- BIBER, D. (1996): «Investigating language use through corpus-based analyses of association patterns», *International Journal of Corpus Linguistics*, vol. 1(2), pp. 171-197.
- BLANCHE-BENVENISTE, C. (1996): «De l'utilité du corpus linguistique», Revue française de linguistique appliquée, 1-2, pp. 25-42.
- BRUNET, É. (1996): «L'hypertexte HYPERBASE», Lexicomatique et dictionnairiques, Actes des IV^{es} Journées scientifiques du réseau Lexicologie, Terminologie, Traduction de l'AUPELF-UREF, Lyon, 1995, pp. 11-30.
- CONDAMINES, A. (1995): «Terminology: new needs, new perspectives», *Terminology*, vol. 2.2, pp. 219-238.
- CONDAMINES, A. (1997): «Langue spécialisée ou discours spécialisé?», Mélanges de linguistiques offerts à Rostislav Kocourek, Halifax, Presses d'ALFA, pp. 171-184.
- DEFRANC, B. et D. WILLEMS (1996): «De l'abstrait au concret. Une réflexion sur la polysémie des noms déverbaux», Les noms abstraits, histoire et théorie, Presses Universitaires du Septentrion, pp. 221-223.
- GUILBERT, L. (1975): La créativité lexicale, Paris, Larousse, collection «langue et langage».
- HABERT, B. et A. SALEM (1995): «L'utilisation de catégories multiples pour l'analyse quantitative de données textuelles», *TAL* (Traitement Automatique des Langues), 36-1/2, pp. 249-275.
- KLEIBER, G. (1994): «Métaphore et déviance: banalisation ou contrainte hiérarchique», Nominales, essai de sémantique référentielle, Paris, Armand Colin, pp. 177-206.
- LEBART, L. et A. SALEM (1994): Statistique textuelle, Paris, Dunod.
- NOAILLY, M. (1996): «Le vide des choses», Cahiers de Praxématique, n° 27, pp. 73-90.
- RASTIER, F. (1995): «Le terme entre ontologie et linguistique», La Banque des mots, numéro spécial 7, actes des premières rencontres «Terminologie et Intelligence Artificielle», pp. 35-64.
- RÉMI-GIRAUD, Sylviane (1996): «Pour une approche notionnelle de la nominalisation», Les noms abstraits. Histoire et théorie, Presses Universitaires du Septentrion, pp. 105-116.
- * Je remercie vivement Max Reinert, statisticien à l'ERSS, qui m'a guidée dans la réalisation de l'analyse statistique.

Annexe : tableau des résultats de l'analyse statistique

Légende: NC: nominalisations dans le corpus

NM: Nominalisations dans Le Monde

VC : verbes dans le corpus VM : verbes dans Le Monde

NsC: Nominalisations au singulier dans le corpus NsM: Nominalisations au singulier dans Le Monde NpC: Nominalisations au pluriel dans le corpus NpM: Nominalisations au pluriel dans Le monde

	NC	NM	VC	VM	chi2 N/V	NsC	NsM	NpC	NpM	chi2 Ns/Np
1er groupe										
achèvement	3	232	15	3413	3,16	3	231	0	1	1,29
adaptation	5	1610	10	3030	(-)1,22	1	1432	4	178	ille le si le s
affinage	4	9	1	268	地開闢6579	4	9	0	0	
ajout	4	87	5	10646	208.82	4	47	0	40	3.28
appartenance	3	662	2	1404	1,78	3	628	0	34	0 16
approvisionnemen	4	364	0	282	3,08	4	307	0	57	0,74
archivage						10	26	0	1	0 38
codage						43	157	0	3	0,81
codification	9	68	0	138	16,83	8	64	1	4	(-10,35
commande									_	
compilation	7	181	. 1,	4 1	0,18	7	128	0	53	2,85
complement						5	380	0	87	1,14
concertation						5	1069	0	32	0,14
configuration	510	255	2	7	7.84	505	226	5	29	48,23
consolidation	3	194	0	1097	16.74	3	189	0	5	7.93
contrôle										
coordination						16	1084	0	32	0,47
couplage						5	31	0	1	- (-)21/4
coûts	12	4004	0	2674						
declenchement	4	274	2	1667	- 103.49	4	264	0	8	0,12
decoupage						19	200	0	_ 26	2,44
definition	69	1512	101	2466	0,45	69	1413	0	99	4,81
demarrage	10	264	0	581		10	258	0	6	0.23
dépendance	5	578	4	1803	4,75	0	535	5	43	(2)(6)
déroulement	6	539	9	2166	3,75	6	525	0	13	0,14
développement	. 381	1274	15	5705	1308,46	374	8748	7	400	5.74
diffusion	19	4611	2	4533	(1886) 18	19	2812	0	50	0,33
documentation	90	486	19	171	3,7	90	483	0	3	0,55
élaboration	25	805	15	1651		25	798	0	7	0.21
enregistrement	10	1005	5	4523		3	659	7	349	(-)5,44
évaluation	6	1338	. 7	1144	(-)0,31	6	1212	0	126	0,62

A. Condamines

evolution	60	5538	1	1991	8.76	35	4856	25	682	(-)46,34
formalisation	4	16	1.4	114	1,32	4	16	0	0	
foundure.	37	361	45	2732	81,52	9	286	28	75	(-)52,78
gestion						322	6519	С	27	1.33
identification	37	399	85	1742	10,14	37	389	Ü	10	0 94
illustration						16	1169	6	332	()0 33
incorporation	6	77	0	196	14,48	6	76	0	1	7.88
incrementation	4	e.	4	9		.:	0.	0	0	
inspection	10	834	0	108	1 29	10	773	0	61	0 78
integration						121	2271	0	16	0.85
interaction	7	142	2	18	0.88	1	81	6	61	(-)4 92
livraison	-+					32	576	18	500	2,1
maillage	- 			_		3	88	0	-1	0.13
maintenance	-+					34	318	0	0	1,52
		-				9	930	0	1	······································
management	27	45		33	14,19	27	40	0	5	3.22
medelisation	3	a	9					_	-	3,22
nommage	_			6188	1547,74	3	0	0	0	
r-orma-sation	8	640	2.	186	3 34	8	640	Ü	O	
optention	7	304	- G	9429	85,37		302	0	2	4 63
parametrage					_	4	- 4	0	0	
phasage						3	2	0	0	
pilotage						5	285	0		
planification	15	450	7	231	4.2	15	456	0	-1	C 13
position	4	8652		175	(-192.84	<u>.</u> .	6668	0	1984	1 18
prestations										
programmation						4	1570	o	37	州河南 9.42
prolongement	5	318	0	1344	20,86	5	248	0	70	ļ.,,
prototypage	9	Ú	C	Ü		9	0	0	0	
qualification	5	1459	8	4197	1,08	5	1207	υ	252	1,04
recommandation	1.3	730	2	989	11,86	0	198	13	532	(-)4.8
redocumentation	4	0	()	0		4	0	0	0	
reférence						140				
regroupement	8	679	20	2413	0.7	8	490	2	189	0,3
re:ecture	12	,113	3	241	14,85	9	104	3	9	(-)3,62
reservation						10	229	0	181	油钾钾发热
sauvegarde										
schematisation	3	4	0	14	7	3	3	0	1	0.87
sequencement						4	0	0	0	
specification with	236	41	16	64	中海河127,09	167	6	69	35	#######\$93
steckage	13	331	5	289	2,49	12	329	1	2	电影机测2 8
structuration				, ,		12	44	0	0	
support	42	665	21	1510	沙型電影6.9	36	444	6	221	6.5
testate real	262	1355	13	742	£105.3		837	194	518	- Ki 11428
traitement	一十					9	2278	21		##6其68
transfert	- 1					22	1272	6	711	
utilisation	55	2081	ĉ1	8969	38,21		2037	0		
				- 2000	301-2010 71-21	L., Y.		~~~		1

validation						108	97	0	1	1.1
vectorisation	3	0	C.			3	0	0	0	
vérification	24	422	22	1438	心的"影",77	14	250	10	172	□世(c)7.25
chi2 moyen					80,33					(-)3,43
										世里 地田田
2ème groupe										
abstraction						4	397	0	38	0.38
acceptation	23	317	9	7372	347,49	23	315	0	2	0.14
acces	13	3962	0	1285						
achat	4	3524	7	2473	·-)2,27	:3	2586	ı	9.38	5,35
acquisition						2	1116	0	443	
analyse										
appel	16	9506	14	10177	U 3	7	8121	9	1385	41222
application						4.7	2871	2		÷==(*)28.8
approbation######	5	429	5	2134	₹\$ ⁶ 5/\$ 7 ,85	4	426	1		(-)20h5
approche	一									
assistance	一					5	1434	O	5	1,74
association				_		10	8414	0	3743	4 44
avancement						26	175	0	6	0.88
changement	\dashv					4	3768	9		海(4)12,62
charge							3,00		74742	7 7 7 7 1 2 1 4 2
Criticis Criticis	33	7524	1.1	8009	12,39				_	
commentaire	- 1	2264	3	2148	0,67	3	1371	1	893	0.34
communication	\dashv					2.4	515	2	55	0.11
comprehension	7	580	25	6660	+ - 8,25	7	578	- 0	2	2 42
canception	241	2113	13	2526	234,77	241	1800	0	313	41:17
construction	7	5687	11	4974	204,77	7	5185	0	502	0.67
consultation	_					1	1154	0	538	1,86
çontenu	_								-	
couverture	5	1201	15	1553	(-12.79	5	1034	0	167	0.8
décision	-					14	8938	5	2571	(~)0,17
decision (🐟 👯	20	11509	5	12458	্ল ই 10,23	15	8938	5	2571	· (-)8,14
decomposition	46	179	8	176	22.88	4 6	177	U	2	0,51
demarche						38		1		
description	119	1903	140	2261	5,88	114	1758	5	145	1,9
detail					0,00	14	1497	1	935	5.37
disposition	21	3684	1.3	6030	8.17	13	1924	8	1760	0.78
echange	-									****
êdition										
	7	244	n	888	24,91		127	2	67	(-)4,22
etude		<u> </u>	— Ť		24,01	8	4888		5616	
execution	11	1630	13	1182	(-)1,43	-	1283			
exploitation		1,00		1102	()1,43	22	2408	0	-	
extension			\vdash			6	1283		_	
	-					17	1393	_	_	
fabrication formation			\vdash			8	7962	0	_	

A. Condamines

genération	16	4051	7	498	(-)8,84	16	2731	С	1320	7.73
information					()-1-	15	6817	18	3984	(-)4,4
metallation sales	15	2213	3	6360	30.93	14	1227	1	986	- 6.66
Introduction	19	1129	12		7,69	19	1085	0	44	0 76
justification	4	404	12	3299	3,23	3	318	1		
lecture						16	2596	3	559	4,82
maîtrise						6	2287	0	85	0,22
modification	131	1697	35	3024	126,46	81	1108	50	589	(-)0,64
note		-								
oifre					<u></u>		-			
organisation	91	11436	12	9131	44,56	83	8113	8	3323	#18,04
précision	9	1590	41	9678	0,62	6	1191	3	399	(-)0,32
préoccupation	4	1621	3	1241	7.22	0	683	4	938	(-)2.9
preparation	28	1903	15		32,34	28	1858	0	45	0,67
présentation	33	1996	92		30,21	31	1939	2	57	(-)1.18
production						36	8099	1	1082	2,93
proposition	10	6089	68	12497	(-)14,03	9	2819	1	3270	1.7.66
rappel	7	808	19	7936	9,61	7	681	Ü	127	1 3
rapport	20	18317	7	2261	(-)6.14	5	15857	15	2460	(364)86
realisations all	98	2161	53	10669	239,79	98	1825	0		17,86
reception						20	515	0	55	2,12
recherche						7	8944	0.	2082	1,62
redaction	58	1683	23	1522	ر 11,56 م	58	1514	0	169	6,45
emarque										Ī
réponse	19	4790	24	5459	(-)0,11	17	3561	2	1229	2,27
reprise						6				
reunion 🔭 🛶	34	7189	1	5122	21,58	21	6169	13	1020	a(a(a)) 5.96
saisie						7		0		Ī
solution	26	5119	0	0						
suppression	3	1806	2	2235	0,47	3	1159	0	647	1.67
surveillance						5	2101	0	25	5 94
synthese	. 6	965	2	101	(-)2,2	6	899	0	66	0.44
transformation	4	1338	2	3858	5 24	4	1090	0	248	0,9
vision						14	2382	1	300	0,3
chi2 moyen					二、信仰:85,8					0,33
3ème groupe										
abreviation	18	32	0	113	45,72	1	23	17	9	(-)20,29
action						3	9986	43	4601	448130
arrêt			Ţ			2	1093	2	327	(-)1,63
assemblage						3	230	1	55	11.0
attribution	4	1420	4	2497	0,65	2	1093	2	327	(-)1,63
calcul	23	563	1	836	301	23	849	3	479	888
conduite						4	2214	0	342	0,61
connaissance	8	2916	4	8881	11.02	8	2174	0	742	
constitution	8	3282	47	8687	(-)4,56	8	3245	0	37	96.5

Analyse des nominalisations dans un corpus spécialisé : comparaison avec le fonctionnement en corpus «général»

contrainte						1	839	28	1285	(-) 15 62
convention	6	3214	7	41	(-)22,21	3	2435	3	779	
demande				_		0	-	8		, ,
dépense						0		- 4		
direction						4	11142	0	679	0,24
écriture						5	2111	0	198	0,46
engagement	4	4632	2	9751	3.26	3	2852	1	1780	0 11
essai	5	6298	0	2785	2,21	0	1800	5	4498	(-)2
existence	6	3691	7	6900	0.73	6	3619	0	72	0,11
expression	5	3349	6	5989	0,43	5	3063	0	286	0.46
fonctionnement	6	2518	0	1973	4,69	- 6	2500	0	18	4,31
instruction	4	3799	0	585	0.61	0	3377	4	422	(*)31.74
interét	4	11806	5	5117	(-)2.73	4	8292	0	3514	1,69
interfaçage	2	0	0			2	0	0	0	
intervention	4	5087	3	3128	(-)6 77	3	3965	1	1122	(-)2.01
listage						2	1	2	0	
mesure										
opération						8	6248	3	3666	0,44
personnalisation	7	45	7	256	HP 11,02	7	45	0	0	
question	1 1	17562	0	167	0.1	5	12466	6	5096	(-)3,47
signification	4	561	0	1629	小湖湖 11,59	4	512	0	49	0.38
Irava-l	17	22454	4	7867	0,51	17	22453	0	1	1.00
chi2 moyen					4,86					0.67

LA MÉMOIRE DES TERMES: ANALYSE CONCEPTUELLE

Manuel Célio CONCEIÇÃO1

UCEH - Universidade do Algarve, Faro, Portugal

L' intérêt manifesté par les terminologues à l'égard des concepts et de la conceptologie a bouleversé les méthodologies de travail traditionnelles en terminologie.

Les termes peuvent être comparés à une pièce de monnaie, ayant deux faces inséparables: la face linguistique et la face cognitive. La face linguistique est un ensemble de morphèmes porteurs de signification qui servent d'étiquette au concept (face cognitive). Celui-ci est une unité d'un système, fonctionnant au niveau cognitif, et constituée par un ensemble de traits conceptuels qui le distinguent de tous les autres dans une conceptologie. Le concept n'est pas le signifié², puisque ce dernier existe au niveau linguistique; il est la manière dont on regarde le concept.

Dans le cadre des recherches en linguistique, l'analyse des dénominations (la face visible des termes) a été l'objet de la plupart des études terminologiques; les études morphosyntaxiques sont les plus fréquentes. Reconnue depuis toujours, et souvent citée, l'autre face du terme n'a commencé à être dépouillée avec soin que depuis quelques années. Il en va de même pour les relations entre les niveaux linguistique et cognitif en terminologie.

Les traits conceptuels qui caractérisent un concept sont «traduisibles» en éléments définitionnels qui constituent, à leur tour, le schéma définitionnel. Ce schéma, qui montre les relations entre les traits constitutifs de ce concept est «l'élément médian entre le terme et le concept» (Thoiron 1995b: 1). À partir de ce schéma définitionnel sont rédigées les définitions terminologiques, ou même encyclopédiques, et peut alors être créée la forme linguistique qui désignera le concept. La formation du mot³ se fera par le choix d'éléments de dénomination qui correspondent à certains (idéalement, tous) éléments définitionnels décrits dans le schéma. Quand les éléments définitionnels sont transposés au niveau linguistique, ils peuvent correspondre à un ou à plusieurs morphèmes et donc un ou plusieurs traits de signification.

¹ Mes remerciements aux professeurs Maria Teresa Lino (Universidade Nova de Lisboa) et Philippe Thoiron (Université Lumière-Lyon II) pour toutes les critiques et suggestions apportées.

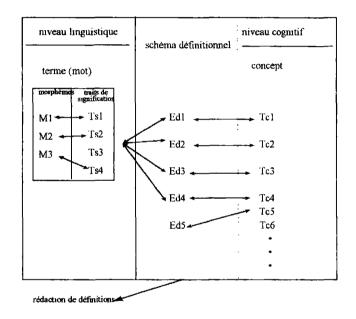
² Pour la distinction concept/signifié voir Thoiron (1995a : 118).

³ Nous entendons ici par *mot* le signe linguistique constitué par une séquence de caractères autorisée dans un système linguistique donné.

Le terme (forme linguistique) est donc un abrègement de forme et de signifié du schéma définitionnel. L'abrègement est, d'ailleurs, l'objectif de la nomination, et il peut provoquer, parfois, la démotivation et, donc, l'opacification.

Le terme (mot) est, ainsi, une synthèse linguistique du concept auquel on accède par l'analyse de ses comportements discursifs et par l'analyse morphosyntaxique de cette dénomination. Cette démarche se fait en associant aux formes des éléments de définition et, par voie de conséquence, des traits conceptuels, ou en les inférant à partir de l'ensemble des connaissances préalables. L'activation des traits conceptuels dépend de la combinaison des données linguistiques et textuelles. Les définitions rédigées seront d'autant plus complètes qu'elles seront fidèles au schéma et qu'elles rendront compte de tous les traits.

Chaque unité terminologique est donc la synthèse de la représentation présentée cidessous, adaptée de Thoiron (1995a).



M- Morphème

Ts- Trait de signification

Ed-Élément de définition

Tc - Trait conceptuel

Puisqu'une terminologie est une structure constituée des termes d'un domaine du savoir ou d'une sphère d'activité socioprofessionnelle et que la répartition de la réalité et du monde en domaines est subjective et socioculturellement déterminée, les termes sont le reflet d'une culture. Nous entendons ici culture au sens le plus large, en tant que processus dynamique dont les individus se servent.

Les concepts sont des unités de connaissance et ne sont donc pas fixes, ils sont relatifs et leur absolue universalité synchronique est discutable⁴. «Au plan sémantique [les concepts] sont des sémies stabilisées par les normes d'une discipline, et déliées de leurs

⁴ «The word "concept" is often taken as indicating a single fixed, well-delimited entity belonging to the intellectual sphere» (Langacker 1997). N'étant pas en accord avec ceci, l'auteur préfère conceptualisations à conceptions.

déterminations contextuelles» (Rastier *et alii* 1994 : 75). Les unités de connaissance sont situationnelles et relationnelles, par conséquent les progrès scientifiques et techniques ainsi que l'affinement des connaissances établissent de nouveaux traits conceptuels qui précisent les concepts. La moindre modification au niveau conceptuel entraîne forcément le changement du terme et l'évolution terminologique doit en rendre compte.

À la suite de ce qui vient d'être dit, nous nous sommes proposé d'analyser l'évolution terminologique/conceptuelle de quelques termes de pharmacologie en portugais. Sans vouloir faire des recherches diachroniques, l'objectif est uniquement de voir comment certains termes fondamentaux de ce domaine se sont «fixés» et de voir les liaisons entre traits conceptuels et éléments de nomination.

Les termes et leurs informations respectives ont été extraits de livres de pharmacologie (surtout pharmacopées) publiés depuis le XVIII^e siècle. Il faut, évidemment, analyser les textes sous l'angle de l'ordre herméneutique de la description, c'est-à-dire qu'il faut tenir compte des «différences de situation historique et culturelle qui séparent la production et l'interprétation» (Rastier *et alii* 1994 : 17). Il est indispensable de partir des textes, puisqu'ils sont un produit linguistique de l'interaction sociale dans laquelle les connaissances sont légitimées et validées.

Le XVIII^e siècle est dans ce domaine, comme dans beaucoup d'autres, le siècle par excellence du début de l'explosion terminologique. Les grandes découvertes dans les domaines connexes comme la chimie (rappelons Lavoisier) ont permis que de nouveaux termes enrichissent les pharmacopées héritées des grecs et des latins.

Contrairement au $XV\Pi^e$ siècle, caractérisé par une piètre fécondité terminologique, les grandes découvertes et les classifications destinées à mettre de l'ordre dans le naturel d'une part, l'expérimentation de plus en plus rigoureuse et ingénieuse et les progrès de tout ordre associés aux perfectionnements des techniques et des technologies d'autre part ont contribué à cette explosion terminologique. Les synonymies se sont multipliées et l'éponymie a été un phénomène très fréquent pour immortaliser les découvreurs.

Au long des siècles on peut constater une certaine pérennité dans le vocabulaire, mais on trouve également quelques modismes éphémères déterminés par toute sorte de changements, sociaux, scientifiques ou techniques.

Bien qu'il y ait fort peu d'études diachroniques en terminologie, et ce domaine ne fait pas exception, l'observation du surgissement ou de la disparition de certaines formes est tâche facile par rapport aux études sur le changement conceptuel. Ceci malgré la très grande difficulté de datation des formes (exception faite des formes éponymiques ou de certaines métaphores très stéréotypées).

Nous sommes, en outre, parti du XVIII^e siècle, parce que les premières pharmacopées écrites en langue portugaise datent de ce siècle — la première est de 1704 — et les premiers manuels de pharmacie (le terme pharmacologie n'est pas encore utilisé) datent de la fin de ce siècle.

Seuls les textes du XX^e siècle ont pu être numérisés, pour les autres la saisie a été faite manuellement. Pour chaque siècle nous avons pris un document officiel. Tous les textes ont été écrits en portugais par des spécialistes portugais et leur public est aussi spécialiste du domaine. Nous évitons ainsi les dérapages de sens liés à la vulgarisation. L'ensemble a été soumis à une lecture descriptive qui «résulte d'une description sémantique qui tient compte du sens immanent à la situation dans laquelle le texte a été produit» (Rastier *et alii* 1994 : 16).

Au début, nous avions l'intention de faire, non seulement une étude monolingue, mais aussi une analyse contrastive, prenant des textes français avec des caractéristiques éditologiques semblables. Cet objectif a été écarté au fur et à mesure de la comparaison des textes. En effet les différences ne sont pas pertinentes, étant donné que les textes sont très proches. Les textes du XVIII^e siècle ne sont souvent que des traductions et des adaptations des mêmes textes sources écrits en latin. À partir du XIX^e siècle, les textes portugais semblent s'inspirer beaucoup des textes français, ce qui témoigne de l'influence des recherches faites en France; influence reconnue par les auteurs qui ont esquissé l'histoire de la pharmacologie au Portugal.

Avec les termes ont été saisis les contextes et, si possible, les définitions présentées par les auteurs (cf. corpus de textes en annexe), vu que c'est dans le discours et dans la communication que s'opère la signification. Elle est le résultat d'une relation entre signifiés établie par la construction faite pendant l'interprétation.

Dans la construction de la signification, outre les éléments textuels considérés, il faudrait aussi voir les formules de reprise comme par exemple les procédés anaphoriques qui montrent les caractéristiques discontinues des traits significatifs (sèmes).

À partir de ces éléments textuels, l'objectif est de partir des désignations et des schémas définitionnels pour arriver aux traits conceptuels et aux concepts respectifs. Cette démarche peut permettre l'établissement d'une conceptologie.

Il s'agit donc de connaître les termes pour maîtriser les concepts. La maîtrise des concepts ne dépend pas seulement de la reconnaissance et de la compréhension des formants, d'autant plus qu'un concept peut avoir plusieurs désignations. L'arrivée au concept se complique aussi car sa connaissance n'est pas exclusivement linguistique, elle est aussi référentielle et propre aux pratiques sociales. C'est ce dernier aspect qui permet l'évolution conceptuelle associée aux événements relevant de l'histoire des sciences et des techniques.

La spécification des concepts peut se faire à partir de l'analyse du comportement discursif. Pour illustrer ce qui a été dit, voyons, à titre d'exemple, l'évolution du concept de médicament, à travers les informations extraites du corpus.

Médicament

traits conceptuels	
1. forme pharmaceutique	
2. produit altérations dans l'organisme	
2.1 guérit	
2.2 prévient	
2.3 diagnostique	
3. appliquée à l'homme et aux animaux	
définitions	traits présents dans la définition
«medicamentum est quod naturam	
alterare potest; vel : medicamentum est	2. «naturam alterare potest»
illud quod corpori humano applicatum	_
afflictiones nostras sanat; vel:	2.1 «afflictiones nostras sanat» ⁶
medicamentum est quod instrumentum,	
quod medici operantur ad extirpandos; e	? instrumentum
sanandus morbis.» (Santo António	
1704 : 1) ⁵	
«he tudo o que tomado por dentro ou	
aplicado por fora produz alterações em	2. «produz alterações»
nossos humores»(Vigier 1716 : 2)	
«agente terapêutico [], substância ou	2. «agente terapêutico»
substâncias, tiradas de algum dos três	
reinos da natureza, não essencialmente	
alimentares, mas que reduzidas a forma	~1. «forma conveniente de
conveniente de administração, são	administração»
capazes de modificar os actos do	2. «modificar actos do organismo»
organismo, tornando-se assim úteis ao	2.1. «úteis ao tratamento de doenças»
tratamento das doenças» (Gomes 1815 :	
2-3)	
«toda a substância ou composição que	~1 «substância ou composição»
possua propiedades curativas ou	2.1 «curativas»
preventivas das doenças e dos seus	2.2 «preventivas das doenças e dos seus
sintomas, do homem ou do animal, com	sintomas»
vista a estabelecer um diagnóstico	3. «do homem ou do animal»
médico ou a restaurar, corrigir ou	2.3 «estabelecer diagnóstico»
modificar as suas funções orgânicas».	2.1 «restaurar»
(Pita, 1993 : 57, d'après le Décret	2. «modificar [] funções orgânicas»
n° 72/91 du 8 février)	<u> </u>

Nous constatons que ce n'est qu'au XIX^e siècle que la nature de médicament se précise. Il commence par être «quod», puis, «tudo», «être ou substance», «agente terapêutico», et devient «forma conveniente de administração». Il est curieux de constater qu'au XVII^e siècle ce concept incluait encore les instruments utilisés par les médecins, «quod medici operantur ad extirpandos, e sanandos morbis».

⁵ Reprend totalement la définition proposée par Galien (liv.I, chap.I).

⁶ Ce trait est aussi repris par Plenck (1786 : 2) «être ou substance qui fait cesser, par sa vertu médicale, la maladie».

La précision et le besoin de transformer la substance en composition et sous la forme d'administration vont de pair avec les changements qui se sont opérés lors du passage de la pharmacie galénique à la pharmacie chirurgicale, déjà présentée par Plenck (1786 : 9) comme «l'art de préparer ou de composer des médicaments».

L'art deviendra pharmacie chimique et industrie pharmaceutique⁷. Les médicaments ont cessé d'être simplement «uma amálgama de drogas onde as de origem vegetal prevaleciam» (Pita 1993 : 205).

Les transformations d'ordre scientifique, technique, et, par conséquent, conceptuel et même social justifient aujourd'hui l'existence de trois dénominations pour chaque médicament : dénomination scientifique, dénomination commerciale et dénomination commune.

Quoiqu'il en soit, le trait conceptuel que l'on infère à partir de la plupart de ces désignations et de ce terme (surtout) est traitement, associé au pouvoir de provoquer des altérations. Ce trait était déjà présent dans l'étymon latin MEDICOR (soigner, traiter) + MENT. Ce terme semble donc avoir une mémoire, même si sa référence a beaucoup changé⁸.

L'analyse conceptuelle facilitera l'identification du concept et sa distinction par rapport à ceux qui lui sont proches. La rédaction des définitions sera aussi facilitée. Dans ce cas, il faudrait analyser les concepts de remède⁹ et de drogue¹⁰, puisqu'ils sont souvent confondus avec médicament¹¹. Un autre terme parfois considéré à tort comme synonyme

^{7 «}O grande comércio de drogas do Renascimento que se viu continuado pelo séc. XVII deu lugar ao grande comércio das especialidades farmacêuticas, fruto da laboração das gigantescas indústrias muitas vezes não apenas destinadas ao medicamento mas igualmente a outras finalidades, como por exemplo as indústrias químicas.» Pita (1993: 205) [Le grand commerce de drogues de la Renaissance continué au XVIIe siècle a cédé la place au grand commerce des spécialités pharmaceutiques, fruit du travail de gigantesques industries souvent destinées non seulement au médicament mais aussi à d'autres finalités, comme par exemple les industries chimiques.]

^{8 &}quot;Curative properties have been attributed to nearly every substance when it was new, unusual, or difficult to obtain and which can be forced into the human system. When the novelty of these substances wear off they cease to be used as medicaments». (Haggard 1946: 388). Comme exemple, citons les pommes de terre qui, lors de leur arrivée en Europe, ont été considérées comme un médicament aphrodisiaque.

⁹ Remède (du latin Res Medeor) est un hyperonyme de médicament, ce n'est pas obligatoirement une forme pharmaceutique, c'est l'ensemble des moyens utilisés avec l'objectif de prévenir les maladies.

¹⁰ Drogue est un terme associé qui désigne la substance utilisée pour la préparation des formes pharmaceutiques. L'utilisation de ce terme est évitée à cause de la connotation sociale actuelle.
11 Cf. Vigier (1716 : 1-2).

[«]Pharmácia, que vem do nome grego Pharmacon, quer dizer medicamento, he a arte que ensina a preparar os remédios. [...] A Galénica he aquella que se contenta da simplez mistura, sem especulação das substâncias que cada hua das drogas he naturalmente composta». Plenck (1786: 2). [Pharmacie, qui vient du grec Pharmacon, veut dire médicament, est l'art qui enseigne à préparer les remèdes. [...] La Galénique est celle qui s'occupe du simple mélange, sans spéculation des substances dont chaque drogue est naturellement composée.]

Cette confusion existe encore et beaucoup d'auteurs rappellent le besoin de correction terminologique (Pita 1993 : 205).

est «présentation pharmaceutique» 12.

Par la mise à jour systématique des traits sémantiques et conceptuels, on pourra voir les organisations conceptuelles successives et l'évolution des connaissances qui imposent une nature dynamique à la langue de spécialité ¹³. Le stockage des résultats de cette analyse était probablement compliqué dans les bases de données terminologiques (BDT), il ne l'est plus dans les bases de connaissances terminologiques (BCT) dans lesquelles la distinction entre données linguistiques et données conceptuelles peut être faite.

L'analyse conceptuelle/linguistique des termes montre que certains traits se maintiennent, quelques-uns disparaissent et que d'autres s'ajoutent au fur et à mesure que les connaissances du domaine évoluent. D'ailleurs, on constate que les traits fondateurs accompagnent le terme, lorsqu'il devient formant d'un terme composé comme «preparação medicamentosa» [préparation médicamenteuse] ou «medicamento essencial» [médicament essentiel].

Du fait que «la scienza [è] come un gioco a tre : una mente che indaga, la natura chi responde e la comunita che debate» (Pera 1991), la mémoire du terme est mise en cause à chaque fois qu'il est utilisé et elle ne résiste que parce qu'elle constitue en quelque sorte son code génétique.

RÉFÉRENCES

- AAVV (1994): Studi sul trattamento linguistico dell'informazione scientifica, a cura di Tullio di Mauro, Roma, Bulzoni Editore, 342 pages.
- CONCEIÇÃO, M. C. (1996): «Terminologias em Português: uma questão de sobrevivência», Mª. H. Mateus (dir.), Revista Internacional de Língua Portuguesa, nº 15, Lisboa, Associação das Universidades de Língua Portuguesa, pp. 25-29.
- CONCEIÇÃO, M. C. (1996): «Socioterminologia», The Linguistic Construction of Social and Personal Identity, Special Conference on Sociolinguistics, Évora, International Sociological Association CIDEHUS, pp. 1-6.
- HAGGARD, H. (1946): Devils, drugs and doctors, the story of Healing from Medicine-Man to Doctor, New York, Pocket Books, 490 pages.
- HARRIS, Z.(1988): Language and information, New York, Columbia University Press, 170 pages.
- JARRONSON, B. (1992): Invitation à la philosophie des sciences, coll. «Point Sciences», Paris, Seuil, 238 pages.
- LANGACKER, R. (1997): "The contextual basis of cognitive semantics", J. Nuyts et E. Peterson (eds), Language and conceptualization, coll. "Language, Culture & Cognition", Cambridge, Cambridge University Press, pp. 229-252.
- LINO, Mª T., R. COSTA et M. CONCEIÇÃO (1996): «Terminologia, informática e multimédia», *Jornada Panllatina de Terminologia, Perpectives i camps d'aplicació*, IV^e Réunion du Réseau Panlatin de Terminologie. Barcelone, IULA, Universitat Pompeu Fabra, pp. 73-81.

¹² Présentation pharmaceutique désigne le produit tel qu'il est délivré au patient ou au client de la pharmacie.

^{13 «}L'énonciation et la compréhension de textes participent comme tous les échanges sémiotiques de l'évolution des cultures, et par là de la phylogenèse dont elle est l'aboutissement» (Rastier et alii 1994 : 22).

- OTMAN, G. (1996): Les représentations sémantiques en terminologie, coll. «Sciences cognitives», Paris / Milan / Barcelone, Masson, 216 pages.
- PERA, M. (1991): Scienza e retorica, coll. «Biblioteca di Cultura Moderna Laterza», Bari, Laterza, 268 pages.
- RASTIER, F., M. CAVAZZA et A. ABILLÉ (1994): Sémantique pour l'analyse, de la linguistique à l'informatique, coll. «Sciences cognitives», Paris / Milan / Barcelone, Masson, 240 pages.
- ROBINSON, E. (1997): "The cognitive foundations of pragmatic principles: implications for theories of linguistic and cognitive representation," J. Nuyts et E. Peterson (eds), Language and conceptualization, coll. "Language, Culture & Cognition", Cambridge, Cambridge University Press, pp. 253-271.
- THOIRON, P. (1995a): «La terminologie multilingue et le traducteur», A. Hermans (dir), Les dictionnaires spécialisés et l'Analyse de la Valeur, coll. «Bibliothèque des Cahiers de l'Institut de Linguistique de Louvain», Louvain-la-Neuve, Peeters, 1997, pp. 117-130.
- THOIRON, P. (1995b): «Concepts, termes et traduction», Communication présentée au 120^e Congrès National des Sociétés Historiques et Scientifiques, 23-29 Octobre 1995, Aixen-Provence, 8 pages.
- WEISSENHOFER, P. (1995): Conceptology in terminology theory, semantics and word-formation, «IITF-Series», Vienne, TermNet, XI + 270 pages.

ANNEXE

- AAVV (1987): Farmacopeia Portuguesa V, Lisboa, INCM. 14
- GOMES, António Bernardino (1815): Elementos de pharmacologia geral ou Princípios geraes de matéria médica e de therapêutica, Lisboa, Typographia da Academia Real das Sciencias.
- PINTO, Agostinho A.S. (1835): Pharmacographia do código pharmacêutico lusitano ou Tratado de Pharmacodinamia, Coimbra, imp da Universidade. 15
- PITA, João Rui (1993): Farmácia e medicamento, noções gerais, Coimbra, Minerva. 16
- PLENCK, M. (1786): Pharmacologie médicale, Paris, s/i.
- SANTO ANTÓNIO, D. Caetano (1704): Pharmacopea Lusitana, methodo prático de preparar & compor medicamentos na forma Galénica com todas as receitas mais uzuais, Coimbra, Joam Antunes. 17
- TAVARES, Francisco (1794): Pharmacopeia geral para o reino e domínios de Portugal, Lisboa, s/i. 18
- VIGIER, Joam (1716): Pharmacopea ulyssiponense, galénica e chymica que contem ..., Lisboa, Pascoal da Sylva.

¹⁴ Texte officiel en vigueur et qui est une adaptation de *Pharmacopée Européenne*.

¹⁵ Deuxième pharmacopée officielle.

¹⁶ En ce qui concerne les définitions, cet auteur suit la loi en vigueur. Il cite souvent la Pharmacopée Portugaise et la législation où les concepts sont définis.

¹⁷ Cet auteur copie beaucoup d'extraits de textes en latin et les traduit. La définition de médicament qu'il présente, par exemple, est celle de Galien.

¹⁸ Première pharmacopée officielle.

DES MOTS QUI ONT LA MÉMOIRE LONGUE : LE GENOU, LA SOURIS ET L'ARBRE

Étienne TIFFOU

Université de Montréal, Canada

INTRODUCTION

La mémoire des mots, thème retenu pour ce congrès, se prête à différentes interprétations. S'agit-il des processus selon lesquels les mots s'ancrent plus ou moins durablement dans la mémoire des sujets parlants? Ces processus sont d'ailleurs multiples. Il peut, en effet, s'agir de clarté de lecture d'un terme, d'associations insolites d'idées qui maintiennent sa vitalité, de l'aisance à en manier le paradigme, etc. Il peut également s'agir de l'art d'ériger un mémorial des mots, ce qui est une façon de leur assurer une certaine pérennité. Cet art intéresse, entre autres, mais pas exclusivement, la conception et l'élaboration de dictionnaires. Il n'est pas interdit non plus de prêter aux mots une mémoire grâce à laquelle il peuvent non seulement raconter des histoires perdues, mais aussi se survivre à eux-mêmes ou se régénérer.

De toutes ces possibilités, c'est la dernière que j'ai retenue. C'est donc une étude diachronique que j'entends mener. Mais celle-ci ne se limitera pas à une simple histoire de mots. Après avoir rappelé celle du genou, celle de la souris permettra de comprendre l'importance de la métaphore dans la création des mots; celle de l'arbre confirmera cette importance en montrant que la structure cognitive l'emporte sur une symbolique en principe d'autant plus contraignante qu'elle est inconsciente. Un tel résultat devrait nous inviter à nous interroger sur l'arbitrarité du signe linguistique pour conclure que celle-ci mérite d'être partiellement reconsidérée.

LE GENOU

On a fait remonter le genou à la même racine que naître et connaître. Selon cette hypothèse, on partirait d'une racine indo-européenne $*g(^e/o)n$ - suffixée soit en $*-(^e/o)H_1$, soit en $*-(^e/o)H_3$. Les radicaux comportant le premier suffixe seraient concernés par l'idée de naissance et ceux qui comportent le second seraient intéressés par l'idée de connaissance. Ainsi sur le thème I de la racine suffixée en $*-(^e/o)H_1$ ($*g^e/on-H_1$) ont été dérivés le grec genevtwr, le latin genitor et la sanskrit janita, alors que sur le thème III ($*gn-H_1$) on a tiré les verbes givgnomai, (g)nascor, jajñe. Tous ces termes sont sémantiquement concernés par l'idée de naissance. La racine suffixée en $*-(^e/o)H_3$ est attestée surtout au thème II (ex.: gignwyskw, co-gnósco, jña-) et réfère toujours à l'idée de connaissance. Quant au

thème du genou, il est très probable qu'il remonte à la même racine, sans qu'on puisse dire si ce thème était suffixé ou pas. Dans le cas où il le serait, nous aurions affaire à un thème I sans que l'on puisse postuler en toute certitude l'une des deux laryngales, puisque cellesci tombent devant voyelles. Si l'on admet qu'il faille poser une développante, il est fort probable qu'il s'agisse plutôt d'un H_1 . Toutefois la forme sanskrite jánu donne à penser que ce terme n'était pas suffixé. En effet, la longue peut s'expliquer par la loi de Brugmann; or, si le -n- était suivi d'une laryngale, la voyelle indo-européenne aurait été en syllabe fermée et la loi n'aurait pu de ce fait s'appliquer. Au reste, le thème pour désigner le genou est fort stable dans les langues indo-européennes, quoiqu'il présente selon cellesci une alternance de timbre : govnu en grec, genú en latin et en hittite genu.

Toutefois, cette parenté ne peut être absolument prouvée. P. Chantraine note bien dans son dictionnaire étymologique de la langue grecque (1968 : 233) qu'on s'est demandé «si les deux familles divergentes de gignwyskw et de givgnomai ne devraient pas être rapportées au nom du genou.» Malheureusement, cette hypothèse, ajoute-t-il quelques lignes plus bas, ne peut se démontrer rigoureusement et il n'est guère possible de faire de givgnomai et gignwyskw des dénominatifs du nom du genou. Quant à Ernout et Meillet (1951: 273B), ils penchent en faveur de ce rapprochement, car, selon eux, il y a suffisamment d'éléments pour rendre cette étymologie tout à fait vraisemblable. D'ailleurs, pour que preuve soit faite, pourquoi est-il nécessaire que givgnomai et gignwyskw soient des dénominatifs du nom désignant le genou ? En outre, les remarques de Meillet (1926 : 54-55) sur l'adjectif anglais genuine évidemment dérivé sur celui-ci, plaide en faveur d'un rapprochement avec l'idée de naissance et de connaissance. Au reste, la forme latine invite à se demander si son vocalisme ne serait pas dû à une influence de geno. Enfin et surtout, d'anciens textes montrent à l'évidence que le genou jouait une place importante dans la reconnaissance du nouveau-né, ce qui le relie clairement à l'idée de naissance et de connaissance.

Le survol des mots dérivés de la racine $*g(^e/o)n$ - ne laisse pas, en effet, d'intriguer. Quels rapports de sens peut-il y avoir entre la naissance, la connaissance et le genou ? En français, les deux premiers termes sont clairement apparentés; ce rapport formel maintient de façon fossilisée une relation entre ceux-ci. L'hébreu présente, au demeurant, une affinité similaire, mais il semble bien que les idées de naissance et de connaissance soient reliées l'une à l'autre de façon différente. La connaissance précède la naissance, puisqu'il fallait qu'Adam connût Ève pour que Caïn et Abel en naquissent. D'après de vieilles légendes, il ressort qu'il en va au rebours en indo-européen et que ce soit la naissance qui précède la connaissance. Cela s'impose si l'on veut intégrer le genou dans le jeu des relations sémantiques. Dans une version hittite éditée par Gaster (1953 : 101-122) d'un conte hourrite, remontant fort probablement lui-même à une version sumérienne plus ancienne, Koumarbi féconde une montagne qui enfante «un énorme galet en forme de nouveau-né» :

«...les déesses s'avancèrent pour le recevoir, le portèrent tendrement à Koumarbi et le placèrent sur ses genoux.

Koumarbi était transporté de joie; prenant l'enfant dans ses bras, il l'éleva en l'air et ensuite le berça sur ses genoux en chantonnant [...] il cessa enfin, installa l'enfant commodément sur ses genoux : "Petit, dit-il, il me faut te donner un nom; tu t'appelleras Oullikoumi"» (1953 : 106-107).

Ce texte revient à plusieurs reprises sur l'installation du nouveau-né sur les genoux de celui qui l'a engendré. À ce propos, Gaster (1953 : 117) note un peu plus loin :

«...l'enfant est placé sur les genoux de son père; c'était là, dans les temps anciens, une façon habituelle de reconnaître la paternité, les genoux étant alors supposés avoir un rapport anatomique direct avec la génération; et l'adjectif anglais *genuine* (véritable, authentique) signifie à l'origine "qui a été mis sur les genoux de son père pour être reconnu"».

Plutôt que de voir dans le genou «un rapport anatomique direct avec la génération», il est préférable de le considérer comme l'organe de l'engendrement juridique. Cette symbolique est sans nul doute bien connue des indo-européens. La naissance de Dionysos qui sort de la cuisse de Zeus après qu'il a été récupéré du sein de sa mère Sémélé foudroyée par la vision du roi des dieux dans sa toute puissance, est une transposition de la reconnaissance juridique de l'enfant qui vient de naître.

Au terme de cette brève présentation, le mot genou ouvre le champ à la mémoire sur des pratiques biologiques et juridiques. L'origine qu'il partage avec le mot *gonade* s'explique ainsi. Les mots portent en eux les clés qui donnent l'accès à la connaissance de temps révolus dont la documentation est pratiquement inexistante. Le cas du genou est intéressant; il diffère, en effet, des deux autres cas que nous allons examiner, car il est contextuel et ne peut être compris sans référence à l'idée de connaissance et de naissance. Cet état complexe sans éradiquer la métaphore sous-jacente l'a déstabilisée. Ce terme porte témoignage à un type de mémoire à moitié effacée à laquelle il assure néanmoins une certaine pérennité.

LA SOURIS

La souris se prête à bien des métaphores relativement récentes qui n'intéressent, de ce fait, que médiocrement notre propos. Ainsi la désignation familière de femme sous le nom de souris n'est pas riche d'enseignement. On appelait souris de rempart les personnes du beau sexe trafiquant de leurs charmes avec les militaires auprès desquels elles se glissaient avec la discrétion furtive d'une souris. Amputé de son déterminant, ce terme devait par la suite s'appliquer à celles qu'on appelait également à cette époque des lorettes; la connotation péjorative qu'il comportait s'est finalement perdue et ce mot désormais s'emploie familièrement pour parler d'une jeune femme (cf. Robert : 1996B). En revanche, confronté à un mot latin différent, mais peu éloigné sémantiquement, le nom de la souris se prête à des analyses beaucoup plus révélatrices qui permettent de raviver de vieux souvenirs.

La souris vient d'un mot latin sorex, soricis attesté dans les Res Rusticæ de Varron. Ce terme est indo-européen et serait dérivé d'une racine suffixée $*s(^e/o)w-(^e/o)r$. Le mot latin serait tiré du thème II (*sw-er- > *swor- > sor-), alors que son correspondant grec u{rax viendrait du thème III (*sw-r- > hur-). Il est parallèle à un nom racine indo-européen attesté notamment en latin (mús), en grec (mu-)et en sanskrit (mú1-); selon Chantraine (1984 : 725B), la longue s'expliquerait par le caractère monosyllabique de ce mot. Celui-ci a survécu en français dans le nom de la musaraigne «la souris/le rat araignée» qui passait pour être venimeuse. C'est la seule attestation que nous ayons de ce

terme avec sa valeur propre. Il est mieux connu dans un autre sens dérivé par métaphore, mais celle-ci n'est plus perçue des sujets parlants. Il s'agit du muscle.

À vrai dire cette métaphore était largement répandue dans bien des langues indoeuropéennes. En s'appuyant sur l'analogie de forme, voire de mouvement, des biceps, on a désigné le muscle comme le «petit rat» ou la «petite souris»; ainsi le nom du rat en grec mu~" désigne également le muscle. Quant au latin, il a adjoint au mot mús un suffixe de diminutif à valeur expressive : -culu(m). Ce suffixe est bien attesté en français où la plupart des mots en -eil remontent à ce suffise : sommeil < somniculu(m); soleil < soliculu(m), etc.). Le muscle signifierait littéralement le «souriceau». Il n'y a pas lieu de distinguer la souris du rat; d'ailleurs l'opposition entre les termes rat et souris n'est pas claire et n'influence pas, à l'origine, la métaphore.

Il n'y aurait pas grand chose à ajouter sur l'étymologie du mot muscle, si la métaphore perdue n'était pas réapparue sous une autre forme. En effet, de nos jours, il est un petit muscle vers le haut du manche du gigot d'agneau que l'on nomme la souris. La métaphore est claire pour les sujets parlants et à ceux qui s'étonnent de cette appellation on explique que la forme de ce muscle fait penser à une petite souris. Nous aurions là un cas de résurrection métaphorique intéressante. Celle-ci n'est pas cependant aussi évidente. Le terme souris avait, en effet, au siècle dernier une signification moins restreinte :

«Par analogie de forme avec le corps de la souris ou un de ses membres, le mot s'est employé comme le latin *mus*, et surtout son dérivé *musculus* (→ muscle), désignant la partie charnue du bras, de la main ou de la jambe (v. 1250 *soris*), l'espace de la main entre le pouce et l'index (1690), sens disparu au XIX^e s.; *souris* est le nom encore usuel (1694) d'un muscle charnu à l'extrémité du gigot, contre l'os.» (Robert 1994 : 1996B)

Il ressort de cette citation que la souris n'avait pas, lorsqu'elle était employée métaphoriquement un sens aussi spécialisé que celui qu'il a de nos jours. Est-ce à dire que la métaphore a été recréée avec ce terme au cours de l'histoire du français ? Ou est-ce à dire que celui-ci avait parallèlement eu les mêmes emplois que *mus* qu'il avait fini par perdre au profit de ce dernier ? Il est certain que dans ce cas, l'histoire de ce mot est beaucoup moins riche d'enseignement et ne ferait que témoigner une fois de plus des combats que les mots se livrent entre eux pour assurer leur survie.

S'il est permis de laisser planer, faute de documentation, un doute sur la solution à retenir, un élément nouveau permet de redonner à ce type de métaphore l'intérêt qu'il mérite. Cette métaphore n'est pas propre à un groupe de langues apparentées, mais se retrouve dans d'autres familles linguistiques. Ainsi, en arabe (exemple donné par A. Mehiri, professeur émérite de l'Université de Tunis), fár désigne la souris mâle et fára la souris femelle. Ce dernier mot est aussi employé pour parler du muscle de la jambe, alors que le premier peut également signifier la «chair» du dos. Le bourouchaski, langue isolée de l'Himalaya, présente également ce type de métaphore. Le mot pour rat, souris se dit gírkis, or on lit dans le lexique de Berger (1974 : 145B) : gírkis cap «Bizeps» et Lorimer, dans son dictionnaire (1938 : 168A), propose pour le dialecte du Hounza le terme píni gírkis qu'il glose en «muscle of calf of leg. Cp. Latin mus : musculum». Il est sûr que les Bourouchos ont élaboré cette métaphore de façon indépendante. Cela montre à l'évidence que nous avons là une image de type universel qui peut s'imposer dans des

milieux totalement différents indépendamment de toute influence extérieure. Un tel type de métaphore fait appel en quelque sorte aussi bien à une mémoire dans le temps qu'à une mémoire dans l'espace.

L'ARBRE

Le nom de l'arbre est bien connu en indo-européen; il repose sur une racine suffixée $*d(^e/o)r-(^e/o)w-$ dont les trois thèmes sont bien représentés; ex.: thème I: gr. dovru, skt. dáru (*dor-w); thème II: gr. devndreon, angl. tree (*dr-ew); thème III: gr. dru~w «le chêne» (*dr-w). Employé de façon métaphorique, il est devenu symbole de la stabilité et de la solidité aussi bien au sens propre qu'au sens figuré. C'est ainsi que nous devons à Hésychius d'avoir conservé le mot droÛovn qu'il glose en ijscurov" «fort, solide». Les langues germaniques offrent de bons exemples du sens figuré, tels le treu de l'allemand «loyal, fidèle» et le truth de l'anglais «vérité».

Il est possible, selon moi, de remonter plus haut en voyant dans le nom de l'arbre une métaphore très ancienne. Cette hypothèse que j'ai conçue il y a plusieurs années, ne laissera pas d'étonner, mais j'ai été rassuré en trouvant dans le livre récent de Gamkrelidze et d'Ivanov (1995) une étymologie proche de celle que j'avais parallèlement élaborée. La racine *d(e/o)r- fait penser à celle que l'on a postulée pour la peau et que l'on retrouve dans le grec devrma «peau». Ce nom comporte un suffixe *-m(e/o)n, suffixe de nom d'action permettant de dériver des déverbatifs. De fait, il dérive d'un thème verbal devrw attesté avec le sens d'«écorcher, dépouiller». On ne s'étonnera pas, dans ces conditions, que le mot devrma désigne la peau que l'on arrache. Ainsi que le note, Chantraine (1968 : 265B) «devrma n. a dû se dire d'abord de la peau dépouillée d'un animal [...] de peaux préparées pour faire des sacs, etc.» Le rapprochement formel entre les noms signifiant respectivement la peau et l'arbre ne manque pas de vraisemblance, mais cette ressemblance peut être fortuite et résulter de neutralisations. C'est un avatar qui se produit fréquemment et qui est bien connu des comparatistes. D'autre part, le rapport sémantique entre les deux mots est loin d'être clair. Ce n'est qu'en trouvant un lien acceptable entre eux que l'hypothèse pourra avoir un certain fondement.

Gamkrelidze et Ivanov (1995 : 202) remarquent qu'il est possible de comparer sémantiquement les thèmes de l'arbre avec celui de la peau, ou plus exactement de son arrachement :

«This base, meaning 'tree; oak' (c'est-à-dire *t'e/oru- et *t'reu-) [...] can be semantically compared to t'er- (Lith. derù, dìrti 'cut off'. Latv. nuõdara 'pole with branches cut off', Russ. dial. dor 'logs that split easily') and *t'r-oH (Olcel. troD(a) 'pole', Slav. drati 'remove skin or bark'...). This may indicate [...] that the form is an extended root in Benveniste's sense with two ablaut states, *t'or-u-/t'r-eu.»

Il ressort de cette citation que le radical **t'er-** s'applique à un élément en bois fragile ou émondé; mais c'est surtout, et c'est également notre hypothèse, l'idée d'enlever l'écorce qui s'impose. L'arbre serait en quelque sorte «l'écorcé» ou «l'écorçable». Cette idée est liée essentiellement, selon moi, à celle du chêne, et ce de deux façons. Il peut s'agir du chêne liège; lorsque le liège a été récolté, les forêts dont il provient semble être formées d'arbres écorchés. On peut également proposer une explication moins restrictive; le bois du chêne est extrêmement solide et entre dans la fabrication de divers objets résistants (ex. : skt

druváya «plat en bois»; drún5am «épée», drun5í «seau», etc.), ce qui suppose que ce bois pour être utilisé devait être le plus fréquemment écorcé. Au reste, ces deux points de vue ne s'excluent pas et ont pu concourir à motiver la désignation de l'arbre.

Si cette hypothèse mérite d'être retenue, il est clair que la métaphore remonte à une date très ancienne et qu'elle a été occultée par d'autres métaphores que nous avons évoquées et à la base desquelles se trouve l'arbre. Où donc est la mémoire de ce mot ? Dans l'émergence de la métaphore dans d'autres termes d'origine différente. C'est ainsi que le français a retrouvé un rapport entre l'arbre et la peau arrachée dans les termes «écorce» et «écorché». D'un point de vue synchronique la ressemblance est frappante et impose à l'esprit un rapport de sens évident. Malheureusement, l'origine de ces deux noms est différente. À propos de l'écorce on trouve dans le *Robert étymologique* (1994 : 654B) :

ÉCORCE n. f. représente l'aboutissement (1176) du latin impérial *scortea* «manteau de peau». Ce mot dérive de *scortum* «peau, cuir», qui se rattache à une racine indo-européenne **sker*- ou **ker*-...

Le mot apparaît dans le sens d'«enveloppe des troncs, des branches, d'arbres» qui correspond au latin *cortex…*

Quant à écorcher, on lit dans la rubrique qui lui est consacrée :

ÉCORCHER v. tr. est issu (1155) du bas latin *excorticare* «écorcer», composé de *ex*- et du latin classique *cortex*, *corticis* «écorce», qui se rattache à une racine indoeuropéenne **sker*- «couper», «séparer», «partager». (1994 : 654B)

Ces étymologies ne sont divergentes qu'en apparence. Il est vrai qu'écorce vient de *scortum*, alors qu'écorcher dérive de *cortex*. Mais ces deux mots ont même origine et remontent, comme il est signalé dans le Robert, à un thème indo-européen *sker-. La métaphore s'était donc reconstituée à date plus ancienne que le français. Celui-ci met, au reste, en évidence la force de la métaphore, en détournant les termes latins de leur valeur propre; en effet, le terme qui s'applique à l'arbre dérive de celui qui signifie «peau», tandis que celui qui s'applique à la peau dérive de celui d'«écorce».

C'est ainsi que cette métaphore, apparemment insolite, se trouve fondée par un jeu de mémoire qui ne restitue pas un état ancien, mais réassigne à des termes nouveaux apparentés une vieille relation métaphorique plus ancienne que celle qui lie la notion d'arbre à celle de vérité.

INTERPRÉTATION

Cette présentation rapide de trois mots à la mémoire longue invite à les caractériser dans un premier temps les uns par rapport aux autres. En fait, le «genou» s'oppose clairement aux deux derniers. Il s'agit d'un cas classique de reconstitution historique à partir de données linguistiques. Là où la documentation en textes ou en vestiges manque, l'histoire des mots peut apporter de précieuses informations. É. Benveniste a donné dans ses Études indo-européennes (1969) une brillante démonstration de ce que cette méthode permettait d'obtenir. On ne pouvait pas ne pas rappeler dans une étude consacrée à la mémoire des mots ce genre de démarche.

La souris et l'arbre introduisent de façon complémentaire, et c'est pour cela que je les ai retenus, à des réflexions d'un ordre différent où synchronie et diachronie finissent par se conjoindre. On verra une des oppositions particulièrement éclairante dans leur valeur symbolique. La souris, en effet, en est très peu chargée. Dans leur Dictionnaire des symboles (1982: 904B), Chevalier et Gheerbrant lui consacrent moins d'une demi colonne et encore n'est-ce que pour faire allusion à des croyances bambaras qui lient la souris à l'excision des jeunes filles. En revanche, ces deux auteurs consacrent à l'arbre treize colonnes (1982 : 62A-68A) tout au long desquelles ils présentent les mythes liés à l'arbre, «l'un des thèmes symboliques les plus riches et les plus répandus; celui également dont la bibliographie, à elle seule, formerait un livre» (1982 : 62A). Or il est significatif que les valeurs symboliques que l'on prête à l'arbre ne soient pas inscrites dans la métaphore qui vient d'être présentée. Celle-ci est de type cognitif et est totalement comparable à celle de la souris et du muscle qui repose sur une analogie de forme et peutêtre de mouvement, et sur rien d'autre. Cela peut mettre en cause la lecture que l'on pourrait faire d'archétypes, dans le sens de Jung, à partir du langage. Les mots sur le divan, pour reprendre l'expression de Nouss, ont beaucoup moins à dire qu'on ne pense. On a plus à apprendre d'eux sur le fonctionnement de notre esprit que sur celui de notre inconscient. C'est du moins la première leçon que je suis tenté de tirer de cette brève, trop brève analyse.

Ce genre de métaphore que je qualifierais de type cognitif permet de tirer quelques informations sur le processus de dénomination. Tout d'abord ce processus tend à une certaine universalité. La liaison de l'arbre et de la peau est généralisée tout au long de l'histoire de l'indo-européen et de ses dialectes, puisqu'elle est sous-jacente dans le rapprochement à faire entre deux racines de sens apparemment très opposé, qu'elle continue de survivre, à vrai dire bien masquée, dans le scortum et le cortex des Latins et qu'on la retrouve enfin remise en lumière en français avec l'écorce et l'écorché. La souris est encore plus significative, car on la retrouve liée au muscle dans des langues qui ne sont pas d'origine indo-européenne. Ce caractère universel invite à considérer que la production lexicale n'est pas toujours de type strictement linguistique et que des schèmes mentaux qui en sont indépendants peuvent la conditionner. Sans prétendre trancher la question inépuisable des rapports de la pensée et du langage, nous avons ici un cas assez clair de préséance de la pensée sur le langage.

Enfin, à la lumière de ce qui précède, on peut rappeler quelques nuances à apporter à l'arbitrarité du signe linguistique. On a déjà mis en évidence la motivation intersigne; il est évident que les ressemblances que l'on décèle entre «jardin» et «jardinage» n'est pas le fruit d'une décision non motivée. Des mots, comme ceux que nous venons d'analyser, montrent à l'évidence qu'il peut y avoir également, dans certains cas, motivation entre le signifié et le référent. C'est dans les structures de notre esprit qu'il faut voir les raisons décisives de cette motivation. Le caractère universel de certaines métaphores en donne une amorce de preuve qui est loin d'être négligeable.

CONCLUSION

De même que l'on se rappelle les mots, les mots se souviennent. Ce souvenir est parfois crypté, mais, clair ou opaque, il est porté par une métaphore. C'est donc à l'étude de la métaphore que renvoie la mémoire des mots, ce que j'ai, au demeurant essayé de mener partiellement, il est vrai, au long de cet exposé. Lakoff et Johnson (1985) se sont attachés à étudier la place de la métaphore dans les situations quotidiennes de communication. D'après eux, il existe des principes selon lesquels les métaphores s'organisent. Ainsi une idée d'agression ou d'offensive est liée aux activités intellectuelles:

«Vos affirmations sont *indéfendables*. Il a attaqué chaque point faible de mon argumentation. Ses critiques visaient droit au but. J'ai démoli son argumentation. Je n'ai jamais gagné sur un point avec lui. Tu n'es pas d'accord? Alors, défendstoi! Si tu utilise cette stratégie, il va t'écraser. Les arguments qu'il m'a opposés ont tous fait mouche.» (1985: 14)

L'approche de Lakoff et Johnson est très intéressante, mais le genre de métaphore étudiée est complexe, car toujours lié à un contexte, et éminemment variable selon le type de société. Mon point de vue, surtout historique, est peut-être plus fondamental, car les métaphores qui m'intéressent sont liées au mot qui les charrie indépendamment de tout contexte. Dans cette mesure, elles portent en elles une certaine universalité; mais avant que de poser les principes sous-jacents aux métaphores marquées de cette universalité, il est nécessaire d'en recueillir le plus grand nombre possible. J'en ai proposé un maigre bouquet; il faudrait travailler plus longtemps aux champs¹. D'autre part, la validité de la métaphore doit être testée diachroniquement; il faut, en d'autres termes, en décrire la mémoire qu'en fait la parole et voir si celle-ci ne la réajuste pas au goût du jour. Si tel est le cas, diachronie et synchronie s'harmonisent pour porter témoignage à la stabilité que l'esprit maintient au delà de ses changements, voire de ses errances. J'espère que la mémoire des trois mots que j'ai évoqués contribuera à montrer comment cette harmonie finit, dans certains cas, par être atteinte.

RÉFÉRENCES

BENVENISTE, É. (1969): Le vocabulaire des institutions indo-européennes, Paris, Les éditions de minuit, 2 vol.

BERGER, H. (1974): Das Yasin Burushaski, Wiesbaden, Otto Harassowitz.

BERT, M. et J. SERME (À paraître): «A la recherche des tendances universelles dans l'évolution sémantique des termes désignant des parties de la tête».

BOISSON, C. (À paraître) : «Sur le trope lexicalisé 'la bouche du couteau' = le tranchant du couteau».

¹ Au reste ce travail est en cours. Témoins l'article de Brown, C. H. et S. R.Witkowski (à paraître) ainsi que l'étude de C. Boisson (à paraître) et celle de Bert Michel et J. Serme (A la recherche des tendances universelles dans l'évolution sémantique des termes (à paraître).

- BROWN, C. H. et S.R. WITKOWSKI (1981): «Figurative language in a universalist perspective», American Anthropologist, 8, 3, pp. 596-615
- CHANTRAINE, P. (1968 et 1984): Dictionnaire étymologique de la langue grecque, Paris, Klincksieck, 2 vol.
- CHEVALIER, J. et A. GHEERBRANT (1982): Dictionnaire des symboles, Paris, Robert Laffont/Jupiter.
- Dictionnaire historique de la langue française (1994): Paris, Dictionnaires LE ROBERT, 2 vol., nouvelle édition.
- ERNOUT, A. et A. MEILLET (1951): Dictionnaire étymologique de la langue latine, Paris, Klincksieck, 3^e édition.
- GAMKRELIDZE, V. et V. IVANOV (1995): Indo-european and the Indo-europeans. A reconstruction and historical analysis of a proto-language and a proto-culture, Berlin, New york, Walter de Gruyter, 2 t.
- GASTER, T. (1953): Les plus anciens contes de l'humanité, Paris, Payot.
- LAKOFF, G. et M. JOHNSON (1985): Les métaphores dans la vie quotidienne, Paris, Les éditions de minuit, traduit de l'américain par M. Defornel avec la collaboration de J.-J. Lecercle.
- LORIMER, D.L.R. (1938): The Burushaski Language. Vocabularies and index, Oslo, Instituttet forsammenlignende kulturforskning.
- MEILLET, A. (1926): BSL, XXVII, p. 54 sq.

LA MÉMOIRE DU COMMERCE : ASPECTS ÉTYMOLOGIQUES ET PHRASÉOLOGIQUES

Jeanne DANCETTE et Christophe RÉTHORÉ

Université de Montréal, Canada

INTRODUCTION

En 1949, en France, Édouard Leclerc, fondateur de la chaîne de grandes surfaces du même nom, décide d'appeler ses magasins «centres de distribution». Dès 1950, les nouveaux commerces qui, comme Leclerc, pratiquent le libre service se désignent également sous le nom de «distributeurs», plutôt que «commerçants». L'habitude est prise et le commerce moderne s'appellera désormais «distribution» (Benoun et Héliès-Hassid 1995: 15).

Ce changement de paradigme, de commerce à distribution, a intrigué les auteurs de cet article, qui travaillent depuis plusieurs années sur le vocabulaire de la distribution (Dancette et Réthoré 1997). Nous avons voulu en examiner les causes et les modalités. Nous sentions que ce changement de direction terminologique ne s'expliquait pas par une différenciation sémantique fondamentale, par des traits qui s'opposeraient tels que le commerce s'oppose à l'industrie, par exemple, mais plutôt par les connotations négatives attachées à l'ensemble {commerce, commerçant, commercial, etc.} et dont il fallait se défaire pour redorer le blason du commerce.

Notre étude de cette question nous a conduits à chercher une explication dans les développements de l'histoire économique à travers les âges. Ensuite, nous avons examiné comment les dictionnaires de français, du 16^e au 20^e siècle, attestaient de l'évolution de ces termes. Enfin, par l'étude lexicométrique de la presse actuelle française et québécoise, nous avons cherché à voir si l'on retrouvait dans la société d'aujourd'hui le jeu de connotations que nous supposions, généralement négatives pour l'ensemble {commerce, commerçant, etc.} et neutres ou positives pour {distribution, distributeur, etc.}, en analysant comment de telles connotations opéraient dans des cultures différentes (France et Ouébec).

1. HISTORIQUE

1.1 Des marchands du Temple au marché noir, en passant par la route des Indes

Depuis ses origines, le commerce suscite des sentiments contradictoires, de l'admiration au mépris. De tout temps et dès l'Antiquité, les grands peuples voyageurs et

marchands, les Crétois, les Phéniciens, les Vénitiens, font naître l'admiration (Attali 1991). Les grands exploits liés à la découverte de la route des épices et de la soie nourrissent l'imagination et donnent lieu à une riche littérature mettant à l'honneur la débrouillardise et l'intelligence du voyageur, ainsi que l'exotisme des contrées lointaines.

«Je me rendis à Bassora, où je m'embarquai avec plusieurs marchands sur un vaisseau que nous avions équipé à frais communs.

Nous mîmes à la voile, et prîmes la route des Indes orientales par le golfe Persique...»

Les Mille et Une Nuits. Histoire de Sindbad le Marin (1996 : 11)

Dans l'histoire des nations, le commerce est reconnu comme facteur indéniable d'évolution, mais il se heurte au conservatisme des forces en place. Les peuples marins, marchands, s'opposent aux peuples terriens, guerriers et paysans. C'est la thèse de Fernand Braudel (1985 : 179) : jusqu'en 146 avant J.-C., Carthage, convoitée et détruite par Rome, symbolise la lutte «d'un peuple essentiellement marin et marchand et d'un peuple essentiellement terrien, guerrier et paysan». Un peu plus tard, poursuit Braudel, l'essor du commerce en Islam fait apparaître le clivage entre une Méditerranée orientale développée, où l'on connaît la monnaie, la lettre de change et le chèque — en arabe *chakk* — et une Méditerranée occidentale, barbare et sous-développée. Les villes situées à l'articulation de ces deux mondes occupent une position pivot. À Carthage, comme à Bagdad ou à Damas, convergent les esclaves d'Afrique, le plomb, le cuir d'Andalousie et de Sardaigne, l'or de l'Afrique noire, le blé d'Afrique du Nord, etc.

«Je veux porter du safran persan à la Chine, où j'ai entendu dire qu'il y a grand prix, et ensuite de la porcelaine de Chine dans la Grèce, du brocard grec dans l'Inde, de l'ac indien à Alep, du verre d'Alep dans le Yémen et des étoffes rayées du Yémen en Perses ...» (Al-AQâsim al-Harîrî 1992)

D'une manière générale, l'essor du commerce entraîne la mise en place d'une économie monétaire, voire d'un capitalisme marchand où les fonctions de banquier et de commerçant tendent à se confondre. D'une part, le commerce est valorisé, car lié au développement des puissances; d'autre part, il fait l'objet d'une position ambiguë, y compris des souverains, qui encouragent ou freinent le commerce, selon les moments et les valeurs religieuses dominantes. Dans les pays d'Europe occidentale à forte tradition catholique, comme l'Italie, l'Espagne, le Portugal et la France, les activités commerciales sont généralement regardées avec mépris, à la différence des pays de religion protestante, et notamment calviniste, qui n'ont pas les mêmes réticences idéologiques face à l'argent (Charmes 1989 : 17-18; Weber 1964; Chung, Shepard et Dollinger 1989 : 309-311). L'histoire et l'origine des grandes fortunes à travers le monde montrent un lien évident entre les valeurs religieuses et culturelles d'un pays et l'enrichissement de ses ressortissants!

¹ L'influence de la religion sur le comportement économique des individus est documentée. Charmes (1989 : 16) donne les chiffres suivants qui datent des années 80. Aux États-Unis, sur un total de 65 milliardaires «recensés», on comptait un seul catholique et 64 protestants, dont 59 calvinistes. En France, parmi les 50 plus grosses fortunes, on trouvait 5 catholiques, 5 juifs et 40 protestants, dont 33 calvinistes. Il y a, selon ses termes, une «disproportion statistiquement significative» entre ces chiffres et la répartition religieuse que cet auteur illustre par le tableau ci-desssous (op. cit., 16) :

Quelle qu'en soit l'origine, les nouvelles fortunes font naître l'envie et revivre l'image des marchands du Temple, agressifs et mercantiles. Toutefois, les clichés, que nous retrouvons jusqu'au 20^e siècle, affectent surtout les moins fortunés des commerçants, présentés comme étant près de leurs sous, voire malhonnêtes et insensibles.

Au cours de la Seconde Guerre mondiale, les événements et le climat qui règnent en France à cette époque (occupation, collaboration, marché noir, etc.) ne font rien pour aider le commerce à se bâtir une nouvelle réputation, ainsi que le montrent ces citations empruntées à divers auteurs.

«À cirer la botte allemande, les bénéfices avaient été reluisants. Que l'on se donne la peine de consulter les chiffres : 10 266 faillites en 1935 pour 1 085 en 1941; celui des liquidations judiciaires passe de 2 222 en 1937 à 48 en 1943 ! Pour la même année, 1 478 épiceries nouvelles s'ajoutèrent à celles qui existaient déjà. Jamais le commerce n'avait été aussi florissant !» (Hallier 1972 : 66)

Même si l'on ne doit pas accorder trop d'importance aux chiffres donnés par cet auteur, notamment à cause du changement intervenu dans le mode de calcul statistique, cette citation atteste de l'image très négative du petit commerçant. Les spécialistes du domaine confirment cette vision :

«Après la Seconde Guerre mondiale, le commerce avait en France une image plus négative que jamais. Le commerçant déjà suspect, comme toujours, de ne pas être moral car gagnant de l'argent sans être productif, c'est-à-dire sans participer à la transformation des biens, se voyait, en plus, accusé d'avoir activement participé au marché noir. À la Libération, quand on voulait insulter un individu ou désigner un parvenu, on le traitait de "BOF" (beurre, œufs et fromage).» (Benoun et Héliès-Hassid 1995 : 14; voir aussi Dutourd 1952)

«... le commerce intérieur, considéré comme un simple rouage technique et juridique dépendant n'a guère suscité d'analyses ou de recherches fondamentales jusqu'au lendemain de la Seconde Guerre mondiale (1). On peut même affirmer que la théorie économique, [...] les sciences humaines et l'université en général cachaient mal une certaine condescendance pour l'univers de la boutique, au reste partagée par une opinion publique manifestement méfiante à l'égard des "intermédiaires", la période 1939-1945 n'ayant pas contribué à atténuer ce sentiment.» (Coquery 1977 : 1-2)

À ce ressentiment politico-social, il faut ajouter dans la période de l'immédiat après-guerre les reproches adressés aux commerçants pour leur côté traditionaliste et leur conservatisme, symbolisé, en France, par le poujadisme. Selon Ducrocq *et al.*, «en 1953, Poujade et le poujadisme confirment que le commerce français n'est pas animé par un esprit de progrès et d'innovation» (1994 : 10).

Religion	États-Unis	France
Catholiques	60 000 000	45 000 000
Protestants	76 000 000	785 000
dont calvinistes	6 000 000	90 000
Musulmans	?	2 000 000
Juifs	?	750 000

Voilà pour l'image peu glorieuse du commerçant. En changeant de nom, le «distributeur» Leclerc indique son intention de s'en démarquer, tant sur le plan éthique que purement commercial. Cette intention explique, au moins partiellement, l'apparition du nouveau paradigme lexical de la distribution. Certes, à ce facteur subjectif s'ajoute un fondement objectif, celui du changement structurel du commerce, lié à la distribution et à la consommation de masse, qu'avait amorcé Boucicaut en France avec la formule des grands magasins (du Closel 1989). Toutefois, le changement des valeurs semble lui aussi indéniable. La théorie économique du 20^e siècle en atteste. Si, au siècle dernier, le Littré rangeait encore le commerce dans les secteurs productifs : «Dans l'économie politique, l'agriculture, la fabrication, le commerce sont les trois branches de la production générale» (1873 : 684), au 20^e siècle, les économistes insistent sur l'image suspecte du commerçant, immoral puisqu'il gagne de l'argent sans être productif. L'I.N.S.E.E., l'office national français de la statistique, s'en est même fait l'écho puisqu'il rangeait jusqu'à récemment le commerce, système intermédiaire entre la production et la consommation, parmi les secteurs improductifs.

1.2 Apparition du paradigme de la distribution

Le divorce est ainsi consommé. La «distribution» est présentée par ses agents comme la solution des problèmes attachés au petit commerce. Ainsi, au «commerce sclérosé, fautif de goulets d'étranglement, aux marges trop élevées», s'opposent «ces industriels des temps modernes maîtrisant totalement l'écoulement de grandes quantités de produits» (Benoun et Héliès-Hassid 1995 : 13). À la boutique du commerçant, «symboliquement représenté par un crémier faisant les comptes des clients sur un carnet à souches avec un crayon qu'il dégage du coin de son oreille», s'oppose «un univers hautement technologique où règnent la lecture optique, l'informatique interactive et la manutention automatisée» et «demain la vente électronique [...] et le prélèvement automatique sur le compte bancaire ...» (Benoun et Héliès-Hassid 1995 : 13-14). Dans l'imaginaire populaire, cet univers impressionnant devient moins malhonnête, moins mesquin. Les distributeurs misent sur la productivité pour s'anoblir.

L'évolution technologique suffit-elle à garantir les valeurs éthiques? Ce n'est certes pas l'objet de cet article d'en discuter, mais on ne peut pas ignorer les problèmes que soulèvent certaines des pratiques souvent en usage dans le milieu de la distribution : relations conflictuelles avec les fabricants, paiements à 90 ou à 180 jours, antisyndicalisme, pressions sur les commissions d'urbanisme commercial. Toutefois, si la réputation de la distribution souffre parfois de telles pratiques, l'image auprès du public est différente, nettoyée des préjugés qui collent encore à la peau du commerçant, comme nous le verrons par l'étude des corpus de la presse actuelle. Mais au préalable, faisons une incursion dans le monde des dictionnaires pour voir comment {commerce, etc.} et {distribution, etc.} sont traités à travers les âges et quelles connotations s'attachent à ces termes et à leurs dérivés.

2. ÉTUDE ÉTYMOLOGIQUE ET LEXICOLOGIQUE DES PARADIGMES COMMERCE ET DISTRIBUTION

2.1 Les sens de commerce sont ambigus

Le mot commerce, qui vient du latin commercium, construit sur cum, avec et merx, marchandises (Dauzat et al. 1964; Rey et al. 1992), désigne «l'échange entre les

hommes des divers produits de la nature ou de l'industrie» (Littré 1873). Toutefois, il ne s'agit là que d'un sens parmi les cinq attestés dans le *Dictionnaire universel* de Furetière (1690). En effet, à ce premier sens neutre, dénotant une activité économique, s'ajoutent dès le 16^e siècle ceux de «relation réciproque» et de «trafic de choses morales». Signalons d'ailleurs que le *Dictionnaire de la langue française du seizième siècle* de Huguet (1925) ne donne «commerce» qu'au féminin, et uniquement dans le sens second de trafic des choses morales, avec une citation de Montaigne: «Nous qui n'avons aucune commerce avec eux. MONTAIGNE, II, 12, var. de 1588 (II, 166, et V, 122)» (Huguet, tome 2, 1925: 367).

Le Littré (1873) relève les connotations négatives attachées au commerce, dans le sens de trafic des choses morales :

```
«Que vois-je autour de moi que des amis vendus,
Qui, choisis par Néron pour ce commerce infâme...»
(Racine, Britannicus, I,4).
```

Tous les dictionnaires consultés sont unanimes lorsqu'ils consignent les expressions «faire un mauvais / un méchant / un vilain commerce» pour dire «se mêler de quelque vilaine affaire», à côté des expressions neutres ou même positives telles que «le commerce des hommes» :

```
«Dans le brillant commerce il se mêle sans cesse
Et ne cite jamais que duc, prince ou princesse»
(Molière, Le Misanthrope, V,1.)
```

ou bien le «commerce des lettres» et le «commerce des muses» pour désigner les occupations littéraires. Il y a donc ambiguïté sur le sens de commerce, mot polysémique, comme le montre, entre autres, l'expression commerce sexuel, qui n'est pas l'équivalent de l'anglais sex shop et qui ne réfère pas aux pratiques que l'on sait, impliquant rémunération.

2.2 Le sens économique de commerce, une évolution

Même dans le sens économique, neutre, du mot commerce, les dictionnaires consultés reflètent une évolution de l'acception qui va faire qu'au 20^e siècle, le terme sera de plus en plus associé au petit commerce, sauf quand il garde son sens de discipline, comme dans «école de commerce», quand il garde son sens macro-économique, comme dans «ministère du Commerce», et dans les expressions «avoir la bosse / le sens du commerce».

Nous illustrons cette évolution par la recherche des phrasèmes haut commerce, grand commerce et petit commerce. Nous ne trouvons le phrasème haut commerce (construit sur le même modèle que haute finance) que dans deux dictionnaires: le Dictionnaire général de la langue française du commencement du XVII^e siècle, jusqu'à nos jours (Hatzfeld et Darmester 1964) et le Grand Robert de la langue française (1985). Ce dernier le consigne pour désigner l'ensemble des grands commerçants en spécifiant que c'est un usage vieilli. De fait, l'expression disparaît au 20^e siècle. Inversement, l'expression petit commerce, absente des dictionnaires anciens, est consignée dans tous les

ouvrages contemporains. Le phrasème grand commerce, quant à lui, se retrouve dans les dictionnaires des 18^e et 19^e siècles, comme dans l'exemple «le grand commerce de Mofcovie eft de fourrure» (Nouveau Dictionnaire de l'Académie françoise, éd. de 1718). Dans les ouvrages ultérieurs il perdure dans le sens de commerce international. Ainsi, il figure en bonne compagnie dans le *Trésor de la langue française*, comme le montre cette citation d'Aragon: «Toutes les élégantes sont là [...]! La haute banque et le grand commerce, l'industrie en jaquette et en tube» (Aragon, Les beaux quartiers, 1936: 491, cité dans le TLF, vol. 5, 1977: 1119).

L'évolution lexicologique que nous avons relevée a sa source dans le glissement du sens du mot commerce, glissement qui s'opère à deux niveaux. Autrefois, «commerce» désignait exclusivement le commerce entre nations, le commerce international, comme dans les expressions commerce de l'Orient ou commerce des fourrures. Cette acception se reflète dans la syntaxe du verbe commercer, attesté dès 1470 : «on commerce la soie quelque part», impliquant donc voyage, commerce extérieur. La réalité intérieure (nationale) du «petit commerçant», celle de la boutique ou de l'échoppe, était désignée par les termes marchands et artisans. Ce n'est qu'au 20^e siècle qu'apparaissent, au moins dans les dictionnaires, la distinction entre commerce intérieur et commerce extérieur et encore plus récemment la notion d'import-export. L'apparition de ces syntagmes libère l'usage de commerce en le restreignant pour désigner principalement le petit commerce.

De plus, en ce qui concerne le commerce intérieur, à cause de l'évolution même du monde commercial, apparaît une autre distinction entre petit commerce et distribution (ou grande distribution; voir supra). L'étude diachronique des termes montre donc une restriction des sens de commerce, ainsi que l'apparition de connotations négatives attachées au concept de commerce. Par exemple, l'adjectif commercial s'entache d'un sens péjoratif — «conçu, exécuté dans une intention lucrative, [...] et pour plaire au plus grand public possible» —, consigné en troisième position (sur cinq acceptions) dans le *Petit Larousse* avec les exemples jazz commercial et peinture commerciale, pour dire «n'étant pas de qualité».

2.3 Le sens de distribution

Alors que pour «commerce», le premier sens est économique, pour «distribution», c'est le contraire. Le terme est général au départ, puis il prend des sens spécifiques (en droit, en imprimerie) et n'apparaît avec le sens moderne «opérations et circuits grâce auxquels les biens de consommation sont acheminés vers les points de vente» que très récemment. Au premier sens consigné à l'entrée distribution, le *Grand Robert de la langue française* renvoie à dons, largesse, libéralité (1985 : 594), au moyen de deux exemples : «Dieu, qui est roy et distributeur des royaumes» et «Dieu distribue la lumière». Ainsi, le sens commercial de distribution n'est consigné ni dans le *Petit Robert* de 1990 ni dans le *Dictionnaire historique de la langue française* (1992). Par contre, il figure dans le *Grand Robert de la langue française* (1985) avec un développement encyclopédique. Le *Trésor de la langue française* (vol. 7, 1979) se montre plus réticent à consigner l'usage moderne du mot distribution puisqu'il n'en donne pas de définition et l'insère au moyen de citations (en précisant distribution commerciale).

Si le sens commercial du terme distribution n'apparaît pas dans le *Petit Larousse* illustré de 1992, il figure par contre dans celui de 1998, avec une définition appropriée et

la collocation grande distribution. Par contre, le dérivé distributeur n'y figure pas encore dans le sens de société de distribution (enseigne), pourtant très employé dans le milieu professionnel. Le *Petit Larousse*, en avance sur les autres dictionnaires, consigne toutefois le sens commercial un peu plus général de distributeur : «personne ou firme qui assure la distribution d'un produit, service, film». On peut donc dater avec une certaine précision l'apparition du sens commercial du triplet {distribution, distributeur, distribuer} et relever le retard des dictionnaires de langue générale sur la réalité, mais aussi sur les encyclopédies, puisqu'une ancienne édition de l'*Encyclopaedia Universalis* comporte déjà un développement de cinq pages sous l'entrée distribution (1985 : 287-291).

Quant aux connotations de distribution, dans ses sens généraux, telles qu'elles apparaissent dans les citations ou les exemples donnés dans les dictionnaires, elles sont toutes positives ou neutres : distribution des revenus, des richesses, du sang dans les artères, des prix, des lettres (le distributeur de lettres); distribuer la communion. On trouve aussi, par plaisanterie, distribuer la raclée (Camus, *La chute* 1956, cité dans le TLF, vol. 7, 1979 : 333) ou «la vieille servante, [...] grande distributrice de taloches et de fessées» (Zola, *Vérité*, 1902, cité dans le TLF, vol. 7, 1979 : 333)

3. ÉTUDE LEXICOMÉTRIQUE DES PARADIGMES COMMERCE ET DISTRIBUTION DANS LA PRESSE CANADIENNE ET FRANÇAISE

L'idée de base consistait à vérifier dans des sous-ensembles de textes extraits de la presse canadienne et française actuelle si les connotations attachées aux formes du paradigme commerce étaient plus nombreuses que celles attachées au paradigme distribution. Pour ce faire, nous avons constitué les corpus «commerc» et «distrib» en utilisant des textes saisis dans la banque de données TEXTUM de l'Université de Montréal. Le contenu de cette banque a déjà été décrit ailleurs (Clas 1996 : 135-137). Nos deux corpus ont été divisés en plusieurs sous-corpus, que nous continuerons toutefois d'appeler corpus dans la suite de cet article (pour des raisons de légèreté évidentes). «Commerc» est donc composé de «commercmo» (Le Monde), «commercof» (Ouest-France) et «commercan» (Presse canadienne). De même, «distrib» se divise en trois : «distribmo», «distribof» et «districan». Ensuite, nous avons segmenté en unités minimales chacun des six nouveaux corpus avec le logiciel Lexico1 et nous avons inséré une clé de partition dans chaque corpus. Nous y avons alors recherché les concordances des formes dérivées des séries {commerce...} et {distribution...}, dans lesquelles nous avons sélectionné les contextes qui révélaient les connotations négatives ou positives attachées aux formes des deux paradigmes (recherche de cooccurrences lexicales, au sens lexicométrique; voir Lafon 1981).

Une étude préalable de la ventilation comparée des différentes formes du paradigme commerce dans nos trois corpus du *Monde*, de *Ouest-France* et de la Presse canadienne indiquait que le lemme² commerce représente entre 41 et 46 % des occurrences du paradigme commerce dans les trois corpus, alors que le lemme commerçant n'en représente que 8 à 19 %. Notons aussi l'emploi rare du verbe commercer (0,1 à 0,5 %), pris en

² Le lemme commerce regroupe tout simplement les formes commerce et commerces. De même, le lemme commercer regroupe les formes conjuguées de ce verbe. Cela dit, nous tenons à préciser que nous ne travaillons pas avec un corpus lemmatisé. Nous avons seulement procédé à un «calcul lemmatisé» ponctuel manuel.

J. Dancette et C. Réthoré

général dans le sens de faire du commerce avec l'étranger. Le même phénomène se reproduit avec le verbe distribuer, beaucoup plus rare que distribution. Le tableau qui suit résume les cooccurrences (commerce; x) dans la concordance de commerce établie à partir du corpus commercemo (*Le Monde*):

1.	accord général sur les tarifs douaniers et le commerce	165
2.	chambre de commerce (et d'industrie)	231
3.	commerce de détail	23
4.	commerce de gros	3
5.	commerce extérieur	474
6.	commerce intérieur	5
7.	commerce international	230
8.	commerce mondial	104
9.	département (américain) du commerce	74
10.	écoles (supérieures) de gestion et de commerce/ grandes	94
	écoles de commerce (et d'ingénieurs)	
11.	fonds de commerce	65
12.	grand commerce	12
13.	ministère de l'industrie et du commerce extérieur/du commerce	46
	et de l'artisanat/du commerce international et de l'industrie	
14.	ministre (délégué au) du commerce international et	217
	de l'industrie /commerce (extérieur)/artisanat	
15.	tribunal de commerce	175

Tableau 1 - Cooccurrences de commerce dans Le Monde (F = 2748)

Ce tableau de cooccurrences appelle trois remarques. La première concerne la primauté des occurrences de commerce extérieur au détriment de commerce intérieur. Ainsi, on observe cinq occurrences de l'expression commerce intérieur, contre environ 800 pour l'ensemble commerce extérieur, commerce international et commerce mondial. Dans la presse canadienne, le tableau des cooccurrences est très semblable, avec pour seule différence notable, une fréquence très inférieure pour l'expression école de commerce. La deuxième remarque porte sur les sens de commerce dans ces contextes, sens qui sont dans la plupart des cas neutres, macroéconomiques, notamment lorsque le commerce désigne le secteur d'activité ou la discipline enseignée. Enfin, notons que les connotations négatives sont absentes de ce relevé de cooccurrences.

Ainsi, la concordance du terme commerce ne nous permet pas de déceler les connotations négatives que nous supposions, au début de cette étude, attachées au petit commerce. Conscients du caractère exploratoire de notre recherche, nous devrons ultérieurement approfondir les raisons de cette absence, notamment en utilisant d'autres

corpus. En effet, humbles et prudents devant les chiffres, nous nous gardons de toute extrapolation hâtive et risquée (à ce sujet, voir Tournier 1980).

L'étape suivante consistait à rechercher les cooccurrences de «commerçant». C'est là que nous avons pu valider en partie notre hypothèse sur les connotations négatives qui expliqueraient, entre autres facteurs, le changement de paradigme de commerce à distribution, et de commerçant à distributeur. Les graphes suivants représentent les réseaux associatifs des cooccurrences de «commerçant» dans les corpus commercmo (France) et commercan (Canada). Nous avons regroupé les contextes canadiens (405) et français (125) que comporte notre concordance, et avons étiqueté (empiriquement) les sept catégories qui se dessinaient : (1) classe sociale; (2) qualificatifs (positifs ou négatifs); (3) actions (positives ou négatives); (4) problèmes (créés ou subis par les commerçants); (5) perceptions de la société (clichés); (6) mouvements-regroupements; (7) origine ethnique. Ces catégories contribuent toutes à révéler les connotations associées à «commerçant». De plus, celle de la classe sociale permet de situer le commerçant par rapport à d'autres groupes sociaux.

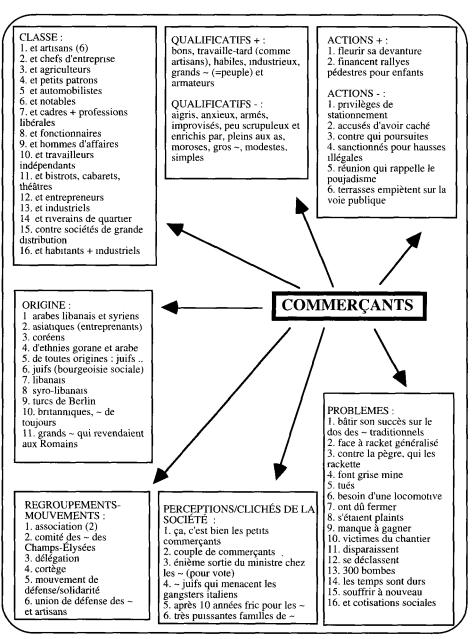


Fig. 1 - Réseau associatif des cooccurrences de commerçants dans le corpus du Monde (commercmo)

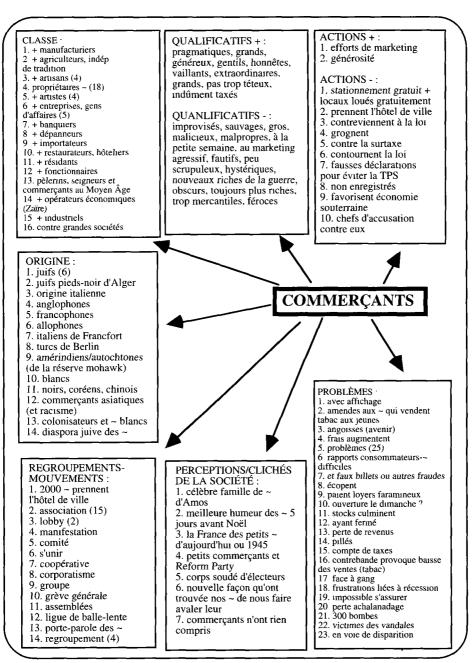


Fig. 2 - Réseau associatif des cooccurrences de commerçants dans le corpus de la presse canadienne française (commercan)

Les réseaux de cooccurrences ou réseaux associatifs permettent d'opérer la jonction entre statistique, terminologie et analyse du discours. Ces outils, très porteurs, sont de plus en plus utilisés (voir Debarbieux et Ravix 1993 : 554-555; Bourion 1995). Les cooccurrences, étudiées catégorie par catégorie, révèlent les associations faites avec «commerçant» dans la presse. Nous pensons que ces associations ne sont pas le fruit du hasard — un hasard théorique qui ferait que les unités de la langue auraient une distribution équiprobable —, mais qu'elles contribuent à forger (ou révéler) l'image du commerçant. Ainsi, pour ne prendre que quelques exemples, nous repérons, dans la catégorie classe, la cooccurrence (positive) de commerçants et artisans. Mais dans les qualificatifs, nous remarquons que les éléments négatifs sont plus nombreux que les positifs. Dans les deux catégories mouvements-regroupements et perceptions de la société, nous remarquons que le côté associatif des commerçants est perçu de façon négative, comme le montrent la référence explicite au poujadisme dans le corpus français et les références au corporatisme dans la presse canadienne française.

La catégorie origine est certainement très révélatrice. Les qualificatifs de nationalité (commerçants juifs, asiatiques, turcs) traduisent un phénomène connu, peut-être basé sur un fondement objectif mais aussi sur des préjugés ou une image sociale stéréotypée : la propension des émigrés à ouvrir un commerce lorsqu'ils arrivent à l'étranger (commerçants turcs à Berlin, italiens à Francfort, portugais de la rue Prince-Arthur de Montréal, etc.). Mais il est intéressant de relever que les nationalités «non marginales» (c'est-à-dire majoritaires) ne sont jamais notées. Ainsi, nous ne trouvons pas commerçants «français», ni «américains», ni «allemands». Dans le corpus canadien, nous retrouvons les marques de deux réalités propres au Québec : les commerçants francophones par opposition aux anglophones ou aux allophones, à cause des problèmes liés aux règlements sur l'affichage, et les cooccurrences de commerçants avec autochtones et amérindiens (dans les réserves) associées aux avantages fiscaux dont ils semblent jouir, selon la presse et l'opinion publique. Quant à l'étude parallèle des cooccurrences de distributeur, elle n'a révélé aucune connotation négative.

CONCLUSION

Nous avons présenté dans cet article certains des résultats tirés d'une étude lexicométrique sur les triplets {commerce, commerçant, commercer} et {distribution. distributeur, distribuer}, en mettant l'accent sur le contraste France-Québec et sur les couples commerce-commerçant et distribution-distributeur. En raison des contraintes matérielles imposées dans cet article, les fréquences inférieures des verbes commercer et distribuer justifiaient ce choix. Cette étude exploratoire devait apporter des éléments de réponse à la question que nous posions sur les connotations négatives comme facteur de changement du paradigme lexical commerce-commerçant à celui de distribution-distributeur.

Nous sommes bien conscients des limites de telles études pour vérifier des phénomènes de langue complexes : un corpus ne révèle que ce qu'il contient. Les articles de journaux sont révélateurs de la presse écrite, d'une volonté d'information sur des sujets d'intérêt général, voire national. Ils sont inscrits dans le temps présent et peuvent faire l'objet d'une auto-censure. De plus, nous n'avons pu incorporer dans notre corpus les articles de journaux datant des périodes «sensibles» telles que l'immédiat après-guerre.

Néanmoins, à l'issue de cette étude et malgré les limites que nous venons d'énoncer, quelques remarques s'imposent. D'une part, les réseaux associatifs du terme commerçant font nettement ressortir les images négatives attachées à ce terme. Ces résultats vont dans le sens de l'hypothèse que nous avions émise comme explication du changement d'appellation chez les grands commerçants qui, pour se distinguer des petits commerçants, adoptent le terme distributeur. D'autre part, nous sommes surpris par le peu de différence, sur le plan quantitatif, entre les listes de connotations négatives au Québec et en France. En supposant que les commerçants installés au Québec avaient un passé moins lourd que leurs homologues français (collaboration), on aurait pu s'attendre à ce que les connotations négatives qui pèsent sur la classe des commerçants fussent moins nombreuses en Amérique du Nord. Mais l'eau est passée sous les ponts depuis la Seconde Guerre mondiale. Notre étude des connotations montre que les préjugés ressortent de toutes les façons sans qu'il y ait d'événements dramatiques les nourrissant. Nous avons également montré que la Seconde Guerre mondiale n'a été qu'un des facteurs, parmi d'autres, qui ont contribué à ternir la réputation des commerçants.

Sur un plan plus général, nous pensons avoir illustré dans cette étude des paradigmes commerce et commerçant le rôle, souvent inconscient, des connotations comme facteur déterminant d'évolution terminologique. Les mots ont une mémoire !

RÉFÉRENCES

- Académie française (1994): Nouveau dictionnaire de l'Académie françoise, édition de 1718, Genève. Slatkine Reprints.
- Al-HARÎRÎ, A. (1992): Le Livre des malias, séances d'un vagabond de génie, Phébus, Paris.
- ATTALI, J. (1991): 1492, Paris, Fayard.
- BALZAC, H. de (1963): La Maison du Chat-qui-pelote; Le bal de Sceaux; La Vendetta, collection «Classiques Garnier», Paris, Garnier.
- BALZAC, H. de (1975): César Birotteau, édition de 1990, collection «Lire et voir les classiques», Paris, Pocket.
- BENOUN, M. et M.-L. HÉLIÈS-HASSID (1995): Distribution, Acteurs et stratégies, 2^e éd., collection «Gestion», Paris, Économica.
- BOURION, É. (1995): «Le réseau associatif de la peur». F. Rastier (dir), L'analyse thématique des données textuelles. L'exemple des sentiments, collection «Études de sémantique lexicale», Paris, Didier Érudition.
- BRAUDEL, F. (dir) (1985): La Méditerranée : l'espace et l'histoire, collection «Champs», Paris, Flammarion.
- CANTIN, A. (1996): «Négociation: la méthode secrète du roi Carrefour», L'Essentiel du Management, n° 19, p. 82-90.
- CHARMES, C. (1989) : Cours de politique générale, fascicule 2 (Politique générale et culture), Angers (France), ESSCA.
- CHUNG, C.H., SHEPARD, J.M et M.J. DOLLINGER (1989): «Max Weber Revisited: Some Lessons from East Asian Capitalistic Development», Asia Pacific Journal of Management, vol. 6, n° 2, pp. 307-321.

- CLAS, A. (1996): «Dictionnaires généraux bilingues: Le Dictionnaire bilingue canadien», Cahiers de lexicologie, n° 69, pp. 127-144.
- CLOSEL, J. du (1989): Les grands magasins français. Cent ans après, collection «Connaissance du commerce et de la consommation», Paris, ICC et Chotard et Associés Éditeurs.
- COQUERY, M. (1977): Mutations et structures du commerce de détail en France: étude géographique, Cergy Village (France), Éditions du Signe.
- DANCETTE, J. et C. RÉTHORÉ (1997): Dictionnaire bilingue de la distribution/Bilingual Dictionary of Retailing, version préliminaire, Université de Montréal, Département de linguistique et de traduction.
- DANCETTE, J. et C. RÉTHORÉ (1997): «Le Dictionnaire bilingue de la distribution: entre dictionnaire de langue et encyclopédie, *Meta*, vol. 42, n° 2. Communication présentée aux IV^{es} Journées scientifiques du réseau «Lexicologie, terminologie, traduction» de l'AUPELF-UREF (Lyon, France, 28-30 septembre 1995).
- DAUZAT, A., DUBOIS, J. et H. MITTERAND (1972): Nouveau dictionnaire étymologique et historique, 3^e éd. rev. et corr., Paris, Larousse.
- DAVID, M. (1995): «Il était une fois le nouveau commerce», LSA, n° 1000.
- DEBARBIEUX, B., et B. RAVIX (1993): «Le territoire en discours: Saint-Gervais-les-Bains», Actes des secondes journées internationales d'analyse statistique de données textuelles, Paris, S.J. Anastex Éditeur, pp. 551-562.
- DUCROCQ, C., JAMIN, N. et S. LAGRANGE (1994): La distribution, collection «Vuibert Entreprise», Paris, Vuibert.
- DUTOURD, J. (1952): Au bon beurre: ou, dix ans de la vie d'un crémier, Paris, Folio.
- Encyclopaedia Universalis (1995-1996): Paris, Encyclopaedia Universalis.
- FAUVEAU, J.-C. (1991): Le monde de la distribution, Noiziel (France), Les Presses du Management.
- FAWCETT, H. (1989): A Text Searching System. PAT 3.3. User's Guide, Waterloo, UW Centre for the New Oxford English Dictionary.
- FURETIÈRE, A. (1690): Dictionnaire universel contenant généralement tous les mots françois tant vieux que modernes, et les termes de toutes les sciences et des arts..., La Haye, Leers.
- HALLIER, J.-E. (1972): La cause des peuples, Paris, Seuil.
- HARDOUIN, P. et C. DUCROCQ (1993): Le commerce indépendant organisé. Rapport du groupe de travail de l'ICC, collection «Connaissance du commerce et de la consommation», Paris, ICC et Éditions Liaisons.
- HATZFELD, A., et A. DARMESTER (1964): Dictionnaire général de la langue française du commencement du XVII^e siècle, jusqu'à nos jours. Précédé d'un Traité de formation de la langue, Paris, Delagrave.
- HUGUET, E. (1925): Dictionnaire de la langue française du seizième siècle, Paris, Champion.

- IMBS, P. (dir) (1971-1994): Trésor de la langue française : dictionnaire de la langue du XIXe et du XXe siècle (1789-1960), Paris, Éditions du CNRS
- LAFON, P. (1980): «Sur la variabilité de la fréquence des formes dans un corpus», *Mots*, n° 1, pp. 127-165.
- LAFON, P. (1981): «Analyse lexicométrique et recherche des cooccurrences. Des rapports syntagmatiques aux cooccurences», *Mots*, n° 3, pp. 95-148.
- LAFON, P. (1984): Dépouillements et statistiques en lexicométrie. Genève-Paris, Slatkine-Champion.
- LEBART, L. et A. SALEM (1994): Statistique textuelle, Paris, Dunod.
- Les Mille et Une Nuits. Histoire de Sindbad le Marin, trad. d'Antoine Galland (1996) : collection «Librio», Paris, Flammarion.
- LITTRÉ, É. (1873): Dictionnaire de langue française, Paris, Hachette.
- Petit Larousse illustré (1992): Paris, Larousse.
- Petit Larousse illustré (1998): Paris, Larousse.
- REY, A. et al. (1992): Dictionnaire historique de la langue française, Paris, Dictionnaires Le Robert.
- ROBERT, P. (1985): Grand Robert de la langue française, 2^e éd., Paris, Le Robert.
- ROBERT, P. (1990): Dictionnaire alphabétique et analogique de la langue française: le Petit Robert 1. Paris/Montréal. Le Robert/Robert-Canada.
- ROBERT, P. (1998): Dictionnaire alphabétique et analogique de la langue française: le Petit Robert 1, Paris/Montréal, Le Robert/Robert-Canada.
- SALEM, A. (1990): Lexico1, version 0.7, Saint-Cloud (France), Lexicométrie et textes Politiques, INaLF-CNRS/ENS de Fontenay-St.Cloud.
- SALEM, A. (1992): Lexico1-Version 0.7. Notice d'utilisation, Saint-Cloud (France), Lexicométrie et textes Politiques, INaLF-CNRS/ENS de Fontenay-St.Cloud.
- SÉDILLOT, R. (1964): Histoire des marchands et des marchés, Paris, Fayard.
- THIL, E. (1966): Les inventeurs du commerce moderne, Paris, Arthaud.
- TOURNIER, M. (1980): «D'où viennent les fréquences de vocabulaire?», *Mots*, n° 1, pp. 189-212.
- WEBER, M. (1964): L'éthique protestante et l'esprit du capitalisme, traduit de l'allemand par J. Chavy, Paris, Plon.
- YDEWALLE, C. d' (1965): Au Bon Marché, de la boutique au grand magasin, Paris, Plon.
- ZOLA, É. (1990): Au Bonheur des Dames, édition de 1883, collection «Lire et voir les classiques», Paris, Pocket.

